

Republic of Iraq
Ministry of Higher Education & Scientific Research
University of Kerbala
College of Engineering
Electrical and Electronic Engineering Department



Real-Time Scheme for Speech Security in Public Communication based on Embedding Method into Audio Channel

**A Thesis Submitted to the Department of Electrical and Electronic
Engineering, University of Kerbala in Partial Fulfillment of the
Requirements for the Degree of Master of Science (MSc) in
Electrical Engineering**

By

Baneen Qasem Abd-Ali

**B.Sc. in Electrical and Electronic Engineering 2015/ University of
Kerbala**

Supervised by

Prof. Dr. Haider Ismael Shahadi

Assist. Prof. Dr. Muayad Saleem Kod

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

وَقُلْ رَبِّيَ زِدْنِي

عِلْمًا

صَدَقَ اللَّهُ الْعَلِيَّ الْعَظِيمَ

سُورَةُ طه الأية ١١٤

SUPERVISOR CERTIFICATE

I certify that this thesis entitled "Real-Time Scheme for Speech Security in Public Communication based on Embedding Method into Audio Channel", which is prepared by "Baneen Qasem Abd-Ali", under my supervision at University of Kerbala in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering.

Signature: 

Prof. Dr. Haider Ismael Shahadi

(First Supervisor)

Date: 26/4 / 2022

Signature: 

Assist. Prof. Dr. Muayad Saleem Kod

(Second Supervisor)

Date: 26/4 / 2022

LINGUISTIC CERTIFICATION

I certify that this thesis entitled "Real-Time Scheme for Speech Security in Public Communication based on Embedding Method into Audio Channel" which is prepared by "Baneen Qasem Abd-Ali", under my linguistic supervision. It was amended to meet the English style.

Signature: 

Name: Assist. Prof. Dr. Ali Jafer Mahdi

Date: 8 / 5 / 2022


EXAMINATION COMMITTEE CERTIFICATION

We certify that we have read the thesis entitled "Real-Time Scheme for Speech Security in Public Communication based on Embedding Method into Audio Channel" and as an examining committee, we examined the student "Baneen Qasem Abd-Ali" in its content and in what is connected with it and that in our opinion it is adequate as a thesis for the degree of Master of Science in Electrical Engineering.

Signature: 

Prof. Dr. Haider Ismael Shahadi
(First Supervisor)

Date: 26/4/2022

Signature: 

Assist. Prof. Dr. Muayad Saleem Kod
(Second Supervisor)

Date: 26/4/2022

Signature: 

Prof. Dr. Hawraa Hassan Abbas
(Member)

Date: 26/4/2022

Signature: 

Assist. Prof. Dr. Mithaq Nama Raheema
(Member)

Date: 26/4/2022

Signature: 

Prof. Dr. Samir Jasim Mohammed
(Chairman)

Date: 26/4/2022

Approval of the Department of Electrical and Electronic Engineering.

Signature: 

Name: Prof. Dr. Haider Ismael Shahadi

(Head of Electrical and Electronic Engineering Dept.)

Date: 8/5/2022

Approval of Deanery of the College of Engineering/ University of Kerbala.

Signature: 

Name: Prof. Dr. Laith Sh. Rasheed

(The Dean of the College of Engineering)

Date: 11/5/2022

DEDICATION

Every challenging work needs
self-efforts as well as guidance and encouragement
from those are very close to our hearts.

My humble effort I dedicate to
my father, mother, husband,
and all my teachers.

ACKNOWLEDGMENT

First, I thank God Almighty, who gave me the strength and patience to complete this project successfully. I also would like to express my special thanks of gratitude to my supervisor Prof. Dr. Haider Ismael, who taught me perseverance in the face of obstacles through his persistent advices and guidance's, which taught me how to rely on myself to find solutions to any problem I face.

Thanks a lot for Assist. Prof. Dr. Muayad and Assist. Prof. Dr. Hameed for all the notes and instructions that you gave me to complete this work within the limited time frame. Our principal Prof. Dr. Ali Abdul Razzaq who gave me the golden opportunity to do this wonderful project on the topic Real-Time Scheme for Speech Security in Public Communication based on Embedding Method into Audio Channel, which also helped me in doing a lot of research and I came to know about so many new things I am really thankful to them.

Abstract

Voice over Internet Protocol (VoIP) is a popular and important internet protocol for real-time voice calling. It is used in several software applications such as Skype, WhatsApp, and Google Talk. However, communications over the internet can easily be exposed to hacking and eavesdropping. In this work, a real-time covert communication approach based on audio steganography is introduced; by generating a secure communication channel into a VoIP network. To implement this approach, two models are introduced.

In the first model, a very fast method based on lifting wavelet transform scheme. Lifting wavelet have important specifications over other transform techniques make it superior in time execution. While the second model is based on internet Low Bit Rate (iLBC) Codec as VoIP standard Codec with auto-correction for speech quality degradation over a lossy channel. For both models, G.711 Codec is used to compress the cover speech at synchronous periods (for each 16 or 20 msec) with respect to the secret speech to meet the VoIP requirements.

After that, the compressed cover data is used to embed the corresponding compressed secret data in real-time for each period individually. At a lossless noisy channel, the quality of the stego-signals is preserved at greater than 41 dB (for the first proposed model) and 40 dB (for the second proposed model) in terms of Signal to Noise Ratio (*SNR*). Also, a high embedding capacity up to 12.5% and 25% from the size of the cover speech for the first and second

proposed models, respectively. Moreover, the secret data is fully retrieved without error. In a noisy channel, the robustness test results show that the proposed approaches can immune additive channel noise up to 35 dB in terms of *SNR* for the first model and 38dB for the second model. Also, the secret data is retrieved without perceptible difference for the human ear, where the value of the Normalized Cross Correlation (*NC*) is not less than 0.92 and the value of Bit Error Rate (*BER*) is less than 0.0035 for each of the two introduced models. At a different levels of packet losses, the quality for the stego-signals and hidden data is maintained at 0.88 and 35dB as average values in terms of *NC* and *SNR* respectively.

CHAPTER ONE INTRODUCTION	1
1.1 Overview	1
1.2 Data Hiding Types	2
1.3 Data Hiding Applications.....	4
1.4 Real-Time Steganography versus Offline.....	5
1.5 Review of the Related Work.....	6
1.5.1 Offline Steganography	6
1.5.2 Embedding before Compression.....	8
1.5.3 Embedding after Compression.....	12
1.6 Problem Statement	18
1.7 Objective of the Study.....	19
1.8 Contributions.....	20
1.9 Thesis Outlines.....	21
CHAPTER TWO BACKGROUND AND THEORETICAL CONCEPTS	23
2.1 Introduction	23
2.2 VoIP Communication Model	23
2.2.1 G.711 Codec.....	25
2.2.2 Internet Low Bit Rate Codec (iLBC).....	26
2.3 Wavelet Transform (WT).....	27
2.3.1 Discrete Wavelet Transform (DWT)	29
2.3.2 Lifting Wavelet Transform (LWT)	31
2.4 Audio Steganography Techniques	35
2.4.1 Least Significant Bit (LSB) technique	36

2.4.2 Phase Coding.....	37
2.4.3 Parity Coding.....	38
2.4.4 Echo Hiding.....	38
2.4.5 Spread Spectrum (SS)	39
2.5 Pseudo-Random Number Generator (PRNG).....	40
2.6 Steganography Requirements.....	41
2.6.1 Data Embedding Rate	41
2.6.2 Perceptual Quality	42
2.6.3 Robustness.....	44
2.6.4 Security.....	45
2.6.5 Computational Complexity	45
2.7 Chapter Summary.....	46

CHAPTER THREE THE PROPOSED MODELS FOR REAL-TIME COVERT COMMUNUCATION 48

3.1 Introduction	48
3.2 Real-Time Covert Communication Approach based on VoIP	48
3.3 The Proposed Model based on Int2Int -Haar LWT CODEC.....	52
3.3.1 Embedding Scheme.....	52
3.3.2 Recovery Scheme	56
3.4 The Proposed Model based on iLBC CODEC	58
3.4.1 Embedding Scheme.....	58
3.4.2 Recovery Scheme	59
3.5 Chapter Summary.....	60

CHAPTER FOUR EXPERIMENTAL RESULTS AND DISCUSSION 62

4.1 Introduction	62
4.2 Results of the Proposed Model based on Int2Int-Haar LWT CODEC.....	62
4.2.1 Tests of Perceptual Quality and Data Embedding Rate.....	65

4.2.2 Tests of Robustness against Additive Noise	69
4.3 Results of the Proposed Model based on iLBC CODEC.....	74
4.3.1 Tests of Perceptual Quality and Data Embedding Rate.....	77
4.3.2 Tests of Robustness against Additive Noise	81
4.3.3 Test of Perceptual Quality in a Lossy Channel with Different Levels of Packet Losses.....	83
4.4 Computational Complexity and Processing Time	85
4.5 Discussion of the Proposed Approach	86
4.5.1 Perceptual Quality	86
4.5.2 Robustness Against Additive Noise.....	88
4.6 Comparison between the Two Proposed Models.....	89
4.7 Comparison with the Related Work.....	90
4.7.1 Perceptual Quality and Data Embedding Rate.....	90
4.7.2 Robustness of the Proposed Models	91
4.8 Chapter Summary.....	92
CHAPTER FIVE.....	93
5.1 Conclusion.....	93
5.2 Future Work	94

Abbreviations

<i>ADC</i>	<i>Analog to Digital Converter</i>
<i>AES</i>	<i>Advanced Encryption Standard</i>
<i>AWGN</i>	<i>Additive White Gaussian Noise</i>
<i>BER</i>	<i>Bit Error Rate</i>
<i>CWT</i>	<i>Complex Wavelet Transform</i>
<i>DAC</i>	<i>Digital to Analog Converter</i>
<i>DFT</i>	<i>Discrete Fourier Transform</i>
<i>DSL</i>	<i>Digital Speech Level Analyzer</i>
<i>DSP</i>	<i>Digital Signal Processing</i>
<i>DSSS</i>	<i>Direct Spread Spectrum</i>
<i>DWT</i>	<i>Discrete Wavelet Transform</i>
<i>FFT</i>	<i>Fast Fourier Transform</i>
<i>FHSS</i>	<i>Frequency Hopping Spread Spectrum</i>
<i>FPGA</i>	<i>Field Programmable Gate Array</i>
<i>GSM</i>	<i>Global System for Mobile communications</i>
<i>HAS</i>	<i>Human Auditory System</i>
<i>HVS</i>	<i>Human Visual System</i>
<i>iLBC</i>	<i>internet Low Bit Rate Codec</i>
<i>Int2Int-Haar LWT</i>	<i>Integer to integer Haar Lifting Wavelet Transform</i>
<i>IP</i>	<i>Internet Protocol</i>
<i>ITU-T</i>	<i>International Telecommunication Union</i>
<i>LSB</i>	<i>Least Significant Bit</i>
<i>LWT</i>	<i>Lifting Wavelet Transform</i>
<i>MELP</i>	<i>Mix Excitation Linear Prediction</i>
<i>MLCG1</i>	<i>Multiplicative Linear Congruential</i>
<i>NC</i>	<i>Normalized Cross Correlation</i>

<i>PCC</i>	<i>Parabolic Curve Cryptography</i>
<i>PCM</i>	<i>Pulse Code Modulation</i>
<i>PESQ</i>	<i>Perceptual Evaluation of Speech Quality</i>
<i>PLC</i>	<i>Packet Loss Correction system</i>
<i>PRNG</i>	<i>Pseudo Random Number Generator</i>
<i>PSV</i>	<i>Partial Similarity Value</i>
<i>PWT</i>	<i>Packet Wavelet Transform</i>
<i>RTCP</i>	<i>Real-time Control Protocol</i>
<i>RTP</i>	<i>Real-time Protocol</i>
<i>SNR</i>	<i>Signal to Noise Ratio</i>
<i>SS</i>	<i>Spread Spectrum</i>
<i>STFT</i>	<i>Short-time Fourier Transform</i>
<i>VAD</i>	<i>Voice Activity Detection</i>
<i>VADDI</i>	<i>Voice Activity detection Dynamic Insertion</i>
<i>VAMI</i>	<i>Value based Multiple Insertion</i>
<i>VODO</i>	<i>Voice Damage Offset</i>
<i>VoIP</i>	<i>Voice over Internet Protocol</i>
<i>WBM</i>	<i>Weighted Block Matching</i>
<i>WCA</i>	<i>Wavelet Coefficient Adaption</i>
<i>WT</i>	<i>Wavelet Transform</i>

Symbols

A	<i>Compression Parameter</i>
A	<i>The Well-chosen Multiplier</i>
A_1	<i>First Approximation Sub-band</i>
A_2	<i>Second Approximation Sub-band</i>
A_3	<i>Third Approximation Sub-band</i>
B	<i>Total Number of Secret Speech Bits</i>
b	<i>Bit order</i>
C	<i>Original Cover Signal</i>
C_i	<i>Cover Frame</i>
C_{ij}	<i>Cover Sample</i>
C_{i_comp}	<i>Compressed Cover Frame</i>
$C_{ik_Candidate}$	<i>Candidate Cover Sample</i>
$C_{ik_Candidate_HB}$	<i>Candidate Cover Sample after Extracting HB</i>
\hat{C}	<i>Modified Cover (Stego) Data</i>
\hat{C}_{ik_HB}	<i>Candidate Stego Sample Without the HB</i>
\hat{C}_{ik}	<i>Candidate Stego Sample With the HB</i>
\hat{C}_i	<i>Stego Frame</i>
d	<i>Secret Digit Order</i>
d_i	<i>Delay for Bit 0 ($i=0$) and bit 1 ($i=1$)</i>
D_1	<i>First Details Sub-band</i>
D_2	<i>Second Details Sub-band</i>
D_3	<i>Third Details Sub-band</i>
$G_0(z), H_0(z)$	<i>Low Pass Filter</i>
$G_1(z), H_1(z)$	<i>High Pass Filter</i>
$h(n)$	<i>Echo Kernel</i>
HB	<i>Number of Hold Bits from LSB Direction</i>

I	<i>Frame Order</i>
J	<i>Sample Order</i>
K	<i>Candidate Cover Sample Order</i>
L_C	<i>Number of Samples in The Cover Frame</i>
L_S	<i>Number of Samples in The Secret Frame</i>
M	<i>The Selected Number for PRNG</i>
M	<i>Number of Candidate Cover Samples</i>
mod	<i>Modulo Operation</i>
N	<i>Number of Frames</i>
N_C	<i>Number of Cover Speech Samples</i>
N_S	<i>Number of Secret Speech Samples</i>
P	<i>Prediction Unit</i>
P	<i>Permutation</i>
P^{-1}	<i>Inverse Permutation</i>
R	<i>Remainder</i>
$S_{inp}(n)$	<i>Input Audio Signal</i>
$S_{out}(n)$	<i>Output Audio Signal</i>
S	<i>Original Secret Signal</i>
\hat{S}	<i>Retrieved Secret Signal</i>
S_i	<i>Secret Frame</i>
S_{i_comp}	<i>Compressed Secret Frame</i>
$S_{i_comp_P}$	<i>Permuted Compressed Secret Frame</i>
$S_{ij_comp_P}$	<i>Permuted Compressed Secret Sample</i>
$S_{ijd_comp_P}$	<i>Permuted Compressed Secret Digit</i>
$sign(x)$	<i>Sign Function</i>
S -Key	<i>Secret Key</i>
Thr	<i>Cover Sample Strength</i>
U	<i>Update Unit</i>

X	<i>Input Signal</i>
$X_n(e)$	<i>Even Indexed Samples</i>
$X_n(o)$	<i>Odd Indexed Samples</i>
x_0	<i>Initial State</i>
x_n	<i>Input Number to PRNG</i>
x_{n+1}	<i>Output Random Number</i>
Y	<i>Output Signal</i>
α	<i>Echo amplitude</i>
$\delta(n)$	<i>Impulse Function</i>
$2\uparrow$	<i>Interpolation by Two</i>
$2\downarrow$	<i>Decimation by Two</i>

List of Tables

Table 2.1 A summary of offline steganography approaches.	7
Table 2.2 A summary of stego-method that achieved before coding process.	11
Table 2.3 A summary of stego-method that achieved after coding process.	16
Table 4.1 Tests the quality of the stego-signals for the first proposed model (the inputs are recorded speeches with a 32khz cover sampling frequency). 66	
Table 4.2 Tests the quality of the stego-signals for the first proposed model (the inputs are speeches from dataset with a 32khz cover sampling frequency).	67
Table 4.3 Tests the quality of the stego-signals for the first proposed model (the inputs are recorded speeches with a 44.1 khz cover sampling frequency).68	
Table 4.4 Tests the quality of the stego-signals for the first proposed model (the inputs are speeches from dataset with a 44.1khz cover sampling frequency).	68
Table 4.5 Robustness test results for the first proposed model (the inputs are recorded speeches with a 32khz cover sampling frequency).....	70
Table 4.6 Robustness test results for the first proposed model (the inputs are speeches from dataset with a 32khz cover sampling frequency).....	71
Table 4.7 Robustness test results for the first proposed model (the inputs are recorded speeches with a 44.1khz cover sampling frequency).....	71
Table 4.8 Robustness test results for the first proposed model (the inputs are speeches from dataset with a 44.1khz cover sampling frequency).....	72

Table 4.9 Some tests for the immunity of the secret data against additive noise for the first proposed model.	73
Table 4.10 Tests the quality of the stego-signals for the second proposed model (the inputs are recorded speeches with a 32khz cover sampling frequency).	78
Table 4.11 Tests the quality of the stego-signals for the second proposed model (the inputs are speeches from dataset with a 32khz cover sampling frequency).	78
Table 4.12 Tests the quality of the stego-signals for the second proposed model (the inputs are recorded speeches with a 44.1khz cover sampling frequency).	79
Table 4.13 Tests the quality of the stego-signals for the second proposed model (the inputs are speeches from dataset with a 44.1khz cover sampling frequency).....	80
Table 4.14 Robustness test results for the second proposed model (the inputs are recorded speeches).	81
Table 4.15 Robustness test results for the second proposed model (the inputs are speeches from dataset).....	82
Table 4.16 Some tests for the immunity of the secret speech against additive noise for the second proposed model.	83
Table 4.17 An example on the varying of the <i>HB</i> factor	89
Table 4.18 A comparison between the two proposed models.	89
Table 4.19 Comparative results for the quality and data embedding rate.	91

List of Figures

Figure 1.1 VoIP steganography model.....	4
Figure 2.1 VoIP network structure	25
Figure 2.2 Time frequency structure (a) STFT and (b) WT.	28
Figure 2.3 Two channel filter bank.	30
Figure 2.4 Three level dwt decomposition.....	30
Figure 2.5 Three level inverse dwt reconstruction.....	31
Figure 2.6 One level LWT, (a) decomposition and (b) reconstruction...	33
Figure 2.7 LSB mechanism.	37
Figure 2.8 Internal components of DSLA.....	43
Figure 2.9 Measurements of stego speech quality.	43
Figure 3.1 The introduced real-time covert communication scheme.	49
Figure 3.2 Flowchart of the embedding scheme for the proposed approach.	50
Figure 3.3 Flowchart of the recovery scheme for the proposed approach.	51
Figure 3.4 The embedding scheme of the first proposed model.....	55
Figure 3.5 Framing process.....	55
Figure 3.6 The recovery scheme of the first proposed model.....	57
Figure 3.7 The embedding scheme of the second proposed model.	59
Figure 3.8 The recovery scheme of the second proposed model.....	60
Figure 4.1 The processed cover speech in the first proposed model (a) original, (b) compressed, (c) stego-data, and (c) decompressed cover speech.	64

Figure 4.2 The processed secret speech in the first proposed model (a) original, (b) compressed, (c) compressed retrieved, and (c) decompressed retrieved.65

Figure 4.3 The retrieved compressed secret data (a, b, c, and d) and decompressed secret speech (e, f, g, and h) from the first proposed model with AWGN = 30dB at $HB=0, 1, 2,$ and 3 respectively.....74

Figure 4.4 The processed cover speech in the second proposed model (a) original, (b) compressed, (c) stego-data, and (c) decompressed cover speech.76

Figure 4.5 The processed secret speech in the second proposed model (a) original, (b) compressed, (c) compressed retrieved, and (c) decompressed retrieved.....77

Figure 4.6 Testing the second proposed system quality at a lossy channel with random packet losses.....84

Figure 4.7 Tests of the effecting of the packet losses on the second proposed model (a) *SNR* of the received stego data vs. Packet losses (b) *NC* of the retrieved secret speech vs. Packet losses.85

CHAPTER One

INTRODUCTION

1.1 Overview

Nowadays, shared data over the internet is widely used. This sharing should be implemented in a secure manner to face the continuous challenges and threats in the field of security and electronic criminals. There are three main methods for security: cryptography, watermarking, and steganography [1]. Cryptography is based on encrypting the transmitted data into ambiguous data for unauthorized people. But it gives a clear indicator to the attackers about how the encrypted data is important for the transmitter [2].

In watermarking, digital data are hidden in the host data in a robust and perceptible or imperceptible manner to convey some details about copyright and ownership. In order to delude eavesdroppers, intelligible data are sent by transmitter but they convey hidden messages. This security approach is called steganography. The word steganography was derived from the Greek words where "stegano" means covered or confident and "graphia" means writing or drawing, so that it is defined as covered or concealed writing [3].

Steganography can be categorized according to the type of the cover signals into image, audio, and videos [4], or according to the type of processing into

offline or online (which requires processing in real-time) [5]. Real-time steganography is more preferred in recent researcher works because it enhances the security level for the confident data. An efficient real-time steganographic system is based on many contradictory requirements [6] including high data embedding rate, perceptual quality, security, robust scheme, low algorithm complexity, and suitable processing time (or latency) proper for real time communication [5].

Real time steganography can be achieved in mobile communication based Global System for Mobile communications (GSM), or internet voice communications based Voice over Internet Protocol (VoIP). VoIP is more popular than GSM because it provides global services with low or free cost and high reliability [7]. VoIP service allows users to make voice calls over the internet protocols. In steganography, these calls can be used as carrier for the secret data to form the term VoIP steganography. The secret data may be any multimedia content offline (previously recorded or saved) or online (in real-time) [8].

1.2 Data Hiding Types

Basically, there are two types of hiding data; blind and non-blind. In blind approach, the original cover is not required for recovery process such as steganography while non-blind approach requires the original cover to retrieve the secret data. Watermarking is an example about non-blind data hiding [5].

In watermarking, the digital watermark is embedded into a carrier in such a way that its removal or manipulation is impossible. Therefore, the robustness against attempts to violation data integrity or copyright is important factor in any watermarking algorithm [9].

While in steganography, high embedding capacity with acceptable imperceptibility are the dominate factors [10]. Moreover, steganography requires that the entire secret data is concealed into a carrier to be invisible, if these data are disclosed, steganography will fail. In contrast, the digital watermark may be visible or invisible [11].

The main challenge in steganography is that it requires high bandwidth to transmit the secret message. The existence of high capacity approaches of steganography may embed up to 25% of the cover size. This leads to making the size of the transmitted data about four times the message size or may much more [12].

Figure (1.1) shows an illustration of a general steganographic model. Message represents the hidden data that the transmitter wants to send it to the receiver in a secure manner. A stego-key is a critical information that is used to ensure only the receiver who has the decoding key can retrieve the secret data. The cover signal that conveys the hidden message is called a stego-signal.

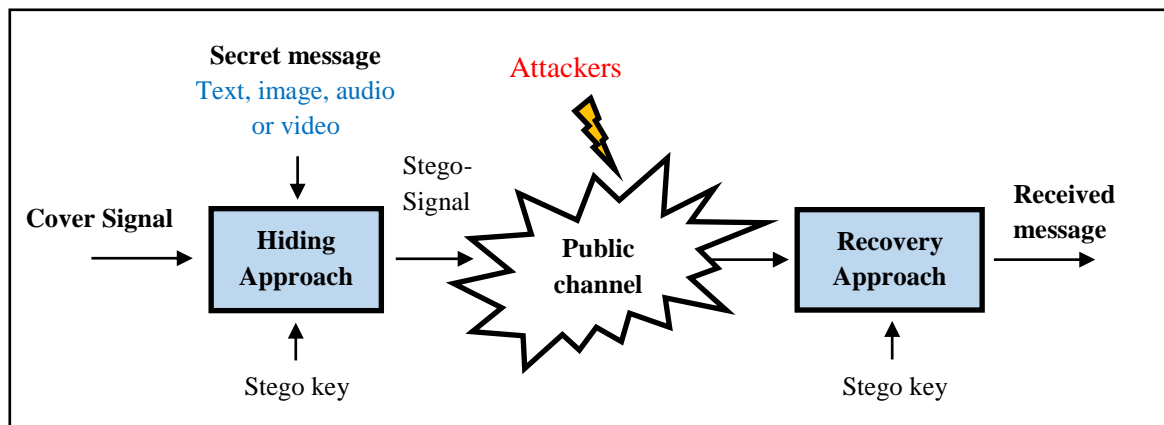


Figure 1.1 VoIP Steganography Model.

