Development of Multilayer New Covert Audio Cryptographic Model

Ziyad Tariq Mustafa Al-Ta'i, *senior member IACSIT*

***Abstract*— New covert cryptography has solved the problem of protecting the anonymity of sender and receiver, as well as the disadvantage of traditional hiding and steganography methods that once their method is known, any one can find embedded message. However, this paper presents multilayer new covert audio cryptography as a developed model in order to get more secrecy. The proposed model uses two covert layers, each layer with different (technique and cover). This model has the ability to cryptographically hide secret audio messages in audio cover using wavelet transform, at first layer. At the second layer, the audio stegocover is hided inside image cover using masking method. The extraction is done using simple cryptographic process. The proposed model has been successfully simulated between two nodes at the Internet.**

***Index Terms*— Information hiding, Masking, New covert cryptography, Wavelet transform.**

I. INTRODUCTION

Broadband communication networks and multimedia data available in a digital format (images, audio, video) opened many challenges and opportunities for innovation[1]. Speech is probably the most fundamental form of communication available to us and our society has become highly dependent on our fast and accurate means of transmitting spoken messages. Usually the aim of communicants is merely to transmit a message as quickly, accurately and cheaply as possible. There are, however, a number of situations where the information is confidential and where an interceptor might be able to benefit immensely from the knowledge gained by monitoring the information circuit. In such situations the communicants must take steps to conceal and protect the content of their spoken message. Of course, the amount of protection will vary. On occasions it is sufficient to prevent a casual listener from understanding the message but there are other times when it is crucial that even a determined interceptor must not be able to deduce it [2]. New covert cryptography is a different trend in cryptography field, because it has the features: (secrecy, covert, and simplicity). This new trend has appeared for two main reasons: first: cryptography has solved the problem of protecting privacy of information content, but it has not protected the anonymity of its sender and receiver. Second: to have a trusted system, one has to build it on his/her own, or the system must be simple enough so that one is able to check its correctness in implementation [3].

In the increasingly connected modern world, one may wish to be able to protect not only secrecy of the communication but also privacy of the communicators. Anonymous communication allows one to communicate without revealing who is communicating. Anonymous communication, the onset of computer technology and the Internet has given new life to information hiding and the creative methods with which it is employed [4]. Information hiding, in general, is covering sensitive information within normal information. This creates a hidden communication channel between the sender and receiver such that the existence of the channel is unnoticeable. Hidden channels have advantages over the encrypted channels that the anonymity of communication is protected [5]. Audio hiding, in particular, is a method for embedding information into an audio signal. It seeks to do so in a robust fashion, while not perceivably degrading the host signal (audio cover). To protect audio files against hacking , the researchers have ensured that the algorithm embeds bits of hidden information in deeper layers of the audio file and alters other bits to decrease the error [6].

In 1995, Naor and Shamir [7] opened the door to new covert cryptography in which cryptographic computation can be done without the use of a computer. Their research has mainly focused on guaranteeing privacy, and the (decryption) requires only primitive technology. Desmedt et al. [8] pointed out that traditional hiding and steganography methods have the disadvantage that once their method is known, any one can find embedded message. Therefore, they have been combined the concepts of new covert cryptography and information hiding to create new covert cryptographic models which are perfectly secure and whose (decryption\extracting) process involves primitive technologies only [9][10].

This paper gives a secure modified model for the simulation of new covert audio cryptographic model [11].

II. NEW COVERT CRYPTOGRAPHY

Since the speech or real pictures are much better means of communications between a human user and a cryptographic device, it is convenient to develop simple cryptosystems whose plaintexts are barely binary digits, but higher level languages such as audio or images. These cryptosystems are preferable to have the following properties:

- Secrecy: the information sent is protected from unwanted eyes.
- Covert: the existence of the secret channel is invisible to others.
- Simple: the decryption involves simple device only.

Therefore, these systems are called New Covert Cryptographic models [5]. One might question why one needs simple models. It is known that an encryption device can securely leaks its private keys to the network without

Manuscript received April 6, 2011.

Ziyad Tariq Mustafa Al-Ta'i is Head Department of Computer Science at College of Science, University of Diyala, Iraq, phone: 00964-17902267553, e-mail: ziyad1964tariq@yahoo.com

creating any trace. Thus a complex black-box system may not be trusted. To have a trusted system, one has to build it on his/her own, or the system must be simple enough so that one is able to check its implementation correctness [5].