1.3 Data Hiding Applications

Steganography and watermarking become widely used in different field and applications. In particular, steganography is used by those require higher levels of security to exchange information such as military communication systems, intelligence agencies, online transactions, online banking, medical services. Furthermore, steganography also can be employed in watermarking to embed a digital watermark in the host file [13].

One of the watermarking applications is detecting fraud and manipulation of data ownership and thereby determining data authentication. By employing watermarking algorithms, data content protection can be achieved and one can distinguish between the copy and original data, where entering the copy file into the recovery watermarking algorithm does not give the embedded digital watermark which means this file is not original [9].

1.4 Real-Time Steganography versus Offline

Embedding data in dynamic nature over time of communication refers to real-time steganography, while embedding static or previously saved data refers to offline steganography. In real-time steganography, a real-time digital media is used as a carrier for the secret data. A real-time carrier gives the transmitted stego-signal an additional level of security where the eavesdroppers do not have the sufficient time to know if there are an embedded secret data or not [14].

While in offline steganography, the secret data is also embedded into digital media but in offline which means after completing the embedding process, the transmitter can send the stego-data at any time not simultaneously (during the embedding [1]).

The digital carrier may be any multimedia content such as an image, audio, or video. Using audio carrier in steganography offers more challenge comparing to other media. This is due to firstly, the Human Auditory System (HAS) is more sensitive than the Human Visual System (HVS). Secondly, an audio signal data is one dimension, so the embedding capacity is lower than an image or video signal [7].

In General, the real-time communications allow to exchange data between two parties immediately or with imperceptible latency based on several protocols. These protocols are employed to process and control traffic, transmission,

delivery, standardized the communication between the end users and other important protocol tasks[15].

1.5 Review of the Related Work

Several research topics in the field of information hiding have been carried out in literature based on offline and online (real-time) communications. Most of these topics have been focused on improving one or more of the steganographic system requirements. These researches can be classified into two categories: offline steganography approaches and real-time steganography approaches which performing the embedding process in uncompressed cover signal or in compressed cover signal. According to these categories, some of the existing hiding approaches are reviewed in the following sub-sections.

1.5.1 Offline Steganography

Different researches focused on embedding offline secret data within offline cover data, some of these researches are listed as follows:

1. *B. Datta and et. al (2015)* [16]: have been embedded a simple string data into an audio signal based on modulo operator. The target string as secret data is converted into an equivalent hexadecimal where only four bits was taken at a time. This scheme offers high embedding capacity with minimum distortion in the recovered secret data.

2. *H. Shahadi and et al (2016) [1]*: have been introduced a lossless audio steganographic approach based on Int2Int- LWT algorithm. It used random positions (adaptive selection) for embedding the secret message blocks. To maintain the cover signal quality, Weighted Block Matching (WBM) process was employed to compute the similarity between message and cover blocks.
3. *L. Voleti and et al (2021) [17]*: have been employed a static image as a carrier to embed an encrypted secret message (encrypted using Vigenere Cipher). The proposed approach is based on LSB technique as a hiding process. It can be used to protect the integrity of data and copyright.
4. *P. Budumuru and et al (2021) [18]*: have been embedded a digital image as secret data within an audio file using LSB method. To overcome the noise that applied to stego-data, the secret bits are embedded in the third and fourth positions from the LSB direction of the cover audio samples.

The previous reviewed approaches are summarized in Table (2.1).

Table 2.1 A Summary of offline steganography approaches.

<i>Reference</i>	<i>Technique</i>	<i>Strengths</i>	<i>Weaknesses</i>
[16]	Modulo operator, wave file.	<ul style="list-style-type: none"> - High imperceptibility. - Minimum error in the retrieved secret data. - Good robustness. 	<ul style="list-style-type: none"> - Data embedding rate unsuitable to embed a real-time secret speech. - Unsuitable to real-time communications. - The hidden data is a simple text.

[1]	Adaptive embedding positions, Int2Int-LWT, WBM	<ul style="list-style-type: none"> - Enhance security and imperceptibility. - High data embedding rate. 	Unsuitable for real-time communications because it needs scanning during matching process. Thus it has high complexity and high processing time.
[17]	LSB technique, and Vigenere Cipher	<ul style="list-style-type: none"> - Enhance the security level. - Good perceptual quality. - High data embedding rate. 	<ul style="list-style-type: none"> - Unsuitable to real-time communications. - Less robustness to noise.
[18]	LSB technique	<ul style="list-style-type: none"> - Enhance the robustness. - Good imperceptibility. 	<ul style="list-style-type: none"> - Data embedding rate unsuitable to real-time communications.

1.5.2 Embedding before Compression

When the embedding stage is implemented before the compression stage of VoIP, the available hiding space (or embedding rate) in the ordinary cover data will be greater than the available hiding space in the compressed cover data because the size of cover data before the compression is greater than the compressed cover data. With increasing the embedding rate, the system robustness becomes more vulnerable to degrade against data compression provided by VoIP. Therefore, the embedding algorithm must have significant robustness to face the compression process which supplies on the modified cover signal (cover signal with the hidden secret data) in a VoIP channel and thereby protecting the hidden secret data from lost. Also, the quality of the stego data must be acceptable or natural and does not cause any suspicion for eavesdroppers.

1. *M. Shirali-Shahreza and S. Shirali-Shahreza (2010)* [19]: have been adopted on revealing the silence intervals of an audio signal to embed the

secret data. An embedding process is achieved by altering the number of the samples for the revealed silence intervals. The scheme has good robustness against noise and MPEG-1 layer III (MP3) compression, but the embedding rate is low.

2. *R. Roselinkiruba and R. Balakirshnanl (2013) [20]*: have been adopted on revealing the active and inactive frames of an audio signal using Voice Activity Detection (VAD) algorithm. Then, the secret data is encrypted with Parabolic Curve Cryptography (PCC) algorithm and compressed with iLBC Codec to be embedded into the inactive frames of a cover signal.
3. *H. Shahadi and et al (2014) [21]*: have been proposed a lossless hiding scheme based on using two level of integer to integer Haar lifting wavelet transform (Int2Int-Haar LWT) to analyze the cover audio into three sub-bands: first details, second details and smooth sub-band. Then, embedding the encrypted secret message (encrypted with an adaptive key) into the first and second details sub-bands using LSB replacements as an embedding algorithm.
4. *S. El-Khamy and et al (2017) [22]*: have been proposed a simple and lossless audio hiding scheme based on Int2Int LWT. To enhance security and robustness against different attacks, the secret data was encrypted by splitting it into blocks. Then, based on chaotic map of hopping method, each block bits XORed with various random sequence. In the embedding

process, the encrypted secret data is embedded into the 2nd details sub-band.

5. *H. Shahadi (2018) [12]*: has been suggested a covert communication approach based on an adaptation method by creating undetectable secret speech similar to another speech signal by changing its wavelet coefficients. Thus, it is an indirect encryption method does not attract the attention of data attackers and does not required additional bandwidth to embed the secret speech.
6. *S. Bharti and et al (2019) [23]*: have been adopted on splitting the secret signal into an amplitude bits and sign bits, the amplitude bits are embedded into the four LSBs of a cover signal while the sign bits are embedded in the 5th to 8th bits of a cover signal.
7. *P. Kumar and K. Srinivas (2019) [24]*: have been employed spread spectrum representation based audio steganography to embed a secret message encrypted with pseudo-random sequences into the frequency space of real-time cover speech using DWT transformation. This scheme has good performance and imperceptibility with low complexity.

The previous reviewed topics are summarized in Table (2.2).

Table 2.2 A Summary of stego-method that achieved before coding process.

<i>Reference</i>	<i>Technique</i>	<i>Strengths</i>	<i>Weaknesses</i>
[19]	Silence intervals	<ul style="list-style-type: none"> - High robustness to MP3 compression, and LSB attacks. - Low complexity. - Suitable to real-time communications. 	<ul style="list-style-type: none"> - Low data embedding rate. - Secret data in offline. - Less robust to re-sampling attacks. - Unsuitable for VoIP communications.
[20]	Active and inactive frames, VAD, encryption with PCC, and iLBC.	<ul style="list-style-type: none"> - Enhance the security. - Good perceptual quality. - Suitable to real-time communications. 	<ul style="list-style-type: none"> - The hidden data is a simple text. - Low data embedding rate. - Unsuitable for VoIP communications.
[21]	Int2Int-Haar LWT, LSB technique, and PRNG.	<ul style="list-style-type: none"> - Simple, fast, and has sufficient security. - Lossless retrieved secret data. - Low complexity and thereby low processing time. - Less storage capacity - Suitable to real-time communications. 	It is unsuitable for VoIP communications.
[22]	Int2Int-LWT and chaotic map using hopping technique	<ul style="list-style-type: none"> - Lossless retrieved secret data. - Low complexity and thereby low processing time. - Less storage capacity - Improve security and robustness. - Suitable to real-time communications. 	<ul style="list-style-type: none"> - The hidden data is a simple static image. - Unsuitable for VoIP communications.
[12]	Indirect encryption, Wavelet Coefficient Adaption (WCA.), Haar DWT	<ul style="list-style-type: none"> - Minimum transmitted bandwidth - Enhance security and robustness. - Suitable to real-time communications. 	It is unsuitable for VoIP communications.

[23]	Amplitude and sign bits, LSB technique.	<ul style="list-style-type: none"> - High data embedding rate. - Enhance security. - More robust to LSB, and resampling attacks. - Suitable to real-time communications. 	<ul style="list-style-type: none"> - Less robust to additive noise (AWGN). - Unsuitable for VoIP communications.
[24]	Spread spectrum representation, DWT.	<ul style="list-style-type: none"> - Enhance imperceptibility and robustness to additive noise. - Low computational complexity. - Suitable to real-time communications. 	<ul style="list-style-type: none"> - Higher storage capacity. - An error may occur in the retrieved data. - Unsuitable for VoIP communications.

1.5.3 Embedding after Compression

The secret information can be hidden into the compressed cover signal (compress with one type of VoIP standard Codecs). In this procedure, there will be a limitation in the embedding rate or embedding capacity where any change in the compressed data will cause distortion in the decompression process. Therefore, the applications in this field are also limited by hiding simple static image, text, or offline speech. Furthermore, the resulting setgo data must have undoubted quality for any attacker on the network with acceptable robustness against removal or distortion of the embedded secret data.

1. *C. Wang and Q. WuIn (2007) [25]*: have been embedded an compressed secret speech (compress with Speex Codec) into the LSBs of the compressed cover speech (compress with G.711 Codec) using the traditional LSB replacements. The proposed scheme is less robust to statistical analysis, additive noise and did not include any additional level of security.

2. *H. Tian and et al (2009)*: have been suggested two-hybrid approaches based on VoIP. One [26] hides the encrypted secret message (encrypted with m-sequence encryption technique) into the compressed cover speech using the LSBs algorithm. To enhance the security, the RSA algorithm was used to encrypt the shared secret key between the sender and receiver. While the other [27] was proposed the notion of partial similarity value (PSV) for matching the similarity between the LSBs of compressed cover speech and secret message to set the proper value for the threshold PSV.
3. *H. Tian and et al (2011)* [28]: have been introduced a steganographic scheme based on a comprehensive adaptive partial matching as extended work to [26] and [27]. The similarity between the cover and hidden data is adopted by employing two thresholds of PSV and an m sequence. To reduce delay and realize real-time requirements, the encryption is integrated with the embedding process.
4. *Z. Wei and et al (2013)* [29]: have been enhanced the security of the traditional LSB algorithm which is utilized to hide the secret data within VoIP streams by adopting three techniques: value-based multiple insertion (VAMI), voice damage offset (VODO), and voice activity detection dynamic insertion (VADDI). The introduced system has better transparent, but the data embedding rate is low around 102.28bps.

5. *T. Yu and et al (2014) [7]*: have been embedded a secret information encrypted using Advanced Encryption Standard (AES) method and symmetric key within the cover speech coded with pulse code modulation (PCM) Codec. Different hiding locations intervals for LSB insertions are used to obtain variable hiding capacities.
6. *W. Zhijun and et al (2015) [30]*: have been created a secure communication channel within the VoIP network. A Mix-excitation linear prediction (MELP) secret speech of low bitrate was embedded into online compressed cover speech (compress with G.729 Codec) by combining matrix coding and interleaving method.
7. *R. Lin (2016)[4]*: has been proposed a novel embedding algorithm based on revealing the inactive frames of low-bitrate audio streams which are more appropriate for hiding secret data than the active frames. To detect the inactive frames, a developed voice activity detection technique was employed.
8. *Y. Jiang and S. Tang (2017) [31]*: have been divided the encrypted secret data (encrypt with a block cipher) into blocks for randomly embedding each block into VoIP streams using chaotic mapping. The system can protect the integrity of secret data by computing the message digest and sending it to the receiver, also it has sufficient security due to use key distribution. However, the embedding bit rate is low between 0.5 and 8 kbps.

9. *F. Li and et al (2018)* [32]: have been adopted on the characteristic of the speech codec, a novel technique in was proposed by altering the codes of excitation pulse positions of G.723.1 speech Codec to hide the sensitive information. This technique offers good security and efficiency.
10. *J. Peng and et al (2020)* [33]: have been proposed a novel theoretical scheme of real-time covert communications over VoIP. It randomly hides an encrypted secret data (encrypted with AES technique) into a VoIP packets at different data embedding rates and different locations. To ensure secure covert communication, the introduced model is adopted on dynamic key updating to embed the secret data.
11. *F. Li and et al (2020)* [34]: have been adopted on the characteristic features of the common VoIP speech codecs (G.723.1 and G.729A) to embed the secret data. This is achieved by modifying the codes of excitation pulse positions of the above speech codecs. The introduced technique offers good imperceptibility and security than similar techniques in the related literatures. But, the data embedding rate is low about 483bps.

A brief summary for the above reviewing literatures is shown in Table (2.3).

Table 2.3 A Summary of stego-method that achieved after coding process.

<i>Reference</i>	<i>Technique</i>	<i>Strengths</i>	<i>Weaknesses</i>
[25]	It used Speex Codec to compress the secret speech and then embedded into cover speech using LSB technique.	<ul style="list-style-type: none"> - High hiding capacity. - Suitable to VoIP communications. 	<ul style="list-style-type: none"> - Less robustness to noise. - Does not contain any additional level of security.
[26]	LSB substitution, M-sequence strategy, RSA algorithm, G.729a	<ul style="list-style-type: none"> - Provide balancing between security, latency and embedding capacity. - Suitable to VoIP communications. 	<ul style="list-style-type: none"> - Less immunity to additive noise. - Limited hiding capacity.
[27]	LSB substitution, PSV, G.729a.	<ul style="list-style-type: none"> - Provides better performance with an acceptable balance between the hiding capacity and imperceptibility. - Suitable to VoIP communications. 	<ul style="list-style-type: none"> - Less immunity to additive noise. - The hiding capacity is limited.
[28]	An adaptive partial-matching strategy between the message bits and the LSBs of the cover, triple M-sequences, G.729a.	<ul style="list-style-type: none"> - Better balance between bandwidth and imperceptibility. - Suitable to VoIP communications. 	<ul style="list-style-type: none"> - Less immunity to additive noise. - The hiding capacity is limited.
[29]	LSB technique, three approaches VAMI, VODO and VADDI, G.711	<ul style="list-style-type: none"> - Improve security and imperceptibility. - Suitable to VoIP communications. 	Low data embedding rate.
[7]	LSB technique, AES encryption, PCM Codec	<ul style="list-style-type: none"> - Enhance security. - Suitable to VoIP communications. 	<ul style="list-style-type: none"> - Less robust to LSB attacks. - Simple offline secret data.
[30]	Matrix coding, interleaving method, Mix-excitation linear prediction, G.729.	<ul style="list-style-type: none"> - Better data embedding rate. - Low complexity. - Suitable to VoIP communications. 	- Offline secret data.

[4]	Active frames, inactive frames, VAD, G.723.1.	- High data embedding rate. - Suitable to VoIP communications.	Stego-speech quality needs to be modified.
[31]	Chaotic maps, message digest, AES encryption, PCM.	- High resistance to statistical steganalysis. - High security. - Suitable to VoIP communications.	- Low data embedding rate. - Embed simple secret data in offline.
[32]	Pulse code positions, G.723.1.	- Offer better security and imperceptibility. - Suitable to VoIP communications.	Low data embedding rate.
[33]	Dynamic key distribution, AES encryption.	- Enhance cryptographic authentication and security of the proposed. - Suitable to VoIP communications.	- Low data embedding rate. - The VoIP Codec needs to be defined.
[34]	Codes of excitation pulse positions, G.723.1 and G.729A speech codecs.	- Improve imperceptibility and security. - Suitable to VoIP communications.	- Low data embedding rate. -System robustness needs to prove.

In real-time steganography based VoIP, there are few proposed approaches in the literatures; however, they still lack features that are considered as critical requirements for an efficient real-time stenographic system. Some of these issues are listed below:

- Lots of research efforts in VoIP steganography are devoted for hiding simple secret data such as small static image, text, or offline speech signal inside real-time cover speech.
- Data embedding rate unsuitable for hiding real-time secret speech.

- Low levels of security and robustness, where most techniques have used least significant bits (LSBs) replacements which is very sensitive to noise, compression, and secret data detection.
- High computational complexity and high latency which are unsuitable for real-time communication requirements.
- Since VoIP network is a connectionless system, then the retrieved secret packets may be dropped or delayed. Therefore, there is necessity to have packet loss correction (PLC) system. This is not analysed in most proposed approaches.

The aim of the proposed approach is to embed a real-time secret speech within a real-time cover speech with improving the basic steganography requirements as mentioned in the next objective section.

1.6 Problem Statement

There are more challenges to the real-time audio steganography which can be summarized as follows:

- a. The sensitivity of the HAS which can sense very small time-delay periods in audio quality, sometimes less than 30 msec for people with what called golden ears.

- b. The standard approaches of real-time voice communication use data compression during sending voice data over the Internet. This compression removes regions of redundancy and low levels of information which used to embed the secret data. As a result, the embedding rate is extremely decreased to be unsuitable for the required capacity for real-time communications.
- c. Large parts of the hidden data in uncompressed cover signal may be lost after data compression.
- d. At noisy channel, the robustness of the steganographic system should be increased by embedding in higher level energy locations in the cover signal. This leads to degradation in the quality of the stego signal.
- e. The packet losses can be occurred due to noise, routing, or/and latency. These losses lead to decreasing the quality of the stego-signal and retrieved secret data.
- f. There is difficulty in introducing an embedding algorithm with more secure operations to hide and protect the secret information because this lead to raising the processing time to be unsuitable to the real-time communication requirements.

1.7 Objective of the Study

The aim of this work is to propose a real-time covert communication approach based on audio steganography by generating a secure communication

channel within the VoIP network. The Proposed approach is hoped to be suitable for the VoIP requirements with the following specifications:

- a. High data embedding rate that meets real-time communication requirements with high imperceptibility.
- b. Improve the security level through using random permutation for the secret data and stego-key for embedding process.
- c. High robustness against Additive White Gaussian Noise (AWGN) in VoIP channel.
- d. Auto-correction for the lost speech packets.
- e. Low computational complexity that meets VoIP imperceptible latency.

1.8 Contributions

The main contributions of this thesis are:

1. Providing insight into an important issue of embedding real-time secret data within VoIP streams.
2. Compressing the input secret speech by using either fast compression algorithm based on lifting wavelet transform or internet Low Bit Rate (iLBC) Codec as VoIP source Codec with auto-correction for speech quality degradation over a lossy channel.
3. Comparing between the two proposed models based on data embedding rate, perceptual quality, robustness against noise.

4. Testing the robustness of the proposed approach in a noisy channel and for different levels of packet losses.

1.9 Thesis Outlines

The rest of this thesis consists of five chapters including this chapter and as follows:

- Chapter two includes the following topics: a brief background for real-time communications-based VoIP, theoretical background about discrete wavelet transform and lifting wavelet transform, the most commonly used audio steganography techniques, the main steganography requirements with some important performance indicators, pseudo-random number generator, and a literature survey for the related work supported by comparisons and pointing the main strength and weaknesses for each related work.
- Chapter three includes the following topics: the proposed real-time covert communication approach, and the embedding and recovery schemes for the first and second proposed models respectively.
- Chapter four includes the following topics: the experimental results and discussion for the two suggested models based on the basic steganography requirements, a comparison between the two proposed models, and a comparison with the most related literatures.

- Chapter five includes the following topics: the conclusion and the suggestions for future work.

CHAPTER Two

BACKGROUND AND THEORETICAL CONCEPTS

2.1 Introduction

This chapter aims to highlight the fundamental concepts, background, and theory that are required to understand the ideas applied in the proposed real-time covert communication approach. The main discussed subjects in this chapter are VoIP steganography, wavelet transforms and lifting wavelet transform, audio steganography techniques, pseudo-random number generator, and the performance parameters with mathematical formulas that are related to the proposed work. Moreover, an intensive literature survey for the published related work is also presented and compared in this chapter.

2.2 VoIP Communication Model

VoIP is a method of conducting a telephone conversation (audio or video) over data packet network like the internet [8]. Unlike traditional phone lines which based on circuit switching networks to conduct telephone calls; VoIP technology is based on a packet switching strategy to deliver voice communications over the internet. Furthermore, it allows to convert the transmitted voice signal into digital form using Analog to Digital Converter (ADC) and then retrieving the original analog signal using Digital to Analog Converter (DAC).

The digital form can be easily transmitted, compressed, converted, stored, and less affected by noise [35].

There are four basis requirements for VoIP which are signaling, encoding, transport, and gateway control. The signaling phase aims to establish and manage the connection between two end-users [20]. After the conversion is established, the transmitted analog signals must be converted into digital form and then encoded or compressed using one type of VoIP standard codecs. Based on the packet switching strategy, the compressed data are split into packets [8].

When signaling and encoding occur, the voice packets will move using two protocols: Real-time Transport Protocol (RTP) and Real-time Control Protocol (RTCP) [5]. After that, gateway control is conducted to convert the voice packets traffic into IP packets and send them to the receiving end. At the reception end, the received voice packets are combined and decoded to the original form (analog voice signals) [36]. Figure (2.1) shows a general structure for VoIP network [20].

In VoIP, there are several standard audio codecs that used to provide the required transmission bandwidth for any VoIP conversation such as G.711, G.722, G.723.1, G.726, G.728, G.729 and iLBC [35]. The next sub-sections will give more details about G.711 and iLBC because they are used in the proposed work of this thesis.

VoIP communication has become increasingly popular due to being cheap or free-cost, more versatile, and flexible because it is based on internet protocols for delivering data instead of landlines with minimum time delay (latency). Also, it processes voice packets in a separate manner where if one or two packets are dropped during the transmission, it does not affect the entire data [37]. However, the basic drawbacks of VoIP service is totally dependent on the internet connection [8].

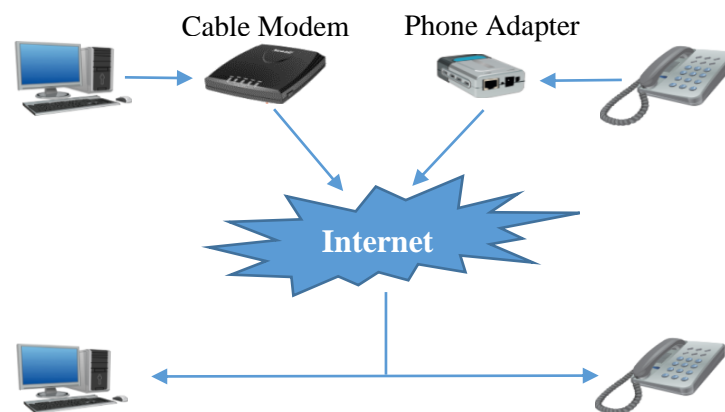


Figure 2.1 VoIP network structure [20].

2.2.1 G.711 Codec

G.711 is one of the International Telecommunication Union-Telecommunication (ITU-T) standards that is used as source coding supported by almost VoIP communications. G.711 Codec is originally designed for use in telephony service. There are two versions of G.711, μ -law (used in America) and A-law (used in Europe and the rest of the world) [38]; each version (or method) based on logarithmic mathematical equations to specify the strength of the input sam-

ples in a companding manner (compressing and expanding). In the implementation, a 16-bit input sample is converted to 8 bits after compression at a compression ratio of 50% for both types which more appropriate to embed the secret data in the proposed approach. However, the coding efficiency in A-law method is better than μ -law [38]. Thus, in this study, an A-law is used as a coding method for the input cover speech.

In A-law compression, an analog input signal is normalized to obtain maximum input sample strength of ± 1 . The A-law encoding (or compressing) equation can be expressed as follows [39]:

$$Y = \begin{cases} \operatorname{sgn}(X) \frac{A^{|X|}}{1+\ln(A)}, & 0 \leq |X| < \frac{1}{A} \\ \operatorname{sgn}(X) \frac{1+\ln(A|X|)}{1+\ln(A)}, & \frac{1}{A} \leq |X| \leq 1 \end{cases} \quad \dots (2.1)$$

Where X and Y are the input and output signals respectively, $\operatorname{sgn}(x)$ is the sign function, and A is the compression parameter which is equal to 86.5.

While the A-law decoding (or expansion) equation [39] is

$$X = \begin{cases} \operatorname{sgn}(Y) \frac{|Y|(1+\ln(A))}{A}, & 0 \leq |Y| < \frac{1}{1+\ln(A)} \\ \operatorname{sgn}(Y) \frac{\exp(|Y|(1+\ln(A))-1)}{A}, & \frac{1}{1+\ln(A)} \leq |Y| \leq 1 \end{cases} \quad \dots (2.2)$$

2.2.2 Internet Low Bit Rate Codec (iLBC)

iLBC is a VoIP source coding used to compress the input speech signal with simple and fast processes suitable for real-time communication requirements [20]. For speech quality degradation across a lossy channel, iLBC Codec has

auto-correction for packet losses. It can be implemented with transmission rates lower than 13.33 kbps or 15.20 kbps for an encoding frame period of 30 msec or 20 msec respectively [20].

The input rate of iLBC is fixed and equal to 128 kbps where the input signal must be a pulse code modulation of speech signal sampled at 8000Hz with a 16 bits per sample resolution. However, the output from this Codec forms 10.4% (for 13.33 kbps mode) or 11.9% (for 15.20 kbps mode) from the original data rate of the input signal [40]. For more details about iLBC Codec see appendix A.

2.3 Wavelet Transform (WT)

Practically, most signals are represented in time-amplitude representation (time domain) but this domain does not give sufficient information about the various frequencies that exist in the signal [41]. Another popular analytical tool is called Fourier transform which can be used to provide spectral analysis or frequency components of the signal that require in some applications [42].

Since the changes in the signal spectrum according to time does not available in this domain, a modified technique known as Short-Time Fourier Transform (STFT) is employed to provide time-frequency representation for the signal as shown in Figure (2.2a). However, non-periodic signals such as image, speech, and video signals are difficultly analyzed using above transformations [10].

Therefore, an alternative mathematical method called wavelet transform used to transform the signal from time domain to wavelet domain as a time-frequency representation by dividing the signal into various frequency components. Each frequency component is analyzed with a resolution suitable to its scale [43] as shown in Figure (2.2b); unlike STFT in which all frequencies have uniform time resolutions.

General structure of wavelet transforms consists of two processing blocks: analysis block and synthesis block. In analysis block, the signal is analyzed into the wavelet coefficients at a process called decomposition process. While the synthesis block acts as the inverse of the decomposition process to reconstruct a signal close to the original signal. In the two blocks, filters are employed [41].

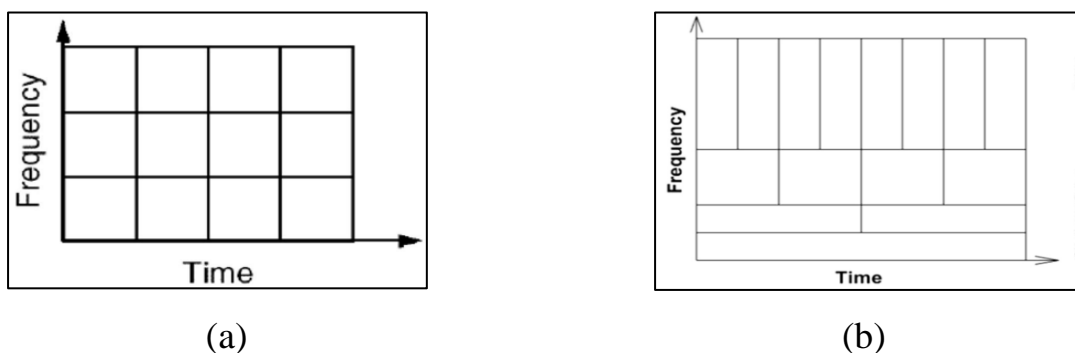


Figure 2.2 Time frequency structure (a) STFT and (b) WT.

Basically, there are several types for wavelet transforms such as Discrete Wavelet Transform (DWT), Lifting Wavelet Transform (LWT), Packet Wavelet Transform (PWT), Complex Wavelet Transform (CWT), and other types. This study focuses on the DWT and LWT that used in the proposed approach where the LWT is an implementation to DWT. Also, the LWT is implemented

with haar filter which based on integer values only. The DWT and LWT are illustrated in the following sub-sections:

2.3.1 Discrete Wavelet Transform (DWT)

The DWT is considered as developed and efficient mathematical method to analysis the non-stationary (non-periodic) signals such as speech. It can be implemented by using discrete filter banks. The main idea of using filter bank is to separate the discrete input signal into two or more sub-signals or to compose two or more signals into one signal [44]. In the first level of DWT, two channel filter bank is employed. The structure of this filter is shown in Figure (2.3) [45]. The input signal is filtered using a filter bank $G_0(z)$ and $G_1(z)$. The outputs are decimated (down-sampled) by a factor of two, and then passed on a processing unit.

The processing unit stores the signal into the memory or transfers it through the channel. To reconstruct the original signal, the modified signals are interpolated (up-sampled) by factor of two and passed through another filter bank $H_0(z)$ and $H_1(z)$. The reconstructed signal is obtained by summing the outputs from $H_0(z)$ and $H_1(z)$. In ideal reconstruction, the reconstructed signal is identical to the input signal. Basically, $G_0(z)$ and $G_1(z)$ called an analysis filter bank while $H_0(z)$ and $H_1(z)$ called a synthesis filter bank [46].

Generally, $G_0(z)$ and $H_0(z)$ are low-pass filters while $G_1(z)$ and $H_1(z)$ are high-pass filters. The purpose of low pass filter is to extract low frequency

components (approximation coefficients) of the input signal after decimation while the high pass filters extract high frequency components (details coefficients). Most input signal power is concentrated at approximation coefficients with little power diffuse over details coefficients [47].

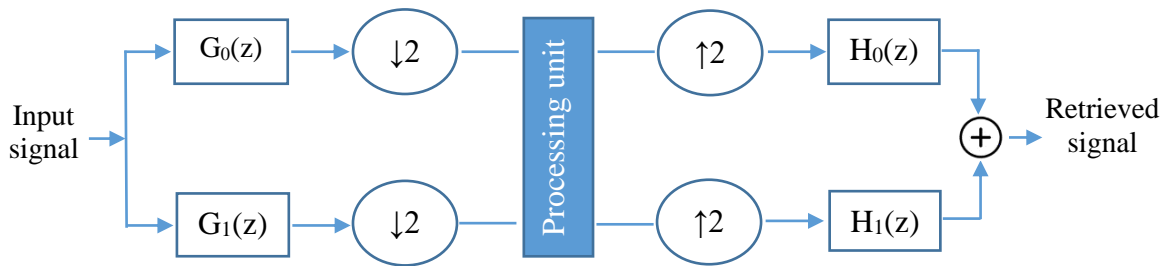


Figure 2.3 Two channel filter bank.

The decomposition process can be continued for multi-levels using cascade of analysis filter banks, where the procedure of first decomposition level DWT is repeated by analyzing either the approximation signal or details signal or both such as in Packet Wavelet Transform (PWT) [46]. Extensively, low frequency components are decomposed for further levels. Figure (2.4) illustrates the decomposition process of three level DWT [45].

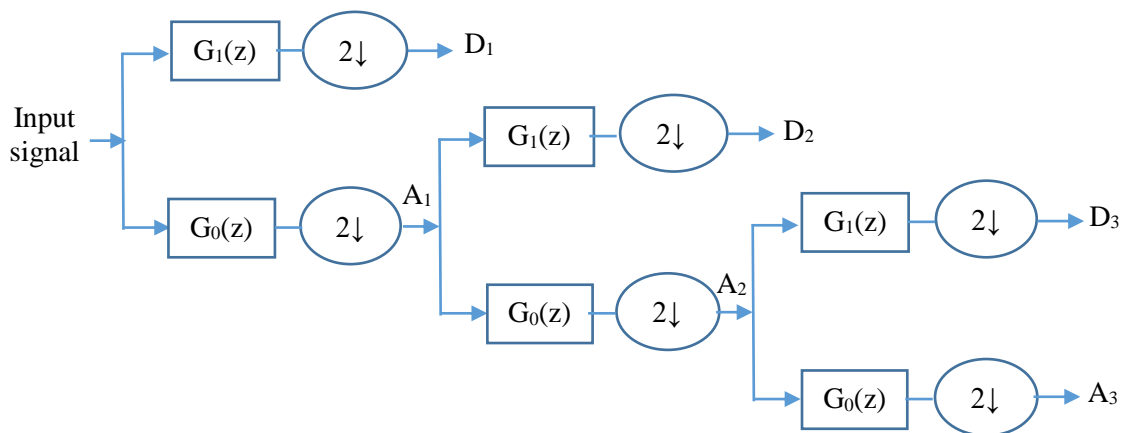


Figure 2.4 Three level DWT decomposition.

The DWT is invertible which means that the inverse of DWT can reconstruct the original input signal. To implement the reconstruction process (inverse DWT), filter banks are also used to implement the inverse of DWT. Figure (2.5) illustrates three levels of the inverse DWT using cascade of synthesis filter banks [45].

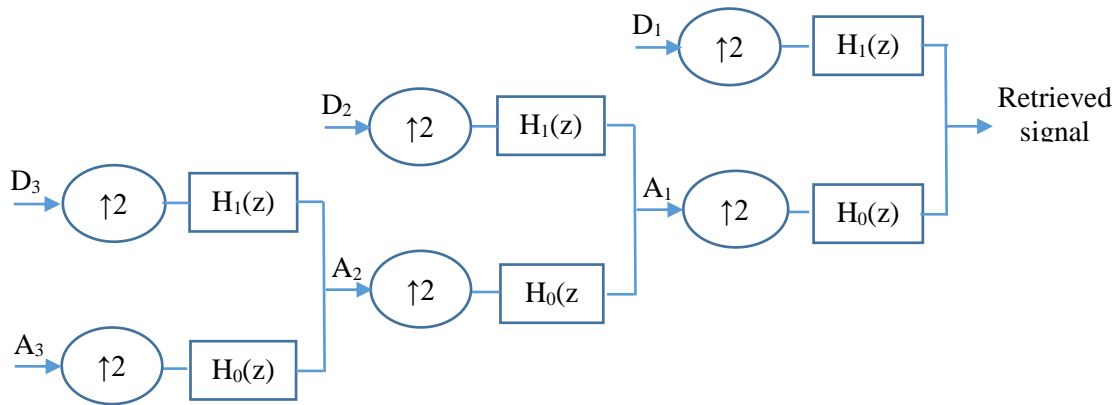


Figure 2.5 Three level inverse DWT reconstruction.

2.3.2 Lifting Wavelet Transform (LWT)

The implementation of DWT using traditional convolution requires memory and computational complexity which can be significantly reduced with obtaining the same results based on lifting transformations. Sweldens (in 1996) introduced a lifting scheme (lifting wavelet transform) to implement the DWT [48].

LWT provides a very fast transform with real number outputs that can be in the same bit resolution of the input. Two signals are produced; one of them is called approximation (smooth coefficients) that is much closed to the original signal with lower resolution, and the other is called detail coefficients that comprise the high-frequency components of the input signal [45].

There are three steps to realize LWT as demonstrated in Figure (2.6a): split, prediction and, update. In the first step, the input signal X is split into a set of even and odd indexed samples $X_n(e)$ and $X_n(o)$. In the prediction (P) step, the odd samples are predicted from even samples to obtain the details coefficients (high frequency components). In the update (U) step, the even samples are updated based on the odd samples to get the approximation coefficients (smooth or low frequency components) [49].

The previous three decomposition steps are defined in Eq. (2.3), (2.4), and (2.5) respectively [50] for one mode of LWT which is called integer to integer. This mode of LWT reduces the required memory of LWT by avoiding floating point results that are normally required higher memory for presentation because higher bit resolution [21].

$$\text{Split//} \quad X_n: X_n(e), X_n(o) \quad \dots \quad (2.3)$$

$$\text{Prediction//} \quad D_{n-1} = X_n(o) - \text{floor} (P[X_n(e)]) \quad \dots \quad (2.4)$$

$$\text{Update//} \quad A_{n-1} = X_n(e) + \text{floor} (U[D_{n-1}]) \quad \dots \quad (2.5)$$

Floor(x) function gives the largest integer number less than or equal to x. This routing function is employed with the integer to integer mode of LWT and it is not employed with floating number LWT. However, the inverse LWT is easy to obtain as shown in Figure (2.6b) by reversing the previous three steps:

undo update, undo prediction and combining odd samples with even samples as given in Eq. (2.6), (2.7), and (2.8) respectively [50].

$$\text{Inverse update//} \quad X_n(e) = A_{n-1} - \text{floor}(U[D_{n-1}]) \quad \dots \quad (2.6)$$

$$\text{Inverse prediction//} \quad X_n(o) = D_{n-1} + \text{floor}(P[X_n(e)]) \quad \dots \quad (2.7)$$

$$\text{Combing//} \quad X_n = \text{Combine}[X_n(e), X_n(o)] \quad \dots \quad (2.8)$$

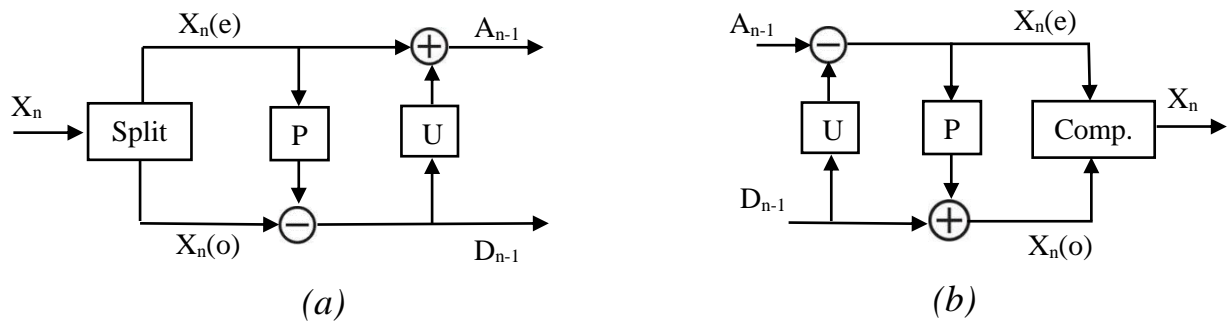


Figure 2.6 One level LWT, (a) decomposition and (b) reconstruction [48].

The LWT is superior to the traditional implementation of DWT due to the following metrics:

- i. It splits the input signal samples into odd and even indexed samples before filtering it, unlike DWT which filters the entire input signal. Thus, instead of processing the whole input data, LWT will process half of it. This leads to decreases in the processing time to half.
- ii. Lossless transformations because it based on integer values.
- iii. It is useful in applications that require lossless data compression [1].

- iv. It has less computational complexity, faster, and can be performed on the hardware (either Field Programmable Gate Array FPGA or Digital Signal Processing DSP chips) with fewer resources.
- v. There is no need for any additional storage capacity, compared to the traditional DWT which requires a higher memory size to store its floating coefficients [21].

2.3.2.1 Integer to Integer- Haar LWT

Haar is a filter bank that used in the implementation of DWT. Two signals are produced from it; one is called approximation that is much closed to the input signal with lower resolution, and the other is called detail coefficients that contain the high-frequency components [45]. The approximation coefficients are obtained by calculating the average of the adjacent samples of the discrete input signal. While the details coefficients are obtained by calculating the difference between each adjacent input samples [50].

In the integer to integer lifting scheme based Haar filter (Int2Int- Haar LWT), the samples of the input signal have only integer values. Thus $P[X_n(e)] = X_n(e)$ and $\text{floor}[X_n(e)] = X_n(e)$. Then, the odd samples are easily predicted to get the detail coefficients as defined in Eq. (2.9) [48].

$$D_{n-1} = X_n(o) - X_n(e) \quad \dots \quad (2.9)$$

Based on D_{n-1} value, the update stage output can be computed. As mentioned earlier, the approximation coefficients' Haar filter are obtained by computing the average of the adjacent samples. Thus,

$$A_{n-1} = \text{floor} [(X_n(e) + X_n(o))/2] \quad \dots (2.10)$$

$$A_{n-1} = \text{floor} [(X_n(e) + X_n(e) + D_{n-1})/2] \quad \dots (2.11)$$

Therefore, the final equation of the update stage can be written as in Eq. (2.12) [48]:

$$A_{n-1} = X_n(e) + \text{floor} [D_{n-1}/2] \quad \dots (2.12)$$

By employing the approximation and details coefficients, the original input signal can be reconstructed. This is achieved with undo-update, undo-prediction as in Eq. (2.13) and (2.14) respectively, then computing the outputs from the two stages together as in Eq. (2.8) [50].

$$X_n(e) = A_{n-1} - \text{floor} [D_{n-1}/2] \quad \dots (2.13)$$

$$X_n(o) = D_{n-1} + X_n(e) \quad \dots (2.14)$$

2.4 Audio Steganography Techniques

In audio steganography, the secret data is embedded into an audio carrier in such way that makes the modifications in the cover file are imperceptible for the human ear. Several techniques are used for audio steganography which can be categorized with respect to the embedding domain into temporal, frequency,

and wavelet domain [51]. In the temporal or time domain, the cover file is directly modified to embed the confidential data. Therefore, the methods in this domain are simple to implement, fast, and can be implemented in real-time; but they have less robustness to noise. While both frequency and wavelet domains transform the audio file (cover) into a set of coefficients to be used as carrier for the secret data. However, audio steganography techniques are accessible to everyone, the researcher must choose a technique that achieved secure communication based on data embedding rate, robustness, security, and other important factors for successful audio steganography. Some of the audio data hiding techniques are offered in the following sub-sections.

2.4.1 Least Significant Bit (LSB) technique

The LSB technique is one of the simplest and most commonly used methods in the temporal domain. Many researchers have adopted this method to embed the secret data in the voice payload packets (VoIP packets). However, the aim of the LSB technique is to transmit the secret information to the receiver without knowing if there is embedded data or not. Traditionally, each bit from the secret message signal is replaced with the LSB of each cover sample as shown in Figure (2.7) [11].

LSB technique is distinguished by high data hiding rate, simplicity to implement, low complexity, and can be used by many techniques. The drawbacks are summarized as it extremely sensitive to additive noise and having weak

resistance to many types of attacks such as scaling, filtering, and lossy compression where all of these processes will lead to destroy the secret data. Furthermore, data attackers can easily discover the embedding data by extracting the entire LSBs of the stego file [52].

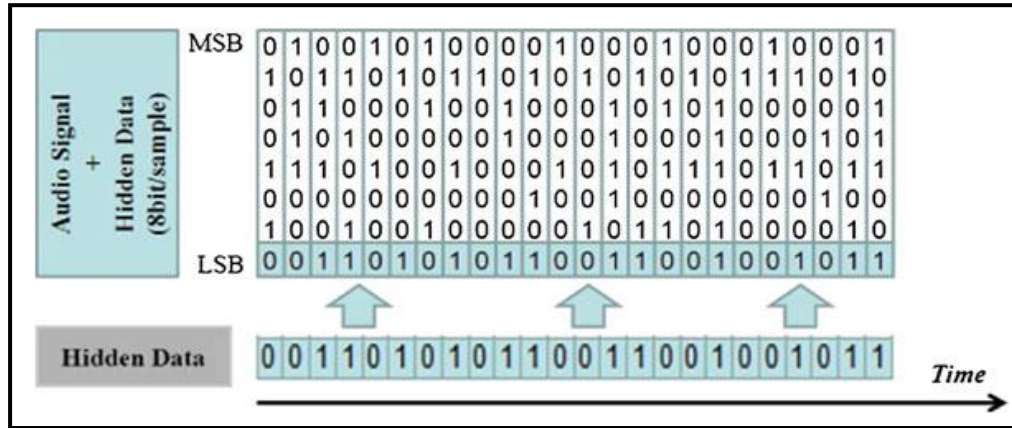


Figure 2.7 LSB mechanism [11].

2.4.2 Phase Coding

Phase coding is another method used to embed secret data in audio hosts. It is based on the fact that the HAS is less sensitive to changes in the phase of various spectral components than noise. To perform this process, the original audio signal is segmented into smaller segments each has a length equal to the length of the secret data. Then, Discrete Fourier Transform (DFT) is applied on each segment to obtain a matrix of phases. In the hiding stage, the phase vector of initial segment is used to embed the secret data according to Eq. (2.15) [15].

$$\text{New phase} = \begin{cases} \text{old phase} + \frac{\pi}{2} & \text{if secret bit} = 0 \\ \text{old phase} - \frac{\pi}{2} & \text{if secret bit} = 1 \end{cases} \quad \dots (2.15)$$

By using the new phase of the initial segment with the original phase matrix, a new phase matrix is generated. The inverse DFT is applied on the new phase matrix and then sequentially combining the resulted audio segments to obtain the output stego file. The drawback of this method is low data embedding rate due to use only the first segment in the hiding process, which means the secret data does not spread over the entire cover audio signal and thus the cropping attack can easily eliminate the hidden data [53].

2.4.3 Parity Coding

This method is operated by dividing the audio signal into separate groups of samples instead of individual samples. Then, the parity bit for each group is obtained; and the secret data is transformed into binary bits. For embedding the secret bit one by one, the parity bit of each group is checked. If there is matching between the secret bit and the parity bit do nothing. Otherwise, change the LSB of one sample in the group to make it equal to the secret bit. So, this will allow to hide the secret bit at a wide range of options and will maintain the modification in the signal to be unobservable [54].

2.4.4 Echo Hiding

This method hides the confident data into the audio signal by introducing an echo to the original signal. This based on three echo parameters to hide the secret data which are amplitude, decay rate and delay time from the host signal. In order to maintain the echo imperceptible, all of those parameters must set at

values not perceptible to human ear [51]. However, a convolution between the echo kernel (as defined in Eq. (2.16)) and the audio signal $S_{inp}(n)$ is employed to obtain the echoed audio signal $x(n)$ as given in Eq. (2.17) [3].

$$h(n) = \delta(n) + \alpha\delta(n - d_i) \quad \dots (2.16)$$

$$S_{out}(n) = S_{inp}(n) + \alpha S_{inp}(n - d_i) \quad \dots (2.17)$$

$i = 0$ or 1 .

d_0 and d_1 : Delay for bit 0 and bit 1 respectively and α : Echo amplitude

Embedding of secret data is achieved by separating the cover audio signal into number of segments equal to the number of secret data bits. Then, echoing each segment according to Eq. (2.17) with a delay equal to the secret bit to be embedded. This method is characterized by its simplicity, ease of implementation and it produces minimal noise [3].

2.4.5 Spread Spectrum (SS)

Spread spectrum is another audio steganography technique that spreads the sensitive data through the frequency spectrum of the cover audio signal. This method is analogous to the implementation of the LSB method which spreads the secret data bits over the whole cover audio file. Unlike LSB, the SS method uses a code during the embedding known for both sender and receiver independent on the original signal. The resulting stego signal has a bandwidth greater than what is actually needed for transmission [53][15].

There are two types of the SS method: Direct Sequence Spread Spectrum (DSSS) and Frequency Hopping Spread Spectrum (FHSS). In DSSS, the secret data is multiplied with a pseudorandom signal based on a constant chip rate and then spreading it over the cover signal. This method provides higher embedding capacity. In FHSS, the frequency spectrum of the carrier signal is varied at different times to make it hop from one frequency to another in a rapid way [55].

2.5 Pseudo-Random Number Generator (PRNG)

PRND is a deterministic technique used to produce a sequence of random numbers based on mathematical formulas. It works by feeding an initial state also called feed state to generate random numbers in a short time. These numbers are deterministic because it can be reproduced if the same starting state in the sequence is fed to the algorithm. PRNG used in several applications such as cryptography, public key generation, games, random numbers which are existed in session keys and other places [56].

Generally, there are three types of random generators which are Multiplicative Linear Congruential (MLCG), shift register, and Fibonacci generators. In this study, MLCG is implemented to randomly permute the secret data. MLCG is one of a simplest and oldest techniques that based on the previous integer number to generate the next integer one as written in Eq. (2.18)[57].

$$x_{n+1} = a x_n \text{ mod } m \quad \dots (2.18)$$

m is the selected number ($m > 0$), a is the well-chosen multiplier ($0 < a < m$), and initial state ($0 \leq x_0 < m$)

2.6 Steganography Requirements

For an effective steganographic system, there are numerous important requirements that are used to measure the performance of any embedding algorithm, which are [58][21]: data embedding rate, imperceptibility, security, robustness, and algorithm complexity. These requirements are contradictory to each other, such that the increase in data embedding rate leads to degradation in the robustness of secret data and imperceptibility of stego-file [21]. Therefore, most researcher works tried to create a balancing trade-off between these requirements.

2.6.1 Data Embedding Rate

Data embedding rate or embedding capacity represents the number of the secret bits that can be embedded into a cover file per unit of time or a maximum amount of secret data that the cover file can carry it without influence on the quality, it is measured by bit per second (bps). Also, it can be evaluated as a ratio of the secret data size to the size of cover data [58].