New covert cryptographic models are previously described in different ideas by different authors such as:

1) Visual Cryptography

The idea of VC is fascinating invention by Noar and Shamir in 1995 [7]. VC is described as visual variant of the k-out of n secret sharing problem. VC is an encryption technique that does not need complex calculations in order to decrypt a message. The ciphertext and the key consist of transparencies. When properly stacked these transparencies reveal the plaintext.

The principle of VC is as follows. The input picture is divided into little small dots called pixels. The value black or white is assigned to every pixel, according to the original picture. The transparencies have again pixels, but now further divided into four subpixels. A fixed amount of these subpixels is colored black (e.g. two) (even if this pixel was white on the original picture). By careful choice of the position of the black subpixels, the difference between black and white is shown by stack the transparencies. (Black pixels have four black subpixels, white pixels have only two black subpixels, so they are grayish in reality) as shown in fig. (1).

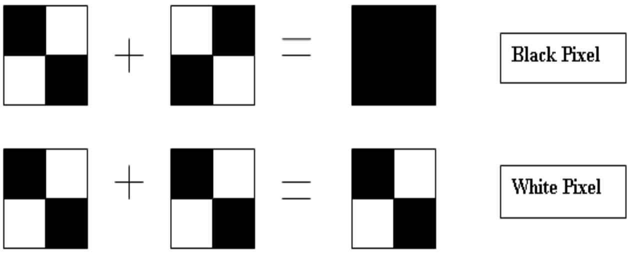

Fig.(1) Getting black and white pixel in visual cryptography

2) Cerebral Cryptography

Shuang Hou et al. [12] presented a new hiding model. This model uses pictures and secret sharing. Two pictures are used as in VC and each picture corresponds to a secret share of the message. This method differs from VC. First, while VC is a cryptosystem, this approach is both a cryptosystem as well as a hiding system. Images are not random, but correspond to real photographs randomly altered. This modification of the original is quite invisible to the naked eye. Secondly, decryption method does not use the subtractive properties of light. The decryption is done by brain using the perceived 3-D properties of the human visual system. A 3-D viewer is the only decryption hardware needed.

In Cerebral Cryptography model, the shares (ciphertext and the key) look like normal pictures. For each share, a subset of lines and columns are deleted from the picture so that when they are combined by looking through a 3-D viewer, a stereo-gram is obtained. The encrypted cleartext will then show up as square blocks, randomly up or down relative to the background. The model uses some interesting tricks to deceive the Human Visual System (HVS) in 3-D perception of some objects. The point is when an object is visible to one eye but the other, the HVS automatically assumes that the object is lower or higher in the scene, compared to those that are visible to both eyes at the same time as shown in fig. (2). Using this knowledge, if each share is seen by one eye separately (i.e. as done in a 3-D viewer), then one can control which portions are going to be seen by each eye. Consequently, one can trick the brain to an illusion that some blocks in the picture are up or down, while some others are normal. These up/down blocks are then used to encode the secret information.

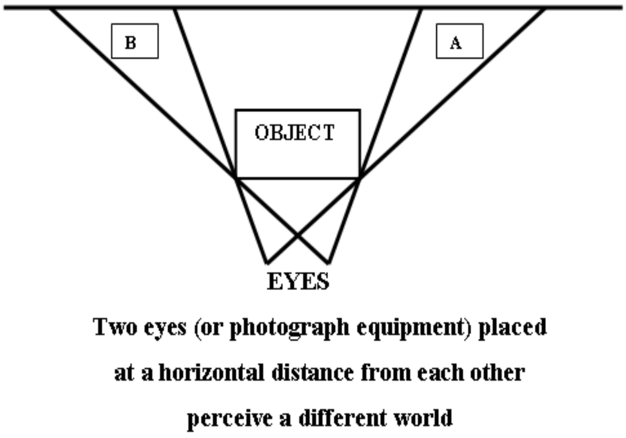
B
A
OBJECT
EYES
Two eyes (or photograph equipment) placed
at a horizontal distance from each other
perceive a different world