2.6.2 Perceptual Quality

In perceptual quality or imperceptibility, the modification in the cover file after the embedding process must be imperceptible to the human ear. In other words, imperceptibility refers to the amount of changes that the cover can withstand without effect on the stego data quality [59]. To measure this requirement, there are two methods, either by hearing the two signals (cover and stego) from several persons whose have perfect hearing or by using performance parameters as mathematical measurements such as Signal to Noise Ratio (*SNR*) and Perceptual Evaluation of Speech Quality (*PESQ*). *SNR* is a popular indicator in audio steganography that measures the quantity of noise in a stego signal. Moreover, it is used to evaluate the imperceptibility of the steganographic system. It can be computed mathematically using Eq. (2.19)[60][61].

$$SNR = 10 \times \log_{10} \frac{\sum_{i=1}^{N_c} C_i^2}{\sum_{i=1}^{N_c} [\hat{C}_i^2 - C_i^2]} \quad \dots (2.19)$$

Where: C is the original cover signal, \hat{C} is the modified cover signal (stego signal), and N_c is the number of cover samples.

PESQ score is an objective measurement technique used to evaluate the quality of the stego speech signal. It can be computed based on either Digital Speech Level Analyzer (DSL) equipment or auditory evaluation of a group of people who have very good hearing (sometimes they named golden ears) [29]. DSL is an efficient equipment with high accuracy that introduced in United Kingdom by Malden Electronics Ltd [31]. Figure (2.8) shows the internal structure of

DSLAs; while Figure (2.9) illustrates a diagram for the required equipment to measure the quality of the stego speech.

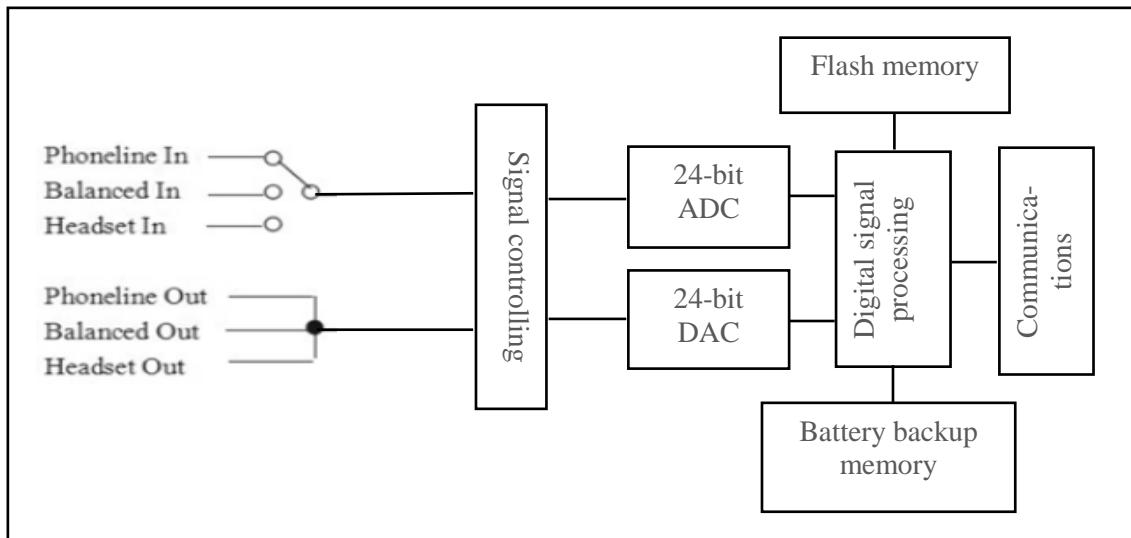


Figure 2.8 Internal components of DSLA [7].

Basically, the recommended value of SNR is 25 dB or more. According to the ITU-T recommendation P.862, the range of $PESQ$ is -0.5 to 4.5 which is considered as 1 instead of -0.5 and 5 instead of 4.5 for more convenience in calculation where 1 indicates the worst sound quality [62].

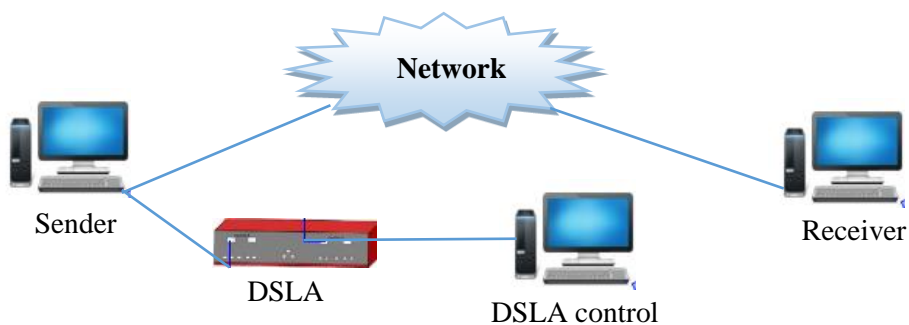


Figure 2.9 Measurements of stego speech quality [7].

2.6.3 Robustness

Robustness refers to the ability of the stego-file to overcome different attacks or the capability of the steganographic system to retrieve the secret data from the stego-file that passed through a noisy channel with a minimum error [58]. The performance parameters use to measure the similarity between the original secret speech and the retrieved secret speech are Normalized Cross-Correlation (NC) as in Eq. (2.20) [21] and Bit Error Rate (BER) as in Eq. (2.21) [55].

$$NC = \frac{\sum_{i=1}^{NS} S_i \times \hat{S}_i}{\sqrt{\sum_{i=1}^{NS} S_i^2} \times \sqrt{\sum_{i=1}^{NS} \hat{S}_i^2}} \quad \dots (2.20)$$

$$BER = \frac{\sum_{i=1}^b \begin{cases} 0, & \hat{S}_i = S_i \\ 1, & \hat{S}_i \neq S_i \end{cases}}{b} \times 100\% \quad \dots (2.21)$$

Where S is the original secret signal, \hat{S} is the retrieved secret signal, N_S is the total number of secret speech samples, and b is the total number of secret speech bits.

$0.85 \leq NC \leq 1$ an acceptable range for NC (the retrieved signal is audible) [13].

There are different types of attacks such as statistical analysis, brute-force, LSB removal, amplification, addition of noise, re-sampling, and compression [23][61]. The steganographic system must have sufficient robustness to face these attacks.

2.6.4 Security

An embedded secret message should be undetectable by any eavesdropper on the network, this refers to the security performance. The complexity of the steganographic algorithm against various attacks or the amount of time that require to destroy the embedding algorithm and extracting the secure data can be used to measure security [63]. Therefore, a secure steganographic scheme indicates statistically undetectable scheme. In general, most stego-algorithms employ simple cryptography with secret or private keys in order to increase the system security [64].

2.6.5 Computational Complexity

The computational complexity can be defined as the amount of computing resources (time and space) that a particular algorithm consumes when it operates. How fast an algorithm will run and how much memory it will use are two mathematical metrics of complexity. Such predictions are crucial for any steganographic system that operates in real-time, as it must have a low level of complexity and, as a result, a low processing time which must be within the human imperceptibility latency [25].

2.7 Chapter Summary

The fundamental concepts and ideas required to understand the proposed real-time covert communication approach are presented in this chapter; which can be summarized as follows:

- a.** In VoIP, the voice signal is digitized, compressed, split into packets, and then transmitted to the destination over the internet.
- b.** Data transmission (voice or any multimedia content) using real-time communications such as the internet protocol networks (VoIP) provides low cost or even free with high reliability.
- c.** Wavelet transform can provide time-frequency representation for non-periodic signals with multi-resolutions.
- d.** Convolution DWT can be implemented with filter banks.
- e.** Lifting scheme is used to implement DWT with simple, fast, and more flexible transformations based on integer to integer LWT mode.
- f.** Haar filter can be used to implement DWT based Integer to integer lifting scheme.
- g.** The basic superiority properties of LWT than DWT are simplicity, fast implementation, and lossless scheme due to utilization integer to integer number which means there are no approximation processes.
- h.** To hide the secret data into audio carriers, several hiding techniques are existed such as LSB, phase coding, parity coding, echo hiding, and spread spectrum.

- i.** One of the simplest audio steganography techniques is the LSB replacements which adopts on spread the secret data over the LSBs of the cover signal.
- j.** In phase coding, the hidden data are replaced with the selected phase components of the cover signal spectrum.
- k.** One of the robust audio hiding techniques is the parity coding; in which the cover audio signal is split into different areas of samples and the secret data is embedded into the parity bit of each sample area. So, the sender has several choices to embed each message bit.
- l.** Based on introducing a short echo signal into an audio carrier, the secret data is hidden in echo hiding technique.
- m.** In spread spectrum, the embedding process is achieved by spreading the secret data over the frequency spectrum of the cover signal.
- n.** Five key features are used to evaluate any steganographic system, which are data embedding rate, imperceptibility, robustness, security and algorithm complexity.

CHAPTER Three

THE PROPOSED MODELS FOR REAL-TIME COVERT COMMUNICATION

3.1 Introduction

This chapter demonstrates the proposed approach for real-time covert communication that suitable for VoIP. Real-time communications have several constraints such as allowable time-latency, limited bandwidth, and sensitivity of the HAS for delay in received signals. These constraints increase the challenges of the proposed system. Therefore, two models are presented in this chapter for the proposed system; the first model is carried out with lifting wavelet transform and the second one is carried out with iLBC Codec for the secret speech compression. G.711 Codec is used to compress the cover speech in both proposed models. The details of the two proposed models are presented in the next sections.

3.2 Real-Time Covert Communication Approach based on VoIP

The proposed approach is based on creating a secure communication channel for the transmitted secret data (in real-time) within a public speech channel such as a VoIP network. As shown in Figure (3.1), the call between two ordinary

persons in a public channel can be considered as a carrier to the secret information exchanged between two important persons.

However, the first important person's speech (represents the secret data) is hidden into the speech of the first ordinary person which is then transmitted via VoIP network as stego-data. At VoIP network, there is no permission for any eavesdropper or even the receiver ordinary person to access the hidden data.

Only the second important person can retrieve the secret speech after using the proper recovery process. Using robust and simple encryption for the secret data, the proposed approach can provide the required protection for the covert communication.

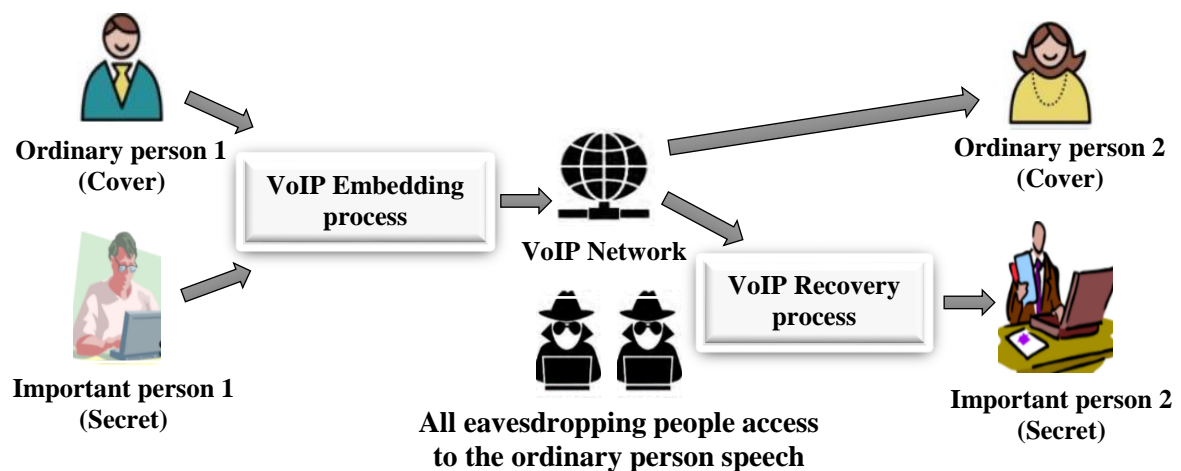


Figure 3.1 The introduced real-time covert communication scheme.

Figures (3.2) and (3.3) show the flowcharts used to implement the embedding scheme and recovery scheme, respectively for the proposed approach.

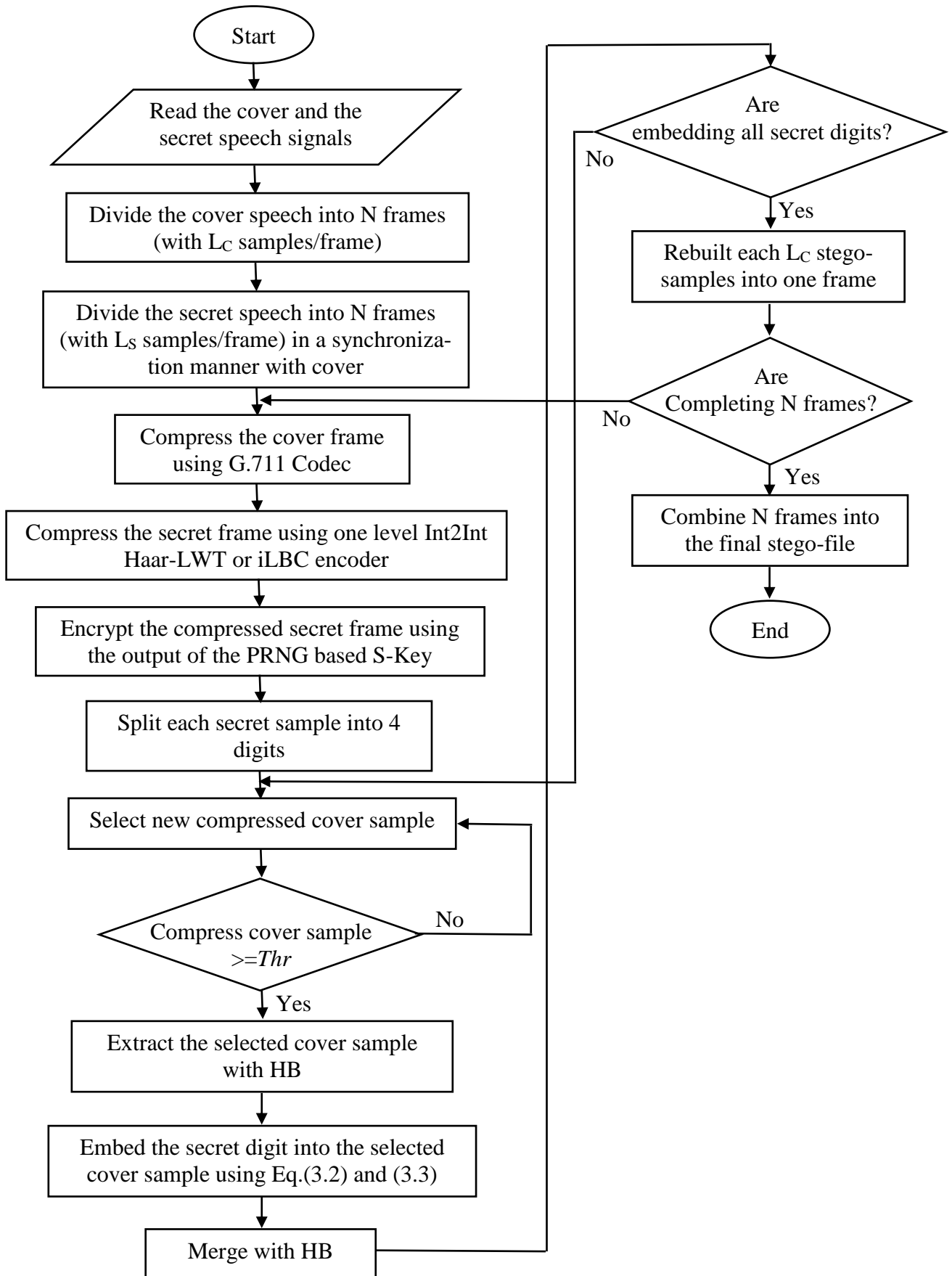


Figure 3.2 Flowchart of the embedding scheme for the proposed approach.

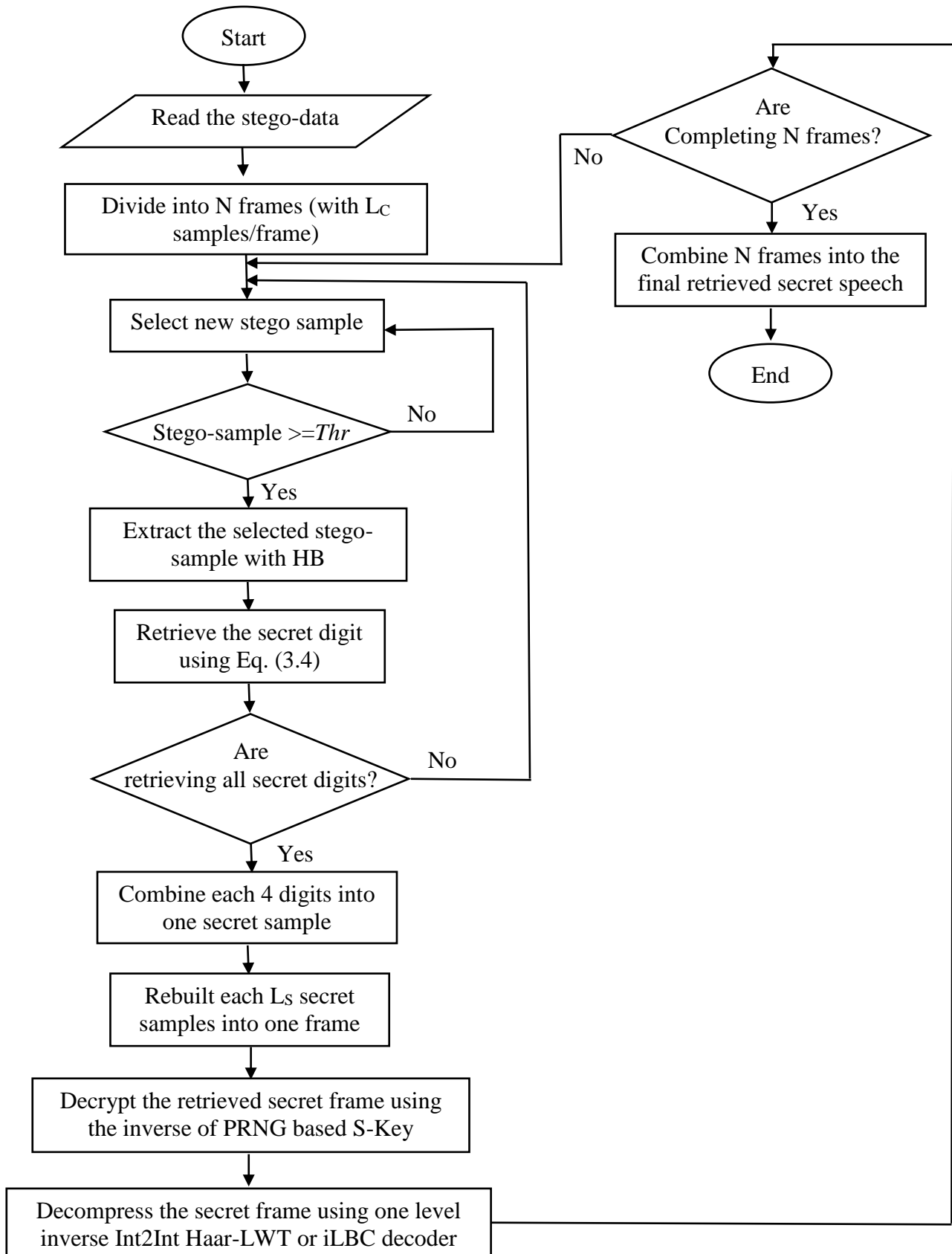


Figure 3.3 Flowchart of the recovery scheme for the proposed approach.

3.3 The Proposed Model based on Int2Int -Haar LWT CODEC

In this model, a real-time input speech signal is compressed by G.711 Codec and used as a carrier for another real-time speech as a secret speech compressed by Int2Int-Haar LWT decomposition. The steps of the embedding and recovery processes with illustrated figures are presented in the following sub-sections:

3.3.1 Embedding Scheme

Each step used in the embedding process are presented in Figure (3.4) and given as follows:

- a. An input cover speech signal (C) sampled at 32 kHz, 44.1 kHz, or more with 16 bits of resolution is divided into N frames. For realizing synchronization between the cover signal and secret signal, an input secret speech signal (S) sampled at 8 kHz or more with a resolution of 8 bits per sample is divided into N frames. Both the input signals are framed without overlap. To maintain the quality of the cover speech after the embedding process, the sampling frequency of the cover speech is chosen as four times (or more) greater than the sampling frequency of the secret speech. The outputs of this step are two frames (C_i and S_i , $i = 1, 2, \dots, N$) in a duration of 16 msec per frame as illustrated in Figure (3.5). The number of samples in the cover frame (L_C) or secret frame (L_S) are obtained by multiplying the sampling frequency of the input speech by 0.016 second.

- b.** G.711 encoder used to compress each cover frame to give (C_{i_comp}) with L_C samples, where each compressed sample has 8 bits $(b_7b_6b_5b_4b_3b_2b_1b_0)$. Also, one level Int2Int-Haar LWT is used to decompose each secret frame into two sub-bands: an approximation and detail as in Eq. (2.9) and (2.10) respectively, where each has $(\frac{L_S}{2})$ samples. Then, calculating the average value of the detail samples which added to the approximation sub-band as one sample. Thus, the compressed secret frame (S_{i_comp}) will contain $(\frac{L_S}{2} + 1)$ samples with a resolution of 8 bits per sample.
- c.** According to PRNG output, the samples of the resulted compressed secret frame (S_{i_comp}) is randomly permuted using Eq. (2.18) to get $(S_{i_comp_P})$. A specific secret key $(S-Key)$ entered by the user as initial state to the PRNG.
- d.** Subsequently, splitting each secret sample $(S_{ij_comp_P})$ into four digits of two bits per digit to get $(S_{ijd_comp_P})$, d is the digit order and may be 1, 2, 3, or 4. Also, L_S is split into 5 digits of two bits per digit and added to the beginning of the first secret frame digits.
- e.** At the embedding process, the cover samples with high energy or strength are selected from (C_{i_comp}) according to a certain threshold (Thr) as input value. This value refers to the strength of the selected cover sample and it may be between 0 and 255. For more clarification:

$Thr = 64$ indicates the position six from the LSB direction in the selected cover sample has a bit 1.

$Thr = 128$ indicates the position seven from the LSB direction in the selected cover sample has a bit 1.

Then, only the candidate samples as defined in Eq. (3.1) will be employed to hide the secret information.

$$C_{ik_{Candidate}} = C_{ij} \text{ if } C_{ij} \geq Thr, \text{ where } k = 1, 2, \dots, M, M \leq L_C \dots (3.1)$$

- f. To determine the positions of embedding in $C_{ik_{Candidate}}$, Hold Bit value (HB) should be chosen between 0 and 3. The increasing in the HB value leads to increasing in the depth of embedding and thereby raising the robustness of the embedded data against adding noise. For more clarification, if HB is 2, then the secret digit will be embedded in the positions of the third and fourth bits from the LSB direction in the candidate sample $C_{ik_{Candidate}}$.

- g. After that, the hold bits are extracted from the selected samples to obtain $C_{ik_{Candidate_HB}}$ which are employed for hiding the secret digits $S_{ijd_comp_P}$. Eq. (3.2) and (3.3) perform the embedding process based LSB technique.

$$R = C_{ik_{Candidate_HB}} \text{ mod } (4) \dots (3.2)$$

$$\hat{C}_{ik_HB} = C_{ik_{Candidate_HB}} - (R - S_{ijd_Comp-P}) \dots (3.3)$$

- h. Sequentially, combining the extracted hold bit/s from each candidate cover sample with the stego-sample to obtain \hat{C}_{ik} . After that, each candidate sample employed in the embedding is replacing with its corresponding stego sample (\hat{C}_{ik}) to give the stego cover frame \hat{C}_i with a length of L_C samples carries the secret frame S_i with a length of $(\frac{L_S}{2} + 1)$ samples.
- i. Finally, combining the N stego-frames into the stego-data (\hat{C}), which represents the compressed cover speech.

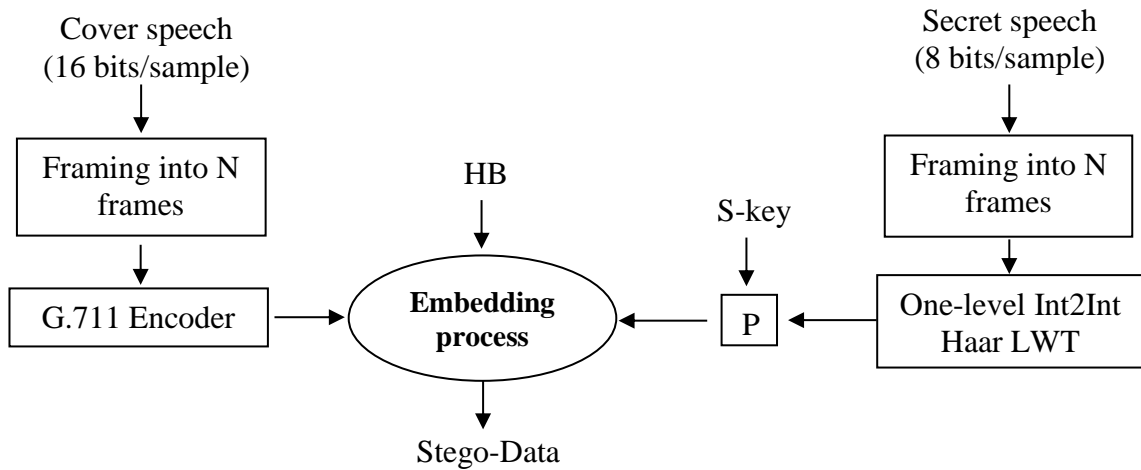


Figure 3.4 The embedding scheme of the first proposed model.

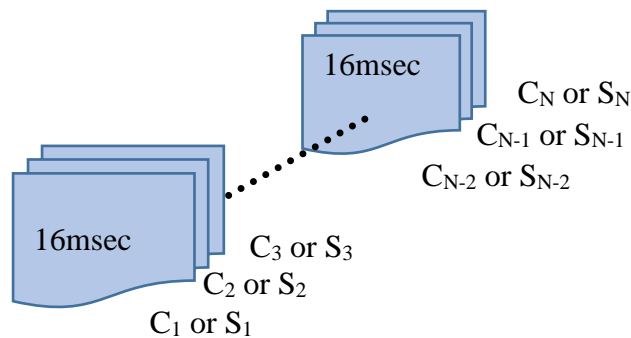


Figure 3.5 Framing process.

3.3.2 Recovery Scheme

The details of each step employed in the recovery process shown in Figure (3.6) are illustrated as follows:

- a. The received stego-data (\hat{C}) is divided without overlap into N frames with a period of 16 msec for each frame.
- b. Selecting high-strength samples greater than or equal to (Thr) as candidate samples, and then extracting the HB from each candidate sample (when HB value greater than zero) to get $\hat{C}_{ik_{HB}}$ (where $k = 1, 2, \dots, M$, $M \leq L_C$).
- c. Retrieving the length of the secret frame L_S from the first five samples in $\hat{C}_{1_{HB}}$ using the following formula:

$$S_{ijd_{comp_P}} = \hat{C}_{ik_{HB}} \bmod (4) \quad \dots (3.4)$$

- d. The retrieving of the secret information is started from $\hat{C}_{16_{HB}}$ until completing $[(\frac{L_S}{2} + 1) \times 4]$ digits using Eq. (3.4).
- e. After that, every 4 secret digits are merged into one sample and each $(\frac{L_S}{2} + 1)$ retrieved secret samples will form one compressed secret frame $S_{i_{comp_P}}$.
- f. The inverse of the permutation is used to decrypt the resulted compressed secret frame into $S_{i_{comp}}$. This is achieved by entering S -Key (which is used by the sender) as an initial state to Eq. (2.18).

- g.** The last sample in S_{i_comp} represents the average value of the original details' samples. Creating $\frac{L_S}{2}$ detail samples with equal values (equal to the retrieved average value). The reminder $\frac{L_S}{2}$ samples from S_{i_comp} will represent the approximation samples.
- h.** The inverse of one level Int2Int-Haar LWT as in Eq. (2.13), (2.14), and (2.8) is used to compose both details and approximation samples into S_i which refers to a retrieved secret frame with a length of L_S samples.
- i.** Finally, each N frames will form the retrieved secret speech.

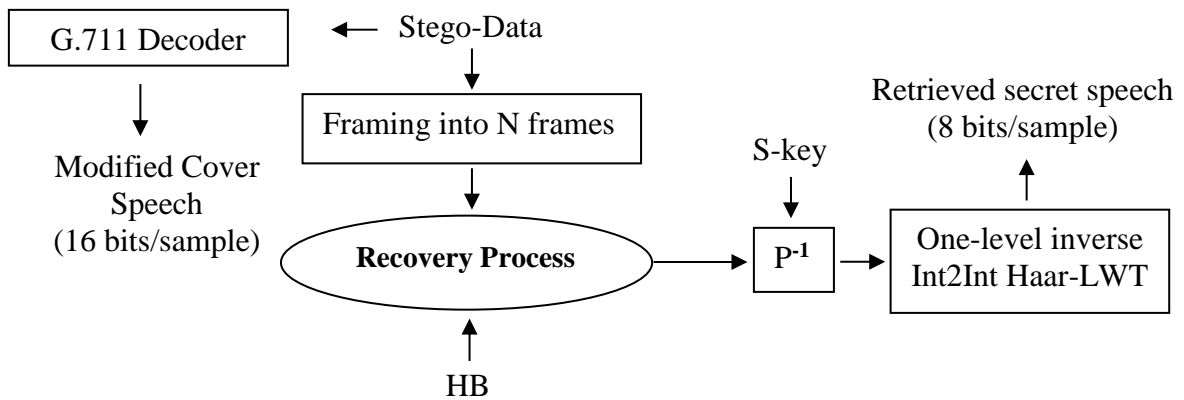


Figure 3.6 The recovery scheme of the first proposed model.

The stego-data is transmitted over public channel and any person can access to it; after applying G.711 decoder on it, he/she can hear a speech signal analogous to the original cover speech without perceptible difference between them

3.4 The Proposed Model based on iLBC CODEC

In this model, a real-time input speech signal is compressed by G.711 Codec and used as a carrier for a 128 kbps real-time secret speech

h signal compressed by iLBC Codec. The steps of the embedding and recovery processes with illustrated figures are presented in the following sub-sections:

3.4.1 Embedding Scheme

The main stages of the embedding process demonstrated in Figure (3.7), are similar to the embedding process of the first proposed model and differ in the following points:

- a. The input secret speech has a resolution of 16 bits per sample due to using iLBC Codec as a compression system for the secret speech.
- b. A period for each cover frame or secret frame is 20 msec because the iLBC operates on two periods: 20 msec or 30 msec. To maintain the audibility within the imperceptible rate, 20 msec have been used.
- c. Compressing the input secret frame ($L_s=160$ samples/frame) using iLBC encoder to 11.9% from its original size with a resolution of 8 bits per sample to get 38 samples/frame.

3.4.2 Recovery Scheme

The stages of the recovery process are demonstrated in Figure (3.8). These steps similar to the recovery process of the first proposed model and differ in the following:

- a. A stego frame period is 20 msec.
- b. The retrieving of the secret information is started from \hat{C}_{16_HB} until completing $[38 \times 4]$ digits using Eq. (3.4).
- c. Each retrieved compressed secret frame S_{i_comp} is decompressed using the iLBC decoder into S_i with a length of L_S samples.

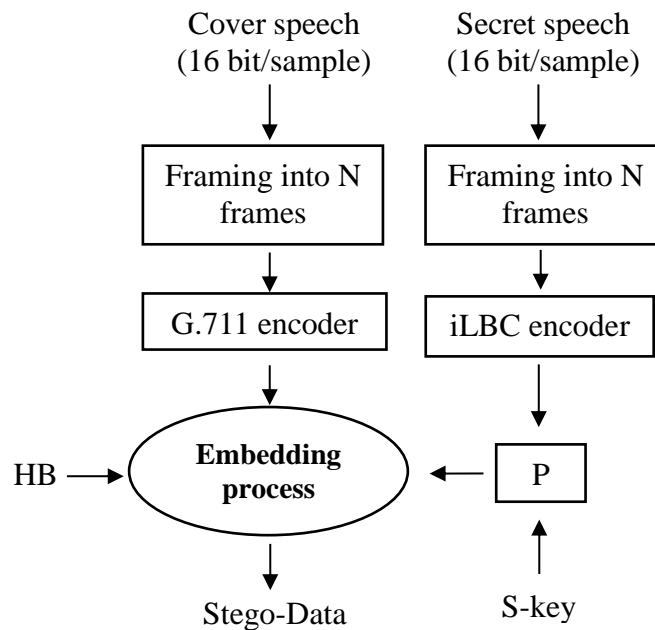


Figure 3.7 The embedding scheme of the second proposed model.

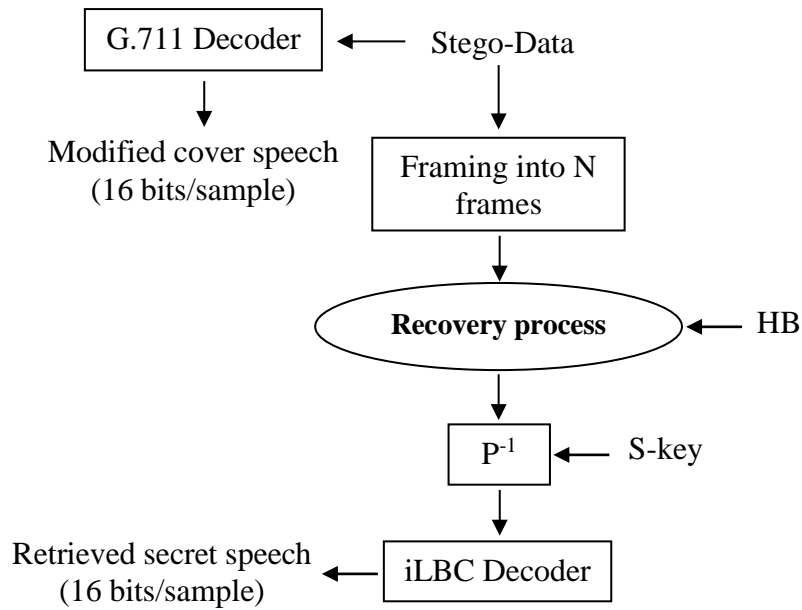


Figure 3.8 The recovery scheme of the second proposed model.

3.5 Chapter Summary

This chapter has proposed a new real-time covert communication approach that can generate a secret channel within a public, standard, and already available channel (VoIP). It exploits the compression provided by VoIP network to embed the secret data. To implement this approach, two models are introduced. The first model encodes a 64 kbps input secret speech using a very simple and fast algorithm called Int2Int-Haar LWT. While the second model adopts iLBC Codec as source Codec to encode a 128 kbps input secret speech signals with auto-correction for speech quality degradation over lossy channel. Both models embed the compressed real-time secret data at a synchronization period into a compressed real-time cover data (compressed with G.711 Codec).

To enhance the security level, pseudo random number generator is utilized as an encryption to the secret data (based on initial value as a secret key) before

the embedding stage and thereby providing the required security for the covert communication. At the embedding process, the secret data are hidden inside the high-strength cover samples. The depth of embedding can be varied in a noisy channel based on *HB* factor. The resulted stego-data which carries the compressed secret data is transmitted over VoIP network to the receiver. The transmitted stego-data can be decoded with G.711decoder to the original cover speech. Also, based on a robust and simple recovery process, the secret data can be fully retrieved from the stego-data in a lossless channel in both models.

CHAPTER Four

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

The main experimental results of the proposed models that described in chapter three are presented in this chapter. The behavior and performance of the proposed models are evaluated by using some assessments tools that are included: *SNR*, *PESQ*, *NC*, and *BER*. Different input speech signals are used as cover and secret speech signals to test the introduced models. These signals have been categorized into two groups: own-recorded speech signals (recorded using the MATLAB audio toolbox) and speeches from the datasets that named "development set, clean speech" [65] and "waves_yesno" [66]. To reproduce a speech signal with a different sampling rate, an online audio converter has been employed. Also, a comparison between the two proposed models and a comparison of the proposed models with other related works are offered.

4.2 Results of the Proposed Model based on Int2Int-Haar LWT CODEC

The performance of the proposed model has been tested using several input speech signals divided into two groups. The first group comprises twenty signals (own-recorded), while the second group comprises forty signals from two datasets: "development set, clean speech" and "waves_yesno". Six signals are

addressed from the first group: three as cover speech signals (Cov_speech1, Cov_speech2, and Cov_speech3) and three as secret speech signals (Sec_speech1, Sec_speech2, and Sec_speech3). From the second group, also six signals are addressed: three as cover signals (Speech1, Speech2, and Speech3) and three as secret signals (Wave1, Wave2, and Wave3). The cover signals are 32 kHz, 16bits/sample resolution, wave format, and mono channel, while the secret signals are 8 kHz sampling frequency, 8bits/sample resolution, wave format, and mono channel.

However, the behavior of the proposed model is illustrated in Figures (4.1) and (4.2) for an input cover and secret frames of a duration of 16 msec per frame. Figure (4.1) shows how the input cover speech is processed from the sender to the receiver. At the sender, the output of the cover speech Codec (G.711 encoder) has compressed cover samples with values between 0 and 255 as seen in Figure (4.1b). These samples are used as carriers for the secret information to produce the stego-data as drawn in Figure (4.1c). There is no obvious difference between the compressed cover data and the stego-data which means that the proposed embedding process doesn't much greatly affect the quality of the cover speech as discussed in the next section. At the receiver end, G.711 decoder is used to decompress the stego-data and hearing the cover speech. There is unperceived degradation from the human ear in the decompressed

cover speech due to the compression and decompression processes which cause some losses in the compressed cover data.

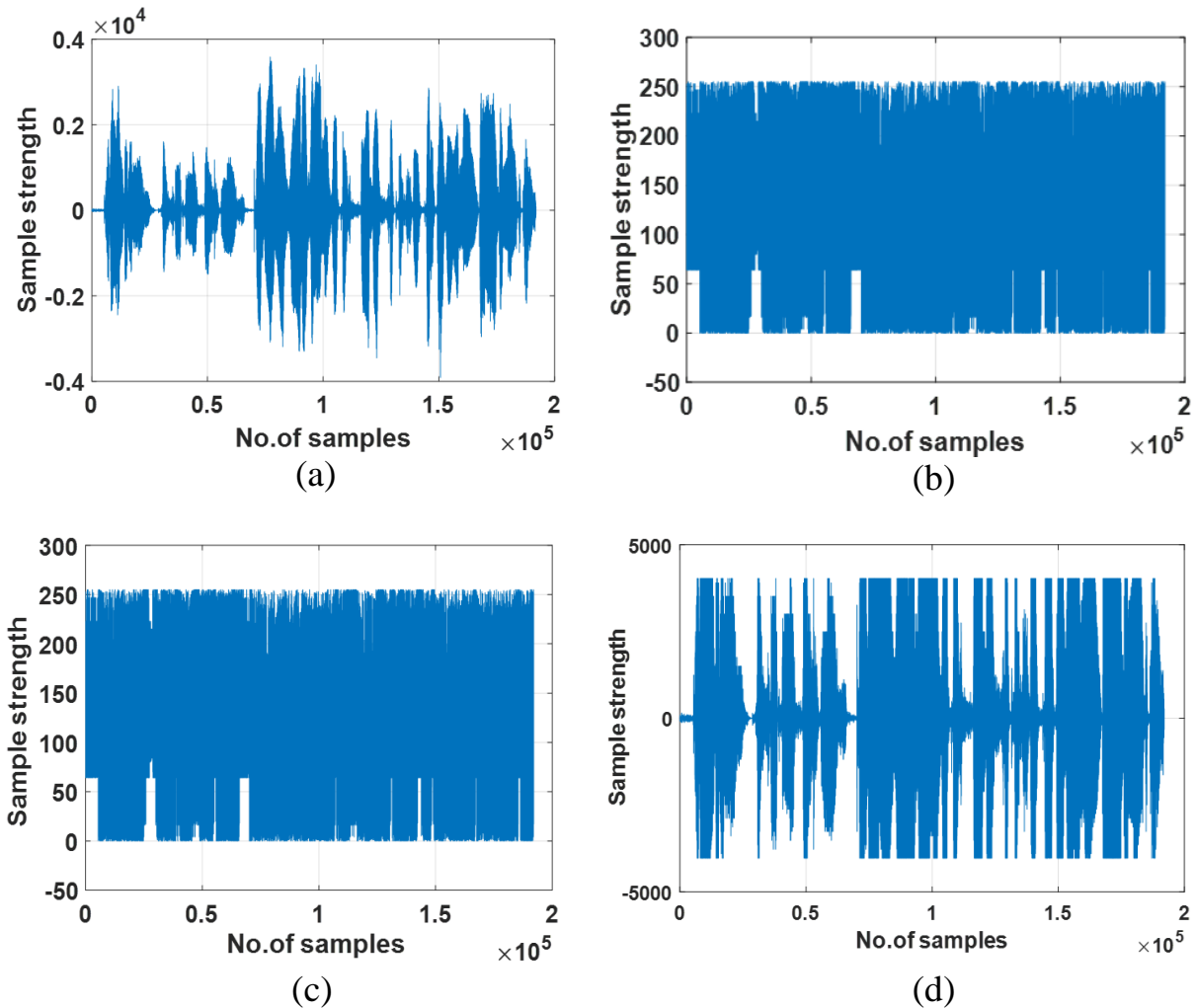


Figure 4.1 The processed cover speech in the first proposed model (a) original, (b) compressed, (c) stego-data, and (c) decompressed cover speech.

Figure (4.2) illustrates the performed processing on the secret speech from the sender to the receiver. At the sender, one level Int2Int-Haar LWT decomposition is applied on each incoming secret frame to reduce the number of secret samples to half as seen in Figure (4.2b). These compressed samples are then embedded in the compressed cover data and sent to the receiver. At the

receiver, the compressed secret data is fully recovered as shown in Figure (4.2c), which is exactly similar to Figure (4.2b). To hear the secret speech, the inverse of one level Int2Int-Haar LWT is applied on the retrieved compressed secret data.

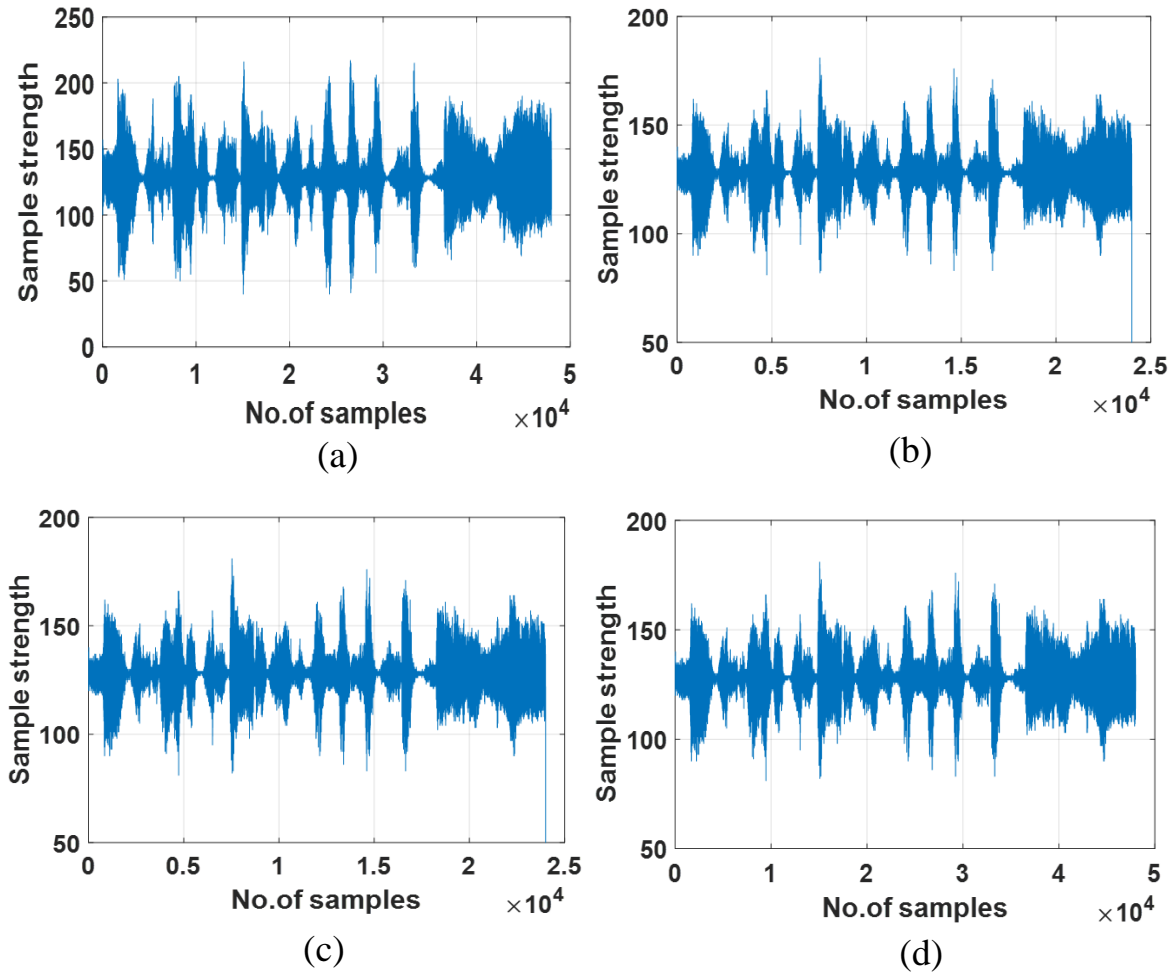


Figure 4.2 The processed secret speech in the first proposed model (a) original, (b) compressed, (c) compressed retrieved, and (d) decompressed retrieved.

4.2.1 Tests of Perceptual Quality and Data Embedding Rate

The perceptual quality of the stego-signals is evaluated based on two quality performance indicators which are *SNR* according to Eq. (2.19) and *PESQ*. The

SNR value is computed between the compressed cover data and the stego-data, while the *PESQ* value is computed between the original cover and the decompressed cover speech signals. However, each secret speech signal is embedded in all of the three cover speech signals as seen in Tables (4.1), (4.2), (4.3), and (4.4) at a lossless noisy channel (*HB* factor=0). The experimental results in these tables show that the quality of the stego-signals is increased with raising the cover sample strength (*Thr*) from 64 to 128. The *PESQ* has very similar values (at *Thr*=64 and 128) for the same cover signal because its operation is similar to the human ear and as a result there is no obvious differences between the two decompressed cover signals.

Table 4.1 Tests the quality of the stego-signals for the first proposed model (the inputs are recorded speeches with a 32kHz cover sampling frequency).

<i>Secret signals</i>	<i>Cover signals</i>	<i>Thr</i>	<i>SNR</i>	<i>PESQ</i>
Sec_speech1	Cov_speech1	64	41.5633	4.1183
		128	41.6392	4.1203
	Cov_speech2	64	41.4469	4.1044
		128	41.5213	4.1050
	Cov_speech3	64	41.9925	4.0989
		128	41.9030	4.0990
Sec_speech2	Cov_speech1	64	41.7442	4.1193
		128	41.8229	4.1195
	Cov_speech2	64	41.6287	4.1045
		128	41.6969	4.1061
	Cov_speech3	64	41.6536	4.0990
		128	41.7825	4.0991
Sec_speech3	Cov_speech1	64	41.5853	4.1186
		128	41.6316	4.1200
	Cov_speech2	64	41.4671	4.1047
		128	41.5343	4.1041
	Cov_speech3	64	41.5281	4.0989
		128	41.6249	4.0998

Table 4.2 Tests the quality of the stego-signals for the first proposed model (the inputs are speeches from dataset with a 32kHz cover sampling frequency).

<i>Secret signals</i>	<i>Cover signals</i>	<i>Thr</i>	<i>SNR</i>	<i>PESQ</i>
Wave1	Speech1	64	41.8339	4.3420
		128	42.2818	4.3571
	Speech2	64	41.7188	4.1494
		128	42.2264	4.1488
	Speech3	64	41.6014	4.0961
		128	42.0264	4.0963
Wave2	Speech1	64	41.5705	4.3444
		128	42.3020	4.3425
	Speech2	64	41.4841	4.1489
		128	42.3070	4.1501
	Speech3	64	41.3435	4.0952
		128	42.0672	4.0956
Wave3	Speech1	64	41.9828	4.3305
		128	42.2446	4.3437
	Speech2	64	41.8787	4.1492
		128	42.1990	4.1494
	Speech3	64	41.7614	4.0956
		128	41.9866	4.0957

The stego-signals have high perceptual quality not less than 41.79 dB and 4.16 as average values for *SNR* and *PESQ* respectively for various input speech signals in the case of the cover speech sampling frequency is 32 KHz. While these values raise to 43.09 dB and 4.22 respectively in the case of raising the sampling frequency for the cover speech to 44.1 KHz as seen in Tables (4.3) and (4.4).

Table 4.3 Tests the quality of the stego-signals for the first proposed model (the inputs are recorded speeches with a 44.1 kHz cover sampling frequency).

<i>Secret signals</i>	<i>Cover signals</i>	<i>Thr</i>	<i>SNR</i>	<i>PESQ</i>
Sec_speech1	Cov_speech1	64	43.0604	4.1949
		128	43.1394	4.1954
	Cov_speech2	64	42.8969	4.2015
		128	42.9154	4.2016
	Cov_speech3	64	43.2894	4.1933
		128	43.3118	4.1963
Sec_speech2	Cov_speech1	64	43.1854	4.1950
		128	43.3301	4.1951
	Cov_speech2	64	43.0645	4.2012
		128	43.1039	4.2017
	Cov_speech3	64	43.4727	4.1966
		128	43.5180	4.1968
Sec_speech3	Cov_speech1	64	43.0625	4.1950
		128	43.1804	4.1952
	Cov_speech2	64	42.9352	4.2014
		128	42.9272	4.2017
	Cov_speech3	64	43.2786	4.1970
		128	43.3538	4.1966

Table 4.4 Tests the quality of the stego-signals for the first proposed model (the inputs are speeches from dataset with a 44.1kHz cover sampling frequency).

<i>Secret signals</i>	<i>Cover signals</i>	<i>Thr</i>	<i>SNR</i>	<i>PESQ</i>
Wave1	Speech1	64	43.0354	4.1909
		128	43.1023	4.1915
	Speech2	64	43.0166	4.2162
		128	43.0239	4.2120
	Speech3	64	43.0190	4.3035
		128	43.1046	4.2987
Wave2	Speech1	64	42.8263	4.1908
		128	42.8581	4.1917
	Speech2	64	42.7357	4.2150
		128	42.7994	4.2126
	Speech3	64	42.7751	4.3032
		128	42.8285	4.2927
Wave3	Speech1	64	43.2293	4.1911
		128	43.2530	4.1915
	Speech2	64	43.1700	4.2155
		128	43.1850	4.2118
	Speech3	64	43.1753	4.2983
		128	43.2514	4.2987

This quality is obtained at a high embedding capacity, which can be computed as a bitrate or ratio as follows:

$$\text{Data embedding rate} = 8000 \frac{\text{sample}}{\text{second}} \times 8 \frac{\text{bit}}{\text{sample}} = 64000 \text{ bps}$$

$$\begin{aligned} \text{Data embedding ratio} &= \frac{\text{hidden data size}}{\text{cover data size}} = \frac{8000 \frac{\text{sample}}{\text{second}} \times 8 \frac{\text{bit}}{\text{sample}}}{32000 \frac{\text{sample}}{\text{second}} \times 16 \frac{\text{bit}}{\text{sample}}} \\ &= \frac{64000 \text{ bit per second}}{512000 \text{ bit per second}} \end{aligned}$$

$$= 12.5 \% \text{ from the size of the cover speech}$$

In the case of a lossy channel, the perceptual quality of the proposed model is decreased at an acceptable range due to increasing the embedding depth or system robustness to avoid losing the embedded information, as will be discussed in the next subsection.

4.2.2 Tests of Robustness against Additive Noise

The degree of resemblance between the original secret speech and the retrieved secret speech has been calculated by utilizing NC and BER , according to Eq. (2.20) and (2.21) respectively. In Tables (4.5), (4.6), (4.7) and (4.8), the robustness test results have been obtained for the embedding stage only and for both embedding and compression stages as overall system in terms of NC and BER at a lossless channel. The obtained results show that the proposed model has good robustness with an error free recovered compressed secret data ($NC=1$

and $BER=0$) after the embedding stage. For the overall system, the allowable decrease in the values of NC and BER is due to the compression and decompression processes on the secret speech.

Table 4.5 Robustness test results for the first proposed model (the inputs are recorded speeches with a 32kHz cover sampling frequency).

Secret signals	Cover signals	Thr	Embedding Stage		Overall system	
			NC	BER	NC	BER%
Sec_speech1	Cov_speech1	64	1	0	0.9925	0.2995
		128	1	0	0.8834	0.4813
	Cov_speech2	64	1	0	0.9925	0.2995
		128	1	0	0.9049	0.5110
	Cov_speech3	64	1	0	0.9925	0.2995
		128	1	0	0.9030	0.5152
Sec_speech2	Cov_speech1	64	1	0	0.9888	0.3506
		128	1	0	0.9091	0.4955
	Cov_speech2	64	1	0	0.9888	0.3506
		128	1	0	0.8850	0.5180
	Cov_speech3	64	1	0	0.9888	0.3506
		128	1	0	0.9092	0.4946
Sec_speech3	Cov_speech1	64	1	0	0.9913	0.3041
		128	1	0	0.8833	0.4802
	Cov_speech2	64	1	0	0.9913	0.3041
		128	1	0	0.8968	0.4982
	Cov_speech3	64	1	0	0.9913	0.3041
		128	1	0	0.8833	0.4803

With increasing the sampling frequency of the cover speech, all the secret data will be embedded in the cover speech even when the strength of the cover samples is raised to 128 as seen in Tables (4.7) and (4.8), where the values of robustness measurements are fixed on one value.

Table 4.6 Robustness test results for the first proposed model (the inputs are speeches from dataset with a 32kHz cover sampling frequency).

Secret signals	Cover signals	Thr	Embedding Stage		Overall system	
			NC	BER	NC	BER
Wave1	Speech1	64	1	0	0.9995	0.2018
		128	1	0	0.8972	0.3723
	Speech2	64	1	0	0.9995	0.2018
		128	1	0	0.8965	0.3731
	Speech3	64	1	0	0.9995	0.2018
		128	1	0	0.8999	0.4368
Wave2	Speech1	64	1	0	0.9996	0.1939
		128	1	0	0.9058	0.4754
	Speech2	64	1	0	0.9996	0.1939
		128	1	0	0.8957	0.3716
	Speech3	64	1	0	0.9996	0.1939
		128	1	0	0.8997	0.3713
Wave3	Speech1	64	1	0	0.9994	0.2098
		128	1	0	0.8958	0.3723
	Speech2	64	1	0	0.9994	0.2098
		128	1	0	0.8903	0.4723
	Speech3	64	1	0	0.9994	0.2098
		128	1	0	0.8951	0.3712

Table 4.7 Robustness test results for the first proposed model (the inputs are recorded speeches with a 44.1kHz cover sampling frequency).

Secret signals	Cover signals	Thr	Embedding Stage		Overall system	
			NC	BER%	NC	BER%
Sec_speech1	Cov_speech1	64	1	0	0.9925	0.2998
		128	1	0		
	Cov_speech2	64	1	0		
		128	1	0		
	Cov_speech3	64	1	0		
		128	1	0		
Sec_speech2	Cov_speech1	64	1	0	0.9888	0.3504
		128	1	0		
	Cov_speech2	64	1	0		
		128	1	0		
	Cov_speech3	64	1	0		
		128	1	0		
Sec_speech3	Cov_speech1	64	1	0	0.9913	0.3045
		128	1	0		
	Cov_speech2	64	1	0		
		128	1	0		
	Cov_speech3	64	1	0		
		128	1	0		

Table 4.8 Robustness test results for the first proposed model (the inputs are speeches from dataset with a 44.1kHz cover sampling frequency).

Secret signals	Cover signals	Thr	Embedding Stage		Overall system	
			NC	BER%	NC	BER%
Wave1	Speech1	64	1	0	0.9995	0.2018
		128	1	0		
	Speech2	64	1	0		
		128	1	0		
	Speech3	64	1	0		
		128	1	0		
Wave2	Speech1	64	1	0	0.9996	0.1939
		128	1	0		
	Speech2	64	1	0		
		128	1	0		
	Speech3	64	1	0		
		128	1	0		
Wave3	Speech1	64	1	0	0.9994	0.2098
		128	1	0		
	Speech2	64	1	0		
		128	1	0		
	Speech3	64	1	0		
		128	1	0		

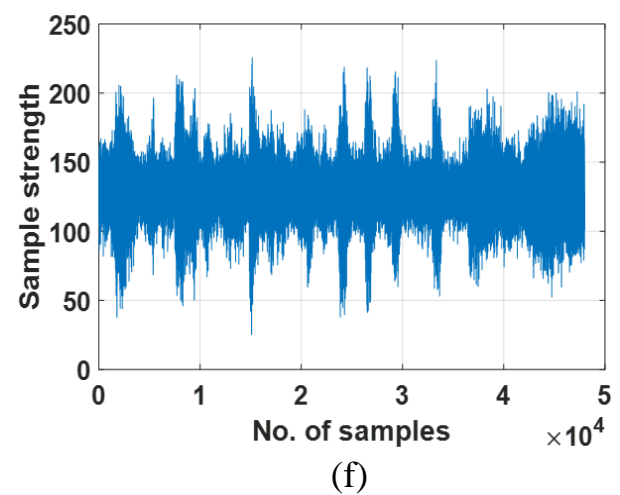
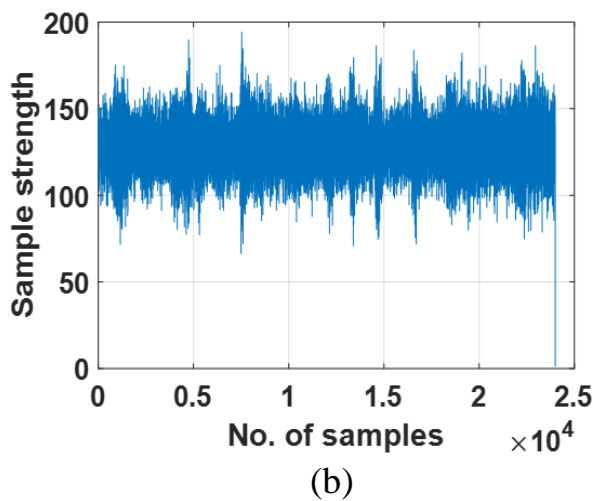
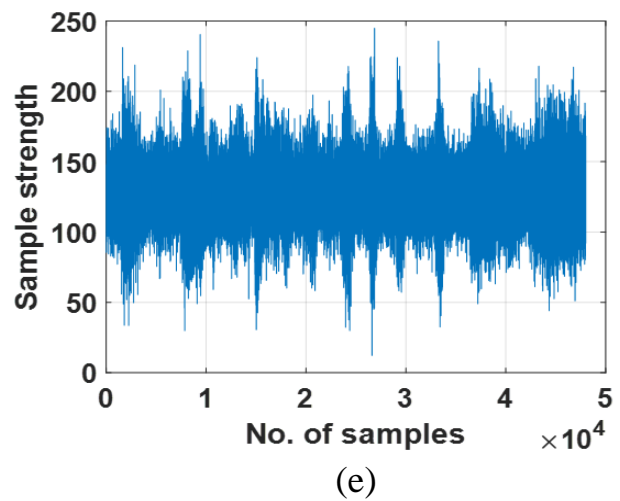
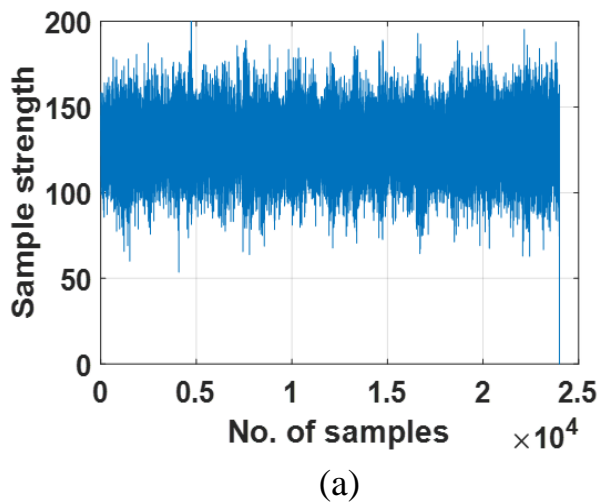
When noise (AWGN) affects the stego data, it will cause a distortion in the hidden data as seen in Table (4.9), where the value of NC is decreased to 0.76. Therefore, the robustness level of the proposed system should be increased. This is accomplished by raising the value of HB which controlled the embedding insertion depth. As a result, the quality of the stego-signals is decreased but it remains within a good or acceptable range.

Table (4.9) demonstrates the robustness tests of the proposed model against AWGN versus HB values, where the immunity of the retrieved compressed secret data and the retrieved decompressed secret speech represented by NC_1 values NC_2 respectively will increase with increasing of HB values. In each

case, the value of NC_2 is less than the value of NC_1 due to the compression and decompression processes. Figure (4.3), illustrates the retrieved compressed and decompressed secret speech signal for different values of HB factor.

Table 4.9 Some tests for the immunity of the secret data against additive noise for the first proposed model.

HB	SNR for stego-data	AWGN in dB							
		50dB		40dB		30dB		20dB	
		NC_1	NC_2	NC_1	NC_2	NC_1	NC_2	NC_1	NC_2
0	41.69	0.975	0.932	0.928	0.852	0.887	0.797	0.847	0.763
1	35.68	0.984	0.956	0.931	0.875	0.890	0.839	0.874	0.787
2	29.68	0.986	0.975	0.954	0.942	0.948	0.905	0.908	0.855
3	26.65	1	0.992	0.980	0.967	0.979	0.950	0.937	0.923



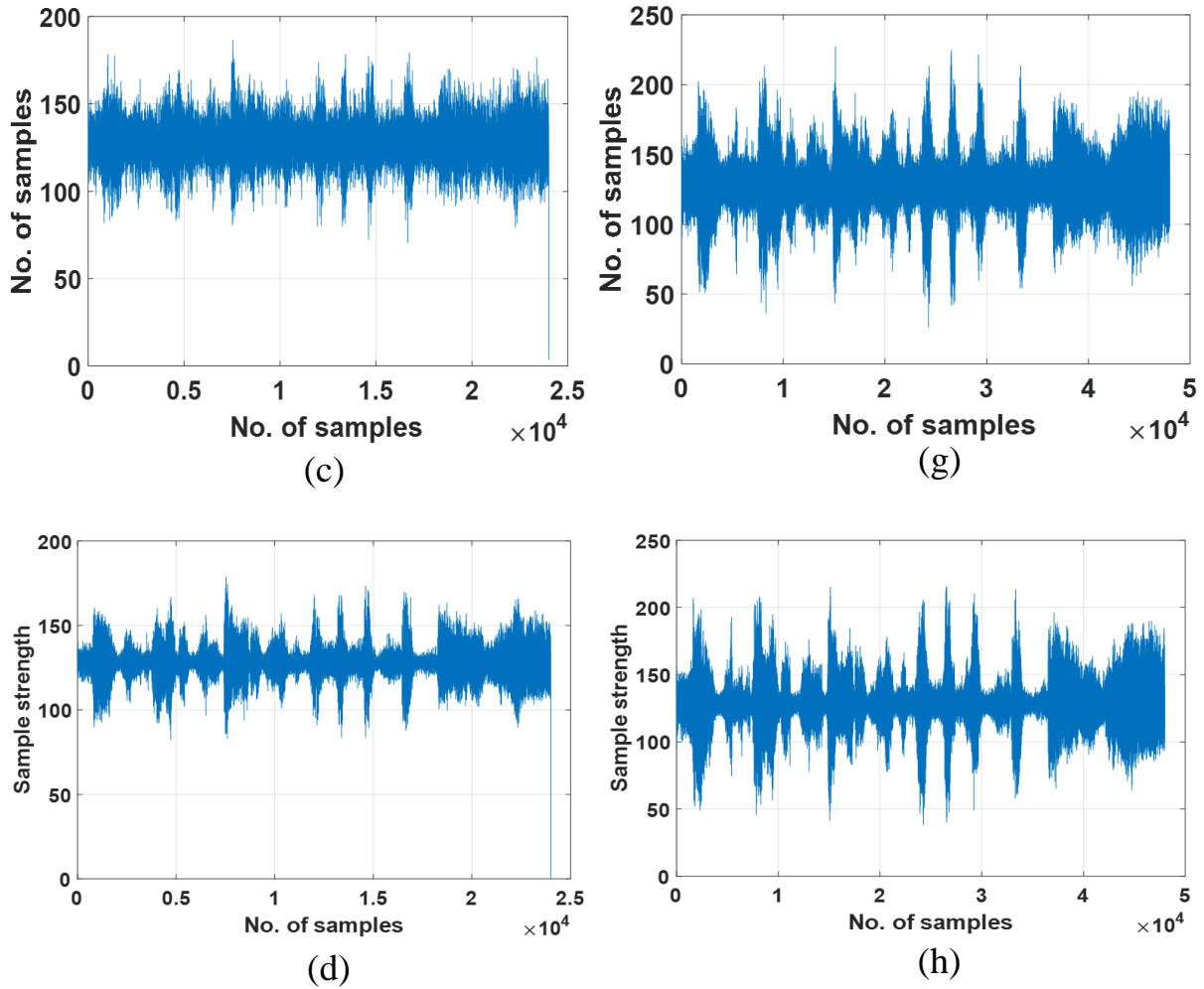
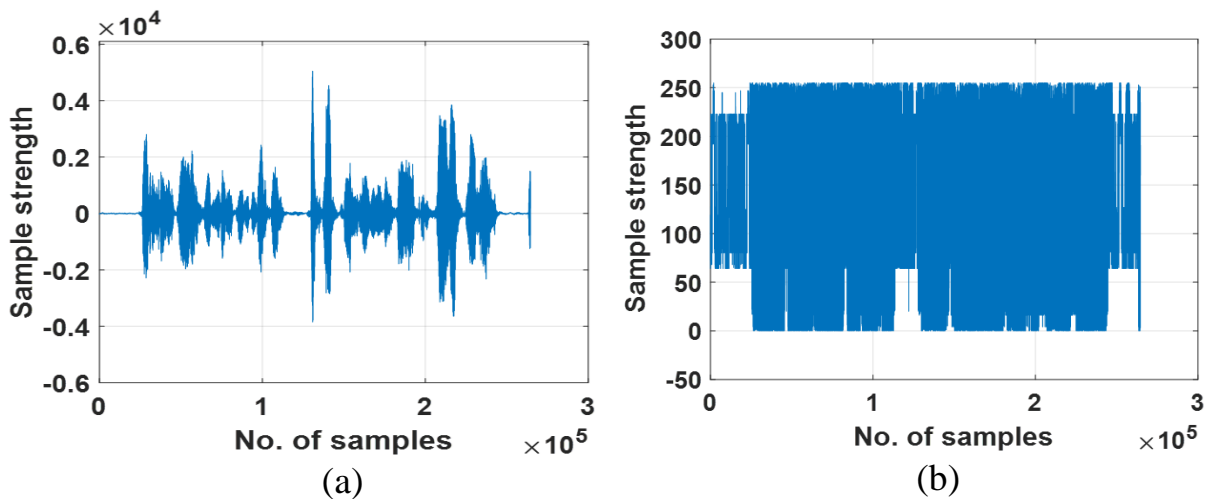


Figure 4.3 The retrieved compressed secret data (a, b, c, and d) and decompressed secret speech (e, f, g, and h) from the first proposed model with AWGN = 30dB at HB=0, 1, 2, and 3 respectively.

4.3 Results of the Proposed Model based on iLBC CODEC

To test the performance of the proposed model, several input speech signals are employed as cover and secret signals. Similar to the first proposed model, the tested signals can be divided into two groups: own-recorded speech signals (twenty signals) and forty speeches from the datasets that are named "development set, clean speech" and "waves_yesno". The addressed signals are six from the first group: three as cover signals (Cov_speech1, Cov_speech2, and

Cov_speech3) and three as secret signals (Sec_speech1, Sec_speech2, and Sec_speech3). From the second group, also six signals are addressed: three as cover signals (Speech1, Speech2, and Speech3) and three as secret signals (Wave1, Wave2, and Wave3). However, the processing running on the incoming cover and secret frames (with a duration of 20 msec per frame) can be summarized in Figures (4.4) and (4.5). In Figure (4.4b), the compressed cover samples (compressed with G.711 encoder) have 8 bits per sample instead of 16 bits per sample with a strength between 0 and 255 for each sample. These compressed cover samples are used to embed the secret information to give the stego data as in Figure (4.4c). There is no noticeable difference between the compressed cover data and the stego data which means that the cover speech quality has been preserved. The decompressed cover speech has undetectable degradation due to the cover speech Codec.



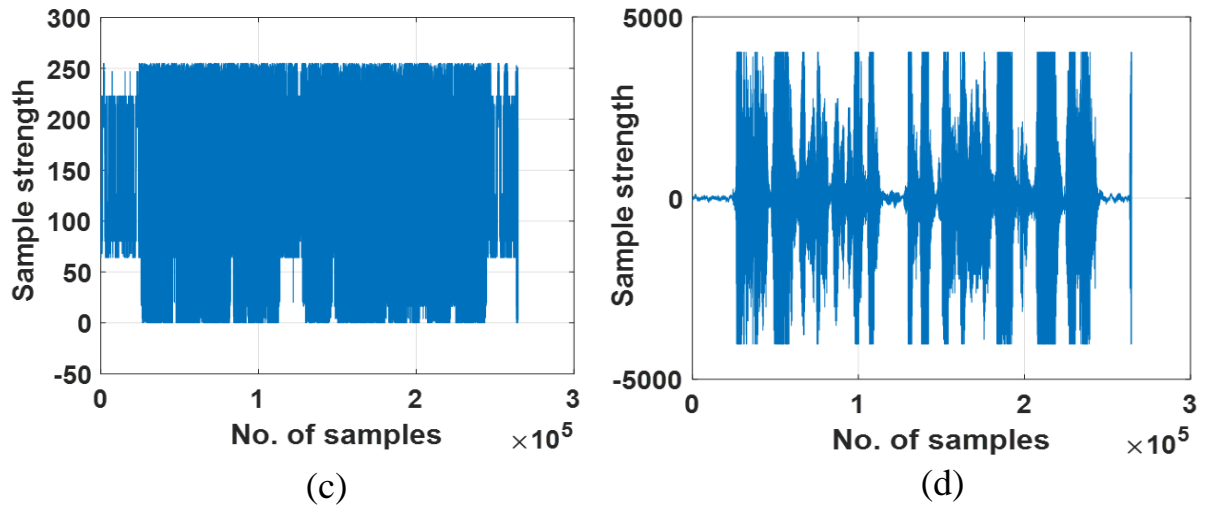
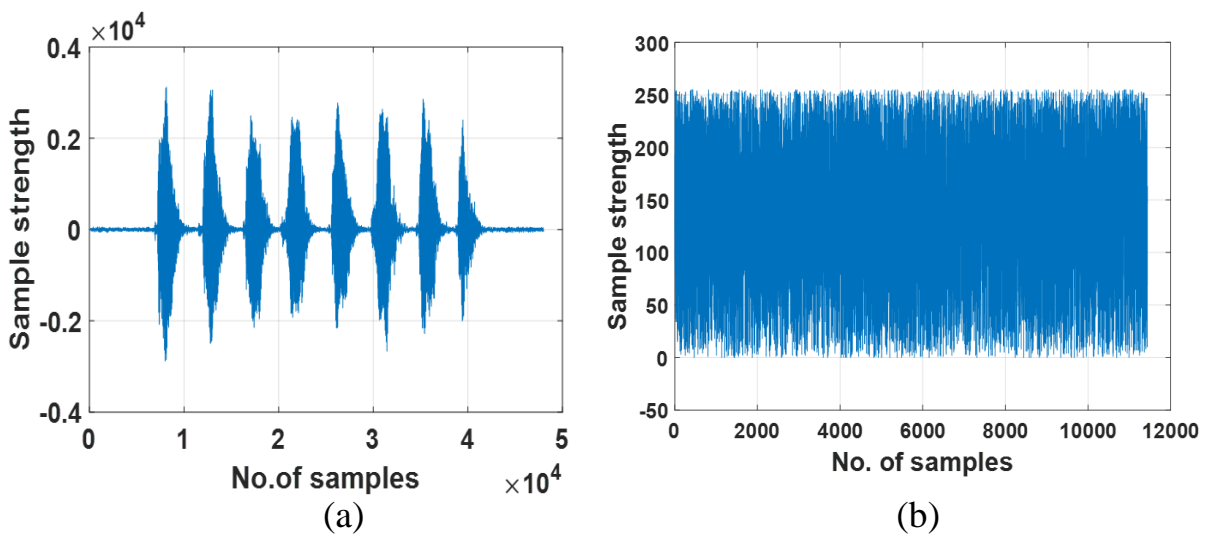


Figure 4.4 The processed cover speech in the second proposed model (a) original, (b) compressed, (c) stego-data, and (c) decompressed cover speech.

Figure (4.5) shows the processed secret speech at the compression and embedding stages. At the compression stage, each secret frame is compressed using iLBC encoder to 11.9% from its original data rate with a resolution of 8 bits per sample, thus the number of secret samples is reduced as in Figure (4.5b). These compressed samples are embedded in the compressed cover data and sent to the receiver.



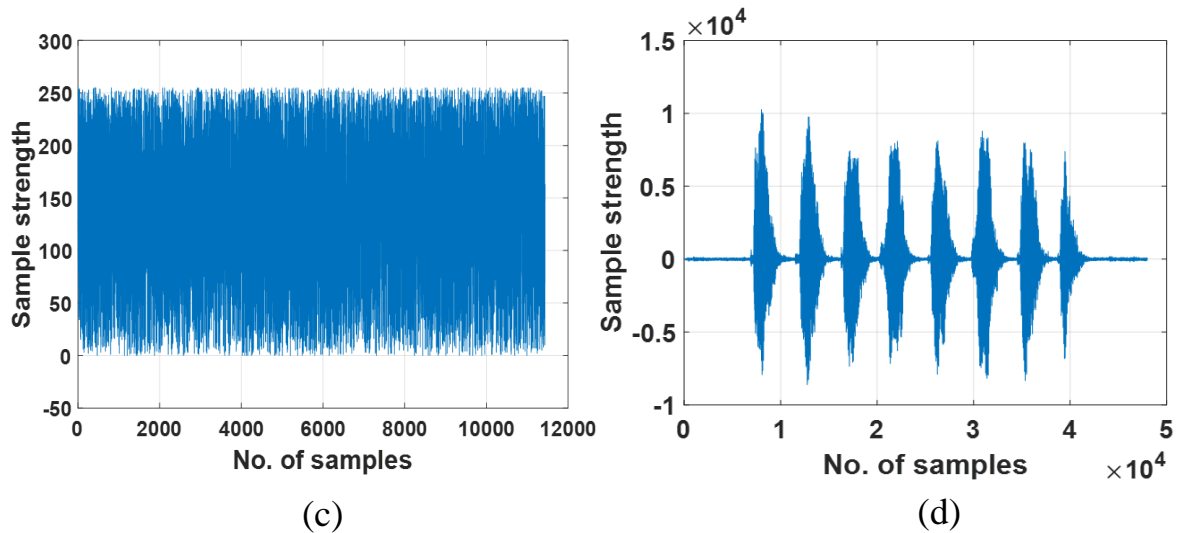


Figure 4.5 The processed secret speech in the second proposed model (a) original, (b) compressed, (c) compressed retrieved, and (c) decompressed retrieved.

4.3.1 Tests of Perceptual Quality and Data Embedding Rate

The perceptual quality or imperceptibility of this model has been measured mathematically using two performance inductors which are *SNR* as in Eq. (2.19) and *PESQ*. However, each secret signal is embedded in all of the three cover signals as seen in Tables (4.10), (4.11), (4.12), and (4.13) at a lossless noisy channel (*HB* factor=0). The obtained results show that the proposed scheme has excellent imperceptibility (45.19 dB and 4.25 in term of *SNR* and *PESQ* respectively) for various cover and secret input speech signals. This is because the introduced embedding process based on high-strength cover samples for hiding the secret data. Thus, the cover speech quality is maintained by minimizing the difference between the cover and modified cover data to lowest values

Table 4.10 Tests the quality of the stego-signals for the second proposed model (the inputs are recorded speeches with a 32kHz cover sampling frequency).

<i>Secret signals</i>	<i>Cover signals</i>	<i>Thr</i>	<i>SNR</i>	<i>PESQ</i>
Sec_speech1	Cov_speech1	64	45.1369	4.2186
		128	45.2041	4.2182
	Cov_speech2	64	44.5455	4.2039
		128	44.5908	4.2038
	Cov_speech3	64	45.1399	4.1991
		128	45.2368	4.1992
Sec_speech2	Cov_speech1	64	45.1276	4.2181
		128	45.2361	4.2179
	Cov_speech2	64	44.6142	4.2040
		128	44.5854	4.2044
	Cov_speech3	64	45.1747	4.1993
		128	45.2677	4.1992
Sec_speech3	Cov_speech1	64	45.1141	4.2182
		128	45.2344	4.2183
	Cov_speech2	64	44.5594	4.2040
		128	44.5605	4.2039
	Cov_speech3	64	45.1774	4.1993
		128	45.2607	4.1991

Table 4.11 Tests the quality of the stego-signals for the second proposed model (the inputs are speeches from dataset with a 32kHz cover sampling frequency).

<i>Secret signals</i>	<i>Cover signals</i>	<i>Thr</i>	<i>SNR</i>	<i>PESQ</i>
Wave1	Speech1	64	45.5348	4.4342
		128	45.5571	4.4393
	Speech2	64	45.2706	4.2439
		128	45.3623	4.2451
	Speech3	64	45.2333	4.1950
		128	45.3369	4.1952
Wave2	Speech1	64	45.5587	4.4346
		128	45.5665	4.4305
	Speech2	64	45.2379	4.2438
		128	45.3388	4.2443
	Speech3	64	45.2794	4.1951
		128	45.3788	4.1952
Wave3	Speech1	64	45.5320	4.4364
		128	45.5617	4.4380
	Speech2	64	45.2044	4.2442
		128	45.3467	4.2441
	Speech3	64	45.2738	4.1950
		128	45.3404	4.1952

Furthermore, the perceptual quality is also improved in the case of raising the sampling frequency of the cover speech from 32 kHz to 44.1 kHz as written in Tables (4.12) and (4.13), where the average values for *SNR* and *PESQ* are raised to 46.63 dB and 4.32 respectively. This is due to increasing the number of samples per second and as a result, all the secret data will be embedded even when the cover sample strength employed in the embedding process is decreased.

Table 4.12 Tests the quality of the stego-signals for the second proposed model (the inputs are recorded speeches with a 44.1kHz cover sampling frequency).

<i>Secret signals</i>	<i>Cover signals</i>	<i>Thr</i>	<i>SNR</i>	<i>PESQ</i>
Sec_speech1	Cov_speech1	64	46.6649	4.3051
		128	46.6840	4.3053
	Cov_speech2	64	45.8279	4.2997
		128	45.9333	4.2995
	Cov_speech3	64	46.7117	4.2978
		128	46.7460	4.2977
Sec_speech2	Cov_speech1	64	46.6901	4.3051
		128	46.7085	4.3054
	Cov_speech2	64	45.8316	4.2996
		128	45.9619	4.2995
	Cov_speech3	64	46.7266	4.2978
		128	46.7584	4.2977
Sec_speech3	Cov_speech1	64	46.6963	4.3051
		128	46.6863	4.3052
	Cov_speech2	64	45.8452	4.2996
		128	46.0067	4.2995
	Cov_speech3	64	46.7217	4.2977
		128	46.7233	4.2979

Table 4.13 Tests the quality of the stego-signals for the second proposed model (the inputs are speeches from dataset with a 44.1kHz cover sampling frequency).

<i>Secret signals</i>	<i>Cover signals</i>	<i>Thr</i>	<i>SNR</i>	<i>PESQ</i>
Wave1	Speech1	64	46.9364	4.4120
		128	47.0421	4.4155
	Speech2	64	46.6539	4.3242
		128	46.7136	4.3256
	Speech3	64	46.6949	4.2889
		128	46.8004	4.2890
Wave2	Speech1	64	46.9819	4.4139
		128	47.0360	4.4150
	Speech2	64	46.6320	4.3243
		128	46.7149	4.3261
	Speech3	64	46.6431	4.2889
		128	46.8034	4.2891
Wave3	Speech1	64	46.9714	4.4111
		128	47.0542	4.4122
	Speech2	64	46.7021	4.3247
		128	46.7568	4.3267
	Speech3	64	46.6587	4.2889
		128	46.8058	4.2892

High embedding capacity is achieved that can be computed as a bitrate or ratio as written bellow:

$$\text{Data embedding rate} = 8000 \frac{\text{sample}}{\text{second}} \times 16 \frac{\text{bit}}{\text{sample}} = 128000 \text{bps}$$

$$\text{Data embedding ratio} = \frac{\text{hiden data size}}{\text{cover data size}} = \frac{8000 \frac{\text{sample}}{\text{second}} \times 16 \frac{\text{bit}}{\text{sample}}}{32000 \frac{\text{sample}}{\text{second}} \times 16 \frac{\text{bit}}{\text{sample}}}$$

$$= \frac{128000 \text{ bit per second}}{512000 \text{ bit per second}}$$

$$= 25 \% \text{ from the cover spech size}$$

4.3.2 Tests of Robustness against Additive Noise

The similarity between the original secret speech and the retrieved secret speech has been evaluated by calculating the *NC* and *BER* parameters with respect to Eq. (2.20) and (2.21) respectively. The value of *NC* or *BER* is computed for the compressed secret data before the embedding process and after the recovery process. The obtained values confirm that the secret data is fully retrieved where the value of *NC* is 1 and *BER* is 0 as seen in Tables (4.14) and (4.15). Also, these indicator measurements are computed for overall system (between the original and retrieved decompressed secret speech), where it not less than 0.92 and 0.25% as average values in term of *NC* and *BER* respectively.

Table 4.14 Robustness test results for the second proposed model (the inputs are recorded speeches).

Secret signals	Cover signals	Thr	Embedding Stage		Overall system	
			NC	BER%	NC	BER%
Sec_speech1	Cov_speech1	64	1	0	0.9301	0.1276
		128	1	0		
	Cov_speech2	64	1	0		
		128	1	0		
	Cov_speech3	64	1	0		
		128	1	0		
Sec_speech2	Cov_speech1	64	1	0	0.9967	0.1344
		128	1	0		
	Cov_speech2	64	1	0		
		128	1	0		
	Cov_speech3	64	1	0		
		128	1	0		
Sec_speech3	Cov_speech1	64	1	0	0.9599	0.3261
		128	1	0		
	Cov_speech2	64	1	0		
		128	1	0		
	Cov_speech3	64	1	0		
		128	1	0		

Table 4.15 Robustness test results for the second proposed model (the inputs are speeches from dataset).

Secret signals	Cover signals	Thr	Embedding Stage		Overall system	
			NC	BER%	NC	BER%
Wave1	Speech1	64	1	0	0.8542	0.3102
		128	1	0		
	Speech2	64	1	0		
		128	1	0		
	Speech3	64	1	0		
		128	1	0		
Wave2	Speech1	64	1	0	0.8826	0.3171
		128	1	0		
	Speech2	64	1	0		
		128	1	0		
	Speech3	64	1	0		
		128	1	0		
Wave3	Speech1	64	1	0	0.8794	0.3239
		128	1	0		
	Speech2	64	1	0		
		128	1	0		
	Speech3	64	1	0		
		128	1	0		

With adding noise (AWGN) to the transmitted stego data, the robustness of the proposed system should be enhanced, this is achieved by increasing the depth of embedding based on the value of *HB* factor. As seen in Table (4.16), where the raising of the *HB* factor causes an increment in the immunity of the retrieved secret speech against AWGN but simultaneously reduces the stego-speech quality at an acceptable range.

Table 4.16 Some tests for the immunity of the secret speech against additive noise for the second proposed model.

<i>HB</i>	<i>SNR for stego-data</i>	<i>NC for different AWGN in dB</i>				
		<i>50 dB</i>	<i>40 dB</i>	<i>30 dB</i>	<i>20 dB</i>	<i>10 dB</i>
0	45.3397	0.985	0.936	0.898	0.817	0.743
1	39.2947	0.987	0.962	0.925	0.825	0.772
2	33.3250	0.991	0.977	0.936	0.857	0.801
3	27.2120	1	0.983	0.959	0.915	0.844

4.3.3 Test of Perceptual Quality in a Lossy Channel with Different Levels of Packet Losses

The quality of the proposed model has been tested in a lossy channel with different levels of packet losses according to the MATLAB model that is presented in Figure (4.6). The packet losses could be occurred due to noise, routing, or/and latency. These losses decrease the quality of the transmitted data. In the proposed system, the increasing in the percentage of packet losses from 2% to 10 % will cause an acceptable degradation in the stego-data and hidden data quality, where the average values of *SNR* and *NC* are 35 dB and 0.88 respectively.

When the percentage of packet losses is more than 10%, the degradation in quality will appear clearly and cause a distortion in both the stego-data and the retrieved secret speech. Figure (4.7a) shows the effecting of packet losses on the *SNR*, while Figure (4.7b) shows the effecting of packet losses on the *NC*.

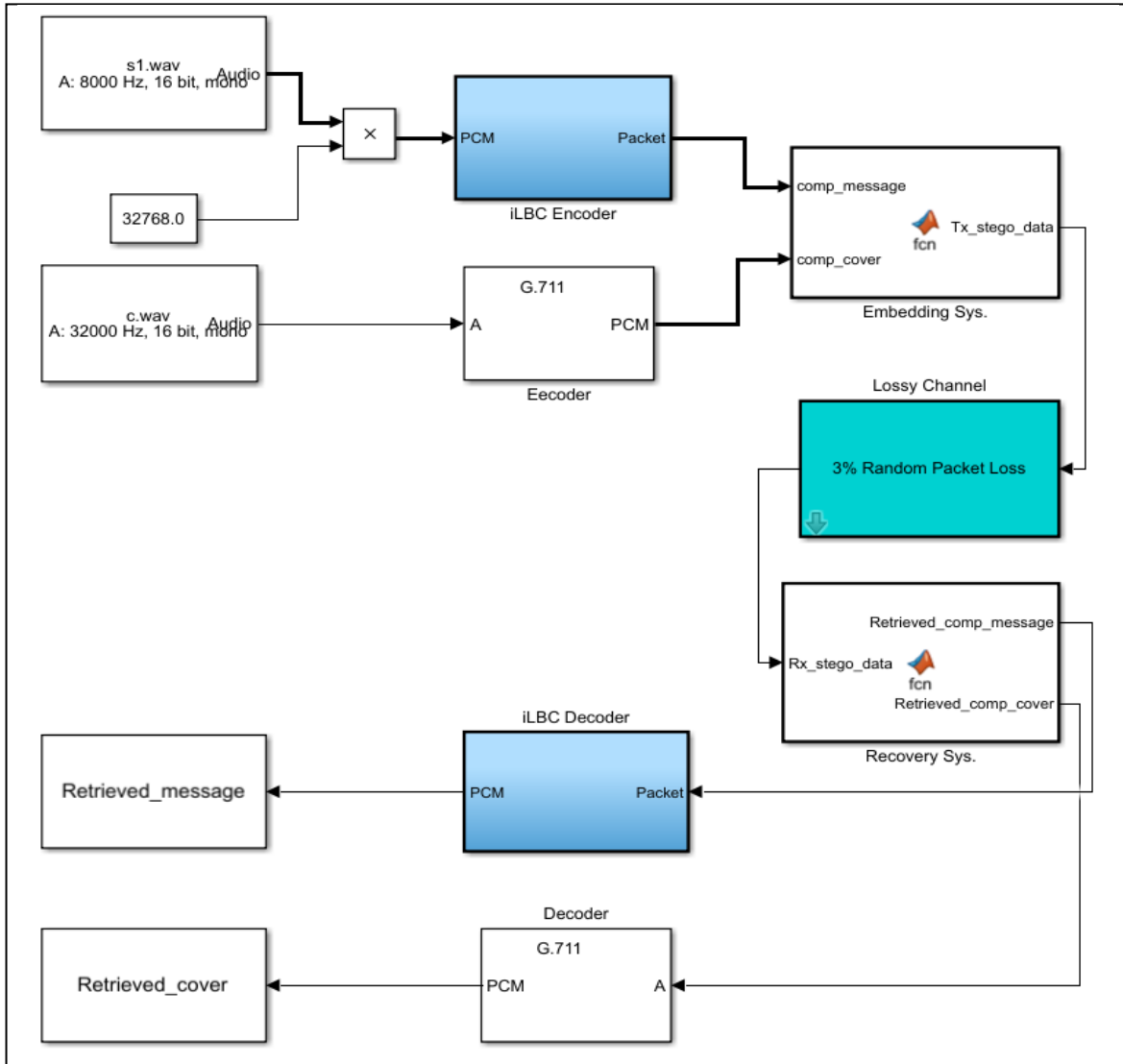
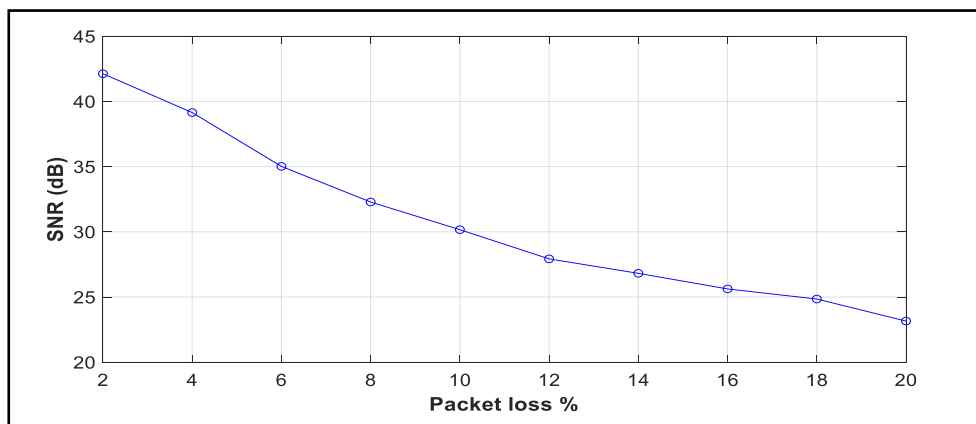
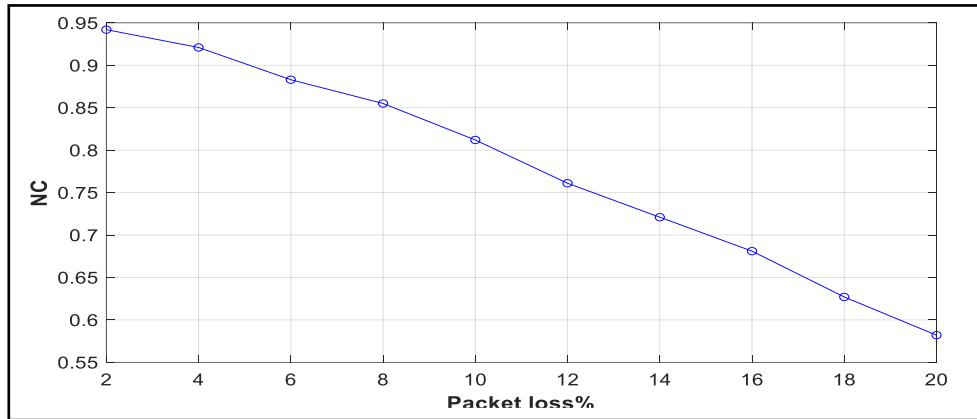


Figure 4.6 Testing the second proposed system quality at a lossy channel with random packet losses.



(a)



(b)

Figure 4.7 Tests of the effecting of the packet losses on the second proposed model (a) SNR of the received stego data vs. packet losses (b) NC of the retrieved secret speech vs. packet losses.

4.4 Computational Complexity and Processing Time

The system complexity is a critical issue for real-time communications. Each of the two proposed models consists of two stages: compression and embedding. The compression stage is a standard that already proved it meets real-time communications. To measure the complexity of the second stage (embedding stage) for a period of 16 msec and for a period of 20 msec, the proposed procedure can be compared with other more complex procedures such as Fast Fourier Transform (FFT) spectrum analyzer that operates in real-time [67] and deals with complex operations. Traditionally, each complex addition operation or multiplication operation is analogous to four times real addition or eight times real multiplication operation respectively.

The introduced embedding algorithm deals with only real operations which are equal to 1024 addition and 512 multiplication operations for the first proposed model and 1280 addition and 640 multiplication operations for the second proposed model. Therefore, the complexity of our embedding stage is about 18.75% of the FFT complexity (for each of the proposed models). These percentages confirm the suitability of the proposed approach to be implemented in real-time with low computational complexity and thereby low processing time within the human imperceptibility rate

4.5 Discussion of the Proposed Approach

The main experimental results offered in the previous sections are discussed in the following sub-sections:

4.5.1 Perceptual Quality

The quality of the stego- signals is not less than 41 dB and 4 in terms of *SNR* and *PESQ*. This is due to several reasons:

- a. The probability of changing in the cover data is 50% at the embedding process because if the determined secret digit is similar to the remainder (R), then the stego sample is exactly the same as the candidate cover sample without any change. For example, if the cover sample is 66 and the secret digit is 2, then according to Eq. (3.2) and (3.3):

$$R = 66 \bmod 4 = 2$$

$$\hat{C} = 66 - (2 - 2) = 66$$

As a result, the probability of errors is decreased to 50%. This percentage is also processed as in the following example:

If the cover sample is 135 (10000111) and the secret bit is **1**, then before the processing, the modified cover sample is 143 (10001111); after the processing, the modified cover sample is 136 (10001000). Therefore, the value of error is reduced from 8 to 1 for this example.

- b. The use of high strength cover samples at the embedding process can preserve the perceptual quality for the stego-signals in terms of *SNR* at 41dB or more in the case of a lossless noisy channel ($HB=0$) and at 35dB or more in the case of a noisy channel ($HB>0$), where the value of *Thr* has been chosen as a 64 or 128, which refers if the selected cover sample has a strength of 64 or more will be employed in the embedding process. However, if the strength of the cover sample is smaller than *Thr*, the quality of the stego data is decreased as defined in the perceptual quality results, where the value of *SNR* is decreased with decreasing the *Thr* value from 128 to 64. Also, the raising of the cover sample strength above 128 will not ensure embedding all the secret data.
- c. Compressing the secret speech before embedding it, can also maintain the quality of the cover speech, where the size of the secret speech after the compression process is much reduced (to 50% or 11.9% from its original

size). As a result, this compressed size does not greatly affect the cover speech quality.

4.5.2 Robustness Against Additive Noise

At a lossless noisy channel ($HB=0$), the retrieved compressed secret data is fully retrieved without errors, where the values of $NC=1$ and $BER=0$. While the decompressed secret speech is retrieved with imperceptible differences to the human ear due to the compression and decompression processes on the secret speech which cause some losses in the compressed secret data; where the average values for the NC and BER are not less than 0.92 and 0.0025 respectively.

At a noisy channel ($HB>0$), the immunity or robustness of the retrieved secret speech is improved by transforming the positions of the embedding from low energy to high energy based on the amount of additive noise (AWGN) that applied on the stego-data.

Table (4.17) illustrates an example on the changing of HB factor, where in the case of $HB=0$ the secret digit will be embedded in a low energy position which is more sensitive to noise. Thus, at varying the value of HB , the secret data will be embedded in a high energy position.

Table 4.17 An example on the varying of the HB factor

<i>Cover sample</i>	<i>Secret digit</i>	<i>HB</i>	<i>Stego-sample</i>
154 (10011010)	3 (<u>11</u>)	0	155 (100110 <u>11</u>)
		1	158 (100111 <u>10</u>)
		2	158 (1001 <u>11</u> 10)
		3	154 (100 <u>1</u> 1010)

4.6 Comparison between the Two Proposed Models

Each of the proposed models is suitable to real-time communications based VoIP, uses G.711 Codec to compress the cover speech, and in a lossless noisy channel, the retrieved compressed secret data is an error free. The main differences between the two models can be summarized by Table (4.18).

Table 4.18 A comparison between the two proposed models.

<i>The first proposed model</i>	<i>The second proposed model</i>
Compress the secret speech using a proposed fast compression algorithm (using Int2Int-Haar LWT). Thus, it provides fast execution.	Compress the secret speech with iLBC Codec which consists of a set processes (see appendix A).
A secret speech is compressed to 50% from its original size.	A secret speech is compressed to 11.9% from its original size. Thus, this model provides excellent quality.
The secret speech has a sampling rate of 8 kHz or more with 8 bits per sample of a resolution.	The secret speech has a sampling rate of 8 kHz with 16 bits per sample of a resolution. Thus, it is more accurate.
The period of the input cover or secret frame is 16 msec.	The period of the input cover or secret frame is 20 msec.
$SNR=42.44$ and $PESQ=4.19$	$SNR=45.91$ and $PESQ=4.28$
Embedding capacity is 12.5% from the size of the cover speech.	Embedding capacity is 25% from the size of the cover speech.

In a noisy channel, it has acceptable robustness.	In a noisy channel, it has good or acceptable robustness. Also, there is ability to recover the lost packet of the secret data.
It does not handle the lost or delayed retrieved secret speech packets.	It handles the lost or delayed retrieved secret speech packets. In other words, it has auto correction system.

4.7 Comparison with the Related Works

To confirm the effectiveness of the introduced real-time covert communication approach, comparisons of the quality performance, data embedding rate, and robustness for the proposed models with other related works are conducted in this section.

4.7.1 Perceptual Quality and Data Embedding Rate

The suggested models are superior as compared with other related works in terms of quality performance, where it is not less than 42dB/45dB (for the first or second proposed model) of *SNR* and 4 of *PESQ* as indicated in Table (4.19). Most existing literatures in the VoIP steganography area are focused to embed offline secret information which may be image, text, or speech into a real-time cover speech [7][20][30] [31][33]. Therefore, their embedding rates are insufficient for hiding a real-time secret speech. To overcome this problem, the secret speech is compressed to 50% from its original size for the first proposed model and to 11.9% from its original size for the second proposed model.

In this study, a 64000 and 128000 bit per second (as data embedding rate for the first and second proposed models respectively) of a real-time secret speech has been hidden inside a real-time cover speech. Thus, the obtained embedding rates are more beneficial than most other related literature. However, the comparison with other related work in Table (4.19) is based on *SNR* and *PESQ* average values and data embedding rate.

Table 4.19 Comparative results for the quality and data embedding rate.

<i>Parameter</i> <i>Reference</i>	<i>SNR</i>	<i>PESQ</i>	<i>Data embed- ding rate (bps)</i>
<i>Zhijun</i> [30]	Not reported	3.11	2400
<i>Jiang</i> [31]	38.7	4.04	8000
<i>Peng</i> [33]	Not reported	4.3	800
<i>Tang</i> [7]	27.5	3.5	3968
<i>Li</i> [34]	Not reported	3.02	483
<i>Lin</i> [4]	Not reported	3.811	2150
<i>Roselinkiruba</i> [20]	Not reported	Not reported	13300
<i>First proposed model</i>	42.44	4.19	64000
<i>Second proposed model</i>	45.91	4.28	128000

4.7.2 Robustness of the Proposed Models

In addition to the superiority of the introduced models in terms of the perceptual quality, they can also fully retrieve the secret speech as confirmed in Tables (4.5), (4.6), (4.7), (4.8), (4.14), and (4.15) which give $NC=0$ and $BER=1$ for different real-time input cover and secret speech signals. With adding noise (AWGN), the introduced models enhance the robustness level by adjusting the

depth of embedding in hiding the secret data thereby can resist channel losses. Whereas [17][18][7][23] and most of the existing real-time steganographic techniques have used only LSB replacements to hide the secret data. Therefore, they are weak or very sensitive to additive noise. Moreover, several literatures embed the secret data before the compression stage [21][22][23][24][12], which makes the hidden data is more vulnerable to damage or loss during the compression. Thus, there is a weakness in their robustness. To avoid this problem, the secret data have been embedded in the compressed cover data (after the compression process).

4.8 Chapter Summary

The experimental results for the two proposed systems are presented in this chapter. Several real-time input cover and secret speech signals are employed to evaluate the performance of this work based on the basic key features for any efficient data hidden system. The results of tests confirm that the introduced systems have excellent perceptual quality, robustness against noise with fully retrieved secret speech when there is no noise. Moreover, they have high embedding capacity of about 12.5% (for the first proposed system) and 25% (for the second proposed system) from the size of the cover speech. Furthermore, they have low complexity with suitable processing time for VoIP communications requirements. Based on these features, the obtained results of the suggested systems are better than similar approaches in the related literatures.

CHAPTER Five

CONCLUSION AND FUTURE WORK

5.1 Conclusion

The most important conclusions that can be derived from the proposed real-time covert communication approaches based VoIP are:

1. A high data embedding rate is obtained that is a 64 kbps/128 kbps in real-time due to compressing the input secret speech and thereby reducing the required size of hiding.
2. The cover speech quality is maintained; firstly, by compressing the secret speech and secondly by using the high-strength cover samples to hide the secret data.
3. For avoiding the losses in the embedded data through the compression provided by VoIP, the secret data are embedded in the compressed cover speech.
4. The embedded data has sufficient immunity against noise (AWGN) due to the embedding at higher positions (based *HB* factor). Also, the security of the proposed approach is improved due to using random permutations for the secret data.
5. In a lossy channel, there is an auto-correction system for the lost speech packets that is provided by the iLBC Codec.

6. The proposed embedding algorithm has low complexity with low processing time suitable for real-time communications.
7. The proposed approach is simple and can be attached to any chatting software that utilizes VoIP without any extra cost to be employed in high secure communications such as military and government communications.

5.2 Future Work

The suggested developments of the proposed approaches are demonstrated for future research:

1. Mobile communication such as Global System for Mobile communications (GSM) can be used as a public channel to embed the secret data.
2. The stego-data can be exposed to a lossy Codec such as G.723 which adds extra non-linear noises.
3. Other audio codecs proper to VoIP, such as G.729, G.726 can be employed.
4. The same proposed approach can be applied to the Walkie-Talkie communications used by the police, so instead of having one person speaking, there are two people.

References

- [1] H. I. Shahadi, R. Jidin, and W. H. Way, “A novel and high capacity audio steganography algorithm based on adaptive data embedding positions,” *Res. J. Appl. Sci. Eng. Technol.*, vol. 7, no. 11, pp. 2311–2323, 2014.
- [2] N. Shetty, “Steganography for Secure Data Transmission,” *Int. J. Comput. Intell. Res.*, vol. 13, no. 10, pp. 2289–2295, 2017.
- [3] H. Dutta, R. K. Das, S. Nandi, and S. R. M. Prasanna, “An Overview of Digital Audio Steganography,” *IETE Tech. Rev. (Institution Electron. Telecommun. Eng. India)*, vol. 37, no. 6, pp. 632–650, 2020.
- [4] R. S. Lin, “A synchronization scheme for hiding information in encoded bitstream of inactive speech signal,” *J. Inf. Hiding Multimed. Signal Process.*, vol. 7, no. 5, pp. 916–929, 2016.
- [5] S. Deepikaa and R. Saravanan, “VoIP steganography methods, a survey,” *Cybern. Inf. Technol.*, vol. 19, no. 1, pp. 73–87, 2019.
- [6] F. Djebbar, B. Ayad, H. Hamam, and K. Abed-Meraim, “A view on latest audio steganography techniques,” *2011 Int. Conf. Innov. Inf. Technol. IIT 2011*, pp. 409–414, 2011.
- [7] S. Y. Tang, Y. J. Jiang, L. P. Zhang, and Z. B. Zhou, “Audio steganography with AES for real-time covert voice over internet protocol communications,” *Sci. China Inf. Sci.*, vol. 57, no. 3, pp. 1–14, 2014.
- [8] Z. Wu, J. Guo, C. Zhang, and C. Li, “Steganography and steganalysis in voice over ip: A review,” *Sensors (Switzerland)*, vol. 21, no. 4, pp. 1–31,

2021.

- [9] A. Khan, A. Siddiqa, S. Munib, and S. A. Malik, “A recent survey of reversible watermarking techniques,” *Inf. Sci. (Ny)*, vol. 279, no. June 2019, pp. 251–272, 2014.
- [10] H. I. Shahadi and R. Jidin, “High capacity and inaudibility audio steganography scheme,” *Proc. 2011 7th Int. Conf. Inf. Assur. Secur. IAS 2011*, pp. 104–109, 2011.
- [11] F. Djebbar, B. Ayad, K. A. Meraim, and H. Hamam, “Comparative study of digital audio steganography techniques,” *Eurasip J. Audio, Speech, Music Process.*, vol. 2012, no. 1, pp. 1–16, 2012.
- [12] H. I. Shahadi, “Covert communication model for speech signals based on an indirect and adaptive encryption technique,” *Comput. Electr. Eng.*, vol. 68, no. December 2017, pp. 425–436, 2018.
- [13] A. H. Ali, M. R. Mokhtar, and L. E. George, “Enhancing the hiding capacity of audio steganography based on block mapping,” *J. Theor. Appl. Inf. Technol.*, vol. 95, no. 7, pp. 1441–1448, 2017.
- [14] A. H. Ali, M. R. Mokhtar, and L. E. George, “Recent approaches for VoIP steganography,” *Indian J. Sci. Technol.*, vol. 9, no. 38, 2016.
- [15] A. Z. Al-othmani, A. A. Manaf, and A. M. Zeki, “A Survey on Steganography Techniques in Real Time Audio Signals and Evaluation,” *Int. J. Comput. Sci. Issues*, vol. 9, no. 1, pp. 30–37, 2012.
- [16] B. Datta, S. Tat, and S. Bandyopadhyay, “Robust High Capacity Audio

- Steganography using Modulo Operator,” *Proc. 2015 Third Int. Conf. Comput. Commun. Control Inf. Technol.*, 2015.
- [17] L. Voleti, R. Balajee, S. Vallepu, K. Bayoju, and D. Srinivas, “A Secure Image Steganography Using Improved Lsb Technique and Vigenere Cipher Algorithm,” *Proc. - Int. Conf. Artif. Intell. Smart Syst. ICAIS 2021*, pp. 1005–1010, 2021.
- [18] B. Budumuru, G. Kumar, and B. Raju, “Hiding an image in an audio file using LSB audio technique,” *Int. Conf. Commun. Informatics, ICCCI 2021*, no. 2329–7190, pp. 5–8, 2021.
- [19] M. H. Shirali-Shahreza and S. Shirali-Shahreza, “Real-time and MPEG-1 layer III compression resistant steganography in speech,” *IET Inf. Secur.*, vol. 4, no. 1, pp. 1–7, 2010.
- [20] R. Roselinkiruba and R. Balakirshnan, “Secure steganography in audio using inactive frames of VoIP streams,” *2013 IEEE Conf. Inf. Commun. Technol. ICT 2013*, no. Ict, pp. 491–495, 2013.
- [21] H. I. Shahadi, R. Jidin, and W. H. Way, “Lossless audio steganography based on lifting wavelet transform and dynamic stego key,” *Indian J. Sci. Technol.*, vol. 7 (3), pp. 323–334, 2014.
- [22] S. E. El-Khamy, N. O. Korany, and M. H. El-Sherif, “Robust image hiding in audio based on integer wavelet transform and Chaotic maps hopping,” *Natl. Radio Sci. Conf. NRSC, Proc.*, no. Nrsc, pp. 205–212, 2017.

- [23] S. S. Bharti, M. Gupta, and S. Agarwal, “A novel approach for audio steganography by processing of amplitudes and signs of secret audio separately,” *Multimed. Tools Appl.*, vol. 78, no. 16, pp. 23179–23201, 2019.
- [24] P. M. Kumar and K. Srinivas, “Real Time Implementation of Speech Steganography,” *Proc. 2nd Int. Conf. Smart Syst. Inven. Technol. ICSSIT 2019*, vol. 23, no. 1, pp. 365–369, 2019.
- [25] C. Wang and Q. Wu, “Information hiding in real-time VoIP streams,” *Proc. - 9th IEEE Int. Symp. Multimedia, ISM 2007*, pp. 255–262, 2007.
- [26] H. Tian, K. Zhou, H. Jiang, J. Liu, Y. Huang, and D. Feng, “An M-sequence based steganography model for voice over IP,” *IEEE Int. Conf. Commun.*, 2009.
- [27] H. Tian, K. Zhou, H. Jiang, Y. Huang, J. Liu, and D. Feng, “An adaptive steganography scheme for voice over IP,” *Proc. - IEEE Int. Symp. Circuits Syst.*, pp. 2922–2925, 2009.
- [28] H. Tian, H. Jiang, K. Zhou, and D. Feng, “Adaptive partial-matching steganography for voice over IP using triple M sequences,” *Comput. Commun.*, vol. 34, no. 18, pp. 2236–2247, 2011.
- [29] Z. Wei, B. Zhao, B. Liu, J. Su, L. Xu, and E. Xu, “A novel steganography approach for voice over IP,” *J. Ambient Intell. Humaniz. Comput.*, vol. 5, no. 4, pp. 601–610, 2014.
- [30] W. Zhijun, C. Haijuan, and L. Douzhe, “An approach of steganography

- in G.729 bitstream based on matrix coding and interleaving,” *Chinese J. Electron.*, vol. 24 no.1, pp. 157–165, 2015.
- [31] Y. Jiang and S. Tang, “An efficient and secure VoIP communication system with chaotic mapping and message digest,” *Multimed. Syst.*, vol. 24, no. 3, pp. 355–363, 2018.
- [32] F. Li, B. Li, L. Peng, W. Chen, L. Zheng, and K. Xu, *A Steganographic Method Based on High Bit Rates Speech Codec of G.723.1*, vol. 11068 LNCS. 2018.
- [33] J. Peng and S. Tang, “Covert Communication over VoIP Streaming Media with Dynamic Key Distribution and Authentication,” *IEEE Trans. Ind. Electron.*, vol. 68, pp. 3619–3628, 2020.
- [34] F. Li, B. Li, Y. Huang, Y. Feng, L. Peng, and N. Zhou, “Research on covert communication channel based on modulation of common compressed speech codec,” *Neural Comput. Appl.*, vol. 0123456789, 2020.
- [35] F. M. Guo, A. Talevski, and E. Chang, “Voice over Internet protocol on mobile devices,” *Proc. - 6th IEEE/ACIS Int. Conf. Comput. Inf. Sci. ICIS 2007; 1st IEEE/ACIS Int. Work. e-Activity, IWEA 2007*, no. Icis, pp. 163–168, 2007.
- [36] W. Mazurczyk, “VoIP steganography and its detection-a survey,” *ACM Comput. Surv.*, vol. 46, no. 2, 2013.
- [37] M. DeSantis, “Understanding Voice over Internet Protocol (VoIP),” *US*

Cert, pp. 1–5, 2008.

- [38] I. T. Union, “ITU-T Recommendation G.711, Pulse code modulation (PCM) of voice frequencies,” 1988.
- [39] R. K. Das, J. Yang, and H. Li, “Data augmentation with signal companding for detection of logical access attacks,” *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2021-June, pp. 6349–6353, 2021.
- [40] I. Since *et al.*, “iLBC – Designed For The Future,” 2004.
- [41] S. Gupta and N. Dhanda, “Audio Steganography Using Discrete Wavelet Transformation (DWT) & Discrete Cosine Transformation (DCT),” *IOSR J. Comput. Eng.*, vol. 17, no. 2, pp. 2278–661, 2015.
- [42] S. Rekik, D. Guerchi, S. A. Selouani, and H. Hamam, “Speech steganography using wavelet and Fourier transforms,” *Eurasip J. Audio, Speech, Music Process.*, vol. 2012, no. 1, pp. 1–14, 2012.
- [43] M. Liao, X. Dong, J. Chen, and D. Zeng, “An audio steganography based on Twi-DWT and audio-extremum features,” *Chinese Control Conf. CCC*, vol. 2019-July, pp. 8882–8888, 2019.
- [44] G. Tzanetakis, G. Essl, and P. R. Cook, “Audio Analysis using the Discrete Wavelet Transform,” *Proc. WSES Int. Conf. Acoust. Music Theory Appl. (AMTA 2001)*, pp. 318–323, 2001.
- [45] M. Pooyan and A. Delforouzi, “LSB-based audio steganography method based on lifting wavelet transform,” *ISSPIT 2007 - 2007 IEEE Int. Symp.*

Signal Process. Inf. Technol., pp. 600–603, 2007.

- [46] T. Acharya and C. Chakrabarti, “A survey on lifting-based discrete wavelet transform architectures,” *J. VLSI Signal Process. Syst. Signal Image. Video Technol.*, vol. 42, no. 3, pp. 321–339, 2006.
- [47] H. I. Shahadi and R. Jidin, “High Capacity and Resistance to Additive Noise Audio Steganography Algorithm,” *IJCSI Int. J. Comput. Sci. Issues*, vol. 8, no. 5, pp. 176–184, 2011.
- [48] H. I. Shahadi, R. Jidin, and W. H. Way, “High performance FPGA architecture for dual mode processor of Integer Haar Lifting-based Wavelet Transform,” *Int. Rev. Comput. Softw.*, vol. 8, no. 9, pp. 2058–2067, 2013.
- [49] F. A. Abdullatif, A. A. Abdullatif, and N. A. Taha, “Data hiding using integer lifting wavelet transform and DNA computing,” *Period. Eng. Nat. Sci.*, vol. 8, no. 1, pp. 58–66, 2020.
- [50] H. I. Shahadi, R. Jidin, W. H. Way, and Y. A. Abbas, “Efficient FPGA Architecture for Dual Mode Integer Haar Lifting Wavelet Transform Core,” *J. Appl. Sci.*, vol. 14, no. 5, pp. 436–444, 2014.
- [51] P. P. Balgurgi and S. K. Jagtap, “Intelligent processing: An approach of audio steganography,” *Proc. - 2012 Int. Conf. Commun. Inf. Comput. Technol. ICCICT 2012*, pp. 1–6, 2012.
- [52] S. M. Alwahbani and H. T. Elshoush, “Hybrid audio steganography and cryptography method based on high least significant bit (LSB) layers and

- one-time pad—A novel approach,” *Stud. Comput. Intell.*, vol. 751, pp. 431–453, 2018.
- [53] P. Jayaram, H. R. Ranganatha, and H. S. Anupama, “Information Hiding Using Audio Steganography - A Survey,” *Int. J. Multimed. Its Appl.*, vol. 3, no. 3, pp. 86–96, 2011.
- [54] D. Tan, Y. Lu, X. Yan, and X. Wang, “A simple review of audio steganography,” *Proc. 2019 IEEE 3rd Inf. Technol. Networking, Electron. Autom. Control Conf. ITNEC 2019*, no. Itnec, pp. 1409–1413, 2019.
- [55] K. A. Anjana, C. C. Satheesh, S. Kamal, and M. H. Supriya, “Spread spectrum based encrypted audio steganographic system with improved security,” *Proceeding Second Int. Conf. Circuits, Control. Commun. IEEE*, pp. 109–114, 2017.
- [56] Ankur, Divyanjali, and V. Pareek, “A new approach to pseudorandom number generation,” *Int. Conf. Adv. Comput. Commun. Technol. ACCT*, vol. 1, no. X, pp. 290–295, 2014.
- [57] F. James, “A review of pseudorandom number generators,” *Comput. Phys. Commun.*, vol. 60, no. 3, pp. 329–344, 1990.
- [58] N. Tiwari, D. Madhu, and D. Meenu, “Spatial Domain Image Steganography based on Security and Randomization,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 5, no. 1, pp. 156–159, 2014.
- [59] M. H. Azam, F. Ridzuan, M. N. Sayuti, and A. Alsabhany, “Balancing

- the Trade-Off between Capacity and Imperceptibility for Least Significant Bit Audio Steganography Method: A New Parameter,” *2019 IEEE Conf. Appl. Inf. Netw. Secur. AINS 2019*, pp. 48–53, 2019.
- [60] M. H. Al-Hooti, T. Ahmad, and S. Djanali, “Audio Data Hiding Using Octal Modulus Function Based Unsigned Integer Sample Values,” *2018 Int. Conf. Comput. Eng. Netw. Intell. Multimedia, CENIM 2018 - Proceeding*, pp. 48–53, 2018.
- [61] A. H. Ali, L. E. George, A. A. Zaidan, and M. R. Mokhtar, “High capacity, transparent and secure audio steganography model based on fractal coding and chaotic map in temporal domain,” *Multimed. Tools Appl.*, vol. 77, no. 23, pp. 31487–31516, 2018.
- [62] T. Installations and L. Line, “ITU, Perceptual evaluation of speech quality (PESQ), and objective method for endto- end speech quality assessment of narrowband telephone networks and speech codecs. ITU-T Recommendation p. 862,” 2000.
- [63] S. Hemalatha, U. D. Acharya, and A. Renuka, “Wavelet transform based steganography technique to hide audio signals in image,” *Procedia Comput. Sci.*, vol. 47, no. C, pp. 272–281, 2015.
- [64] S. S. Verma, “A Novel Technique for Data Hiding in Audio Carrier by Using Sample Comparison in DWT Domain,” pp. 0–4, 2014.
- [65] “Librispeech: An ASR corpus based on public domain audio books,” <https://www.openslr.org/12/>.

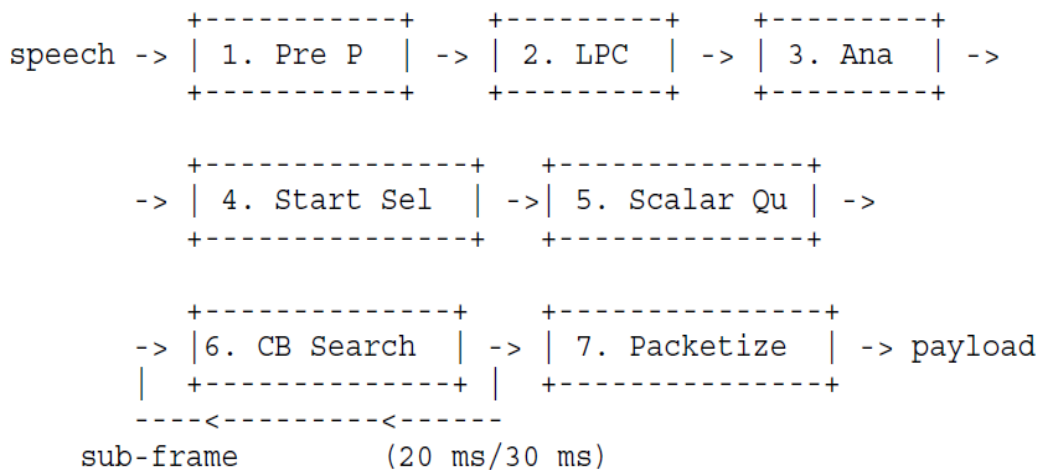
- [66] “waves_yesno,” <https://www.openslr.org/1/>.
- [67] W. Lv, C. Shen, F. Gui, Z. Tian, and D. Jiang, “Real-time spectrum analyzer based on all phase FFT spectrum analysis,” *Proc. - 2013 4th Int. Conf. Digit. Manuf. Autom. ICDMA 2013*, pp. 966–969, 2013.

APPENDIX A

Internet Low Bit Rate Codec (iLBC)

1. iLBC Encoder

All the components of the iLBC encoder is presented in the following block diagram:



1. Pre-process speech with a HP filter.
2. Compute LPC parameters, quantize, and interpolate.
3. Use analysis filters on speech to compute residual.
4. Select position of the start state sample.
5. Quantize the start state sample with scalar quantization.
6. Search the codebook for each sub-frame, then encode sub-blocks forward in time, and then encode sub-blocks backward in time.
7. Packetize the bits into the payload as the compressed speech.

الخلاصة

يعتبر بروتوكول نقل الصوت عبر الإنترنت (VoIP) وسيلة لربط المحادثات الصوتية عبر شبكة الإنترنت. حيث يتم استخدامه في العديد من التطبيقات البرمجية مثل Skype، WhatsApp و Google Talk. بشكل عام، يمكن أن تتعرض الاتصالات عبر الإنترنت بسهولة للقرصنة والتنصت. لضمان سرية المعلومات المرسله وعدم كشفها تقترح هذه الدراسة نهج اتصال سري في الوقت الفعلي يعتمد على توليد قناة خطاب سرية داخل قناة خطاب عامة (شبكة VoIP) بناءً على تقنية إخفاء الصوت. تتطلب العملية قدرة تضمين عالية بطريقة الوقت الفعلي. تم تنفيذ نموذجين لإخفاء الصوت وهما:

النموذج الأول يعتمد على ضغط الكلام السري باستخدام تقنية سريعة وبسيطة مبنية على تقنية تحويل الموجات الراجعة والتي تتميز بسرعتها العالية، اما بالنسبة للنموذج الثاني فيستخدم برنامج ترميز الإنترنت (iLBC) لضغط الكلام السري مع التصحيح التلقائي لتدهور جودة الكلام المنتقل عبر قناة اتصال مدمجة بالضوضاء و فقد لبعض البيانات المرسله خلالها. في كلا النموذجين المقترحين ولأجل تلبية متطلبات الاتصال الصوتي عبر الأنترنت، يتم ضغط الموجة الصوتية المستخدمة كغلاف بواسطة مشفر G.711 في فترات زمنية متزامنة (لكل ١٦ أو ٢٠ ملي ثانية) مع موجة الصوت السرية المراد تضمينها لاحقاً. بعد ذلك، يتم تضمين بيانات موجة الصوت السرية المضغوطة داخل بيانات الغلاف المضغوط في الوقت الفعلي لكل فترة زمنية لما يناظرها.

في قناة اتصال عديمة الخسائر او الفقد للبيانات، تم الحفاظ على جودة الإشارات الناتجة بعد عملية التضمين بأكثر من ٤١ dB (للمنموذج المقترح الأول) و ٤٥ dB (للمنموذج المقترح الثاني) من حيث نسبة الإشارة إلى الضوضاء (SNR). أيضاً، تم تحقيق قدرة تضمين عالية تصل إلى ١٢,٥% و ٢٥% من حجم خطاب الغلاف للنموذجين المقترحين الأول والثاني. علاوة على ذلك، تم استرداد البيانات السرية بالكامل دون أخطاء. كذلك تم فحص متانة الطريقة المقترحة من حيث مناعتها ومقاومتها ضد وجود ضوضاء في قناة الاتصال وكذلك ضد وجود فقد جزء من الموجة بسبب عمليات التوجيه الخاطئة

في الشبكة. في قناة مدمجة بالضوضاء، أظهرت النتائج ان النهج المقترح يمكنه التغلب على الضوضاء بنسبة تصل الى أكثر من 35 dB من حيث SNR للنموذج الأول و 38 dB للنموذج الثاني. كما تم استرجاع البيانات السرية بدون فرق محسوس للأذن البشرية ، حيث لم تقل قيمة مقياس التشابه (NC) عن 0,92، في حين كانت قيمة معدل الخطأ في عدد البت المسترجعة (BER) أقل 0.0035 لكل من النموذجين المقترحين. عند نسب مختلفة من الخسائر او الفقد للبيانات الناتجة بعد عملية التضمين، تم الحفاظ على جودة الاشارات الناتجة والبيانات المخفية عند 0,88 و 35dB كمتوسط قيم من حيث NC و SNR على التوالي.



جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة كربلاء/كلية الهندسة
قسم الهندسة الكهربائية والإلكترونية

منظومه لأمن الاتصالات الصوتية في الوقت الفعلي اعتمادا على التضمين في قناة صوتية عامة

رسالة مقدمة الى

قسم الهندسة الكهربائية والإلكترونية في كلية الهندسة/ جامعة كربلاء
كجزء من متطلبات نيل شهادة الماجستير في علوم الهندسة الكهربائية

من قبل

بنين قاسم عبدعلي

(بكالوريوس في علوم الهندسة الكهربائية والإلكترونية ٢٠١٥)

تحت اشراف

دكتور حيدر اسماعيل اشهادي

دكتور مؤيد سليم كود

٢٠٢١/2022