Fig.(2) Different views seen by two eyes in cerebral cryptography

3) Binary Audio Cryptography

Desmedt et al. [8] introduced new covert cryptography model, which is a secret sharing model that guarantees perfect privacy as well as high quality. In binary audio cryptography, the ciphertext and the key correspond to music whose phases have been changed. If the corresponding bit in cleartext is 1, then the two pieces of music in the two respective shares will be in phase, otherwise they will be out of phase. Hence, when heard together, a bit 1 in the cleartext will correspond to a constructive interference of the two shares, whose net effect is an increase in the amplitude of the combined music. And a bit 0 in the cleartext will then correspond to a destructive interference of the two shares. This results in a decrease of amplitude in the combined music. So the decryption method consists of playing the first share on speaker one and the second share on speaker two, and observing the increase and the decrease in amplitude.

4) Optical Cryptography

Desmedt et al. [8] gave another new covert cryptographic model which is based on the interference property of light wave. This model uses images to hide information. This approach is completely different from the one used to obtain visual cryptography. As in cerebral cryptography, the shares are not suspicious. The privacy of the model is perfect and the stego images are of high quality. Using a Mach-Zehnder interferometer on two shares, the embedded image can be seen. The model has advantage of providing larger bandwidth over Cerebral Cryptography. The idea of Optical Cryptography is as follows. The plaintext is 1 bit/pixel digital image. A high quality n bit/pixel image is chosen, which has a larger size than that of plaintext. The plaintext is padded to make it the same size of the cover image. The share1 is generated by randomly flipping the least (mth) significant bit of each pixel in the cover image. Share1 is copied to share2 as its initial value. Then if in plaintext a pixel has the value1 then the least (mth) significant bits of this corresponding pixel in share2 are flipped. So, now the (mth) significant bits in the generated shares are uniformly random bits. The two shares only differ in the least (mth) significant bits. Now a machine called Mach-Zehnder interferometer can be used to reconstruct the plaintext.

5) *Non Binary Audio Cryptography*

Quisquater et al. [9] presented Non Binary audio Cryptography model to overcome the problems of Binary audio cryptography model. These problems are: non user friendly, low bandwidth, or untested. In this model the plaintext can be speech, or any other audio signal. This approach guarantees perfect secrecy by introducing variations of the one time pad. The ciphertext is non suspicious, since it sounds with high quality as normal music. The goal of this model is to seek for some hiding place in the human auditory system to hide another signal. One of the good candidates for such a place is the masking effect happening in the human brain. The decryption method of Binary Audio Cryptography has been tested with this model and it did not work. There are two problems. First, the mixing obtained is not good enough. In order to solve this problem an inexpensive audio mixer suffices. Secondly, even by using an audio mixer, one is still unable to hear the message. This is solved using an (old fashioned) amplifier.

6) *Moire' Cryptography*

Desmedt and Tri [10] applied concepts of steganography to create secret sharing models whose shares are realistically looking images. This new technique is based on an idea of employing Moire' patterns for producing images. The advantage of this model over others is that it does not require a complicated algorithm, thus a computer, to decrypt the ciphertext. The cleartext can be read simply by putting the ciphertexts one onto the other. Therefore, a solution for random nature of secret shares in VC is given, with a novel type of visual secret sharing models, whose secrecy and anonymity are both satisfied. The main idea of this model is as follows. When two transparencies are stacked together, the result is an (and) operation of the two transparencies. Unfortunately, this does not provide a group operation on the set {0,1}, thus it can not provide perfect secrecy. VC overcomes this by encoding 0 and 1 with random black-white matrices of different averaged gray levels. In this approach, to encode a 1 bit, two squares are superimposed on the two shares whose dots are oriented with different angles. To encode a 0 bit, two small squares are superimposed on the two shares whose dots are oriented by same angle. Hence, the resulting picture appears with one Moire' pattern that forms the embedded picture; not the gray level of the squares as done in VC. The decryption process is relatively simple. The two transparencies are stacked onto each other to create Moire' patterns.

III. THE PROPOSED MODEL

A.. *Primitive Models*

The primitive models for this paper are shown in fig. (3) [9], and fig. (4). These primitive models are depended on two weak points in Human Auditory System (HAS)and Human Visual System (HVS) , which are the masking effect and the phase shift effect.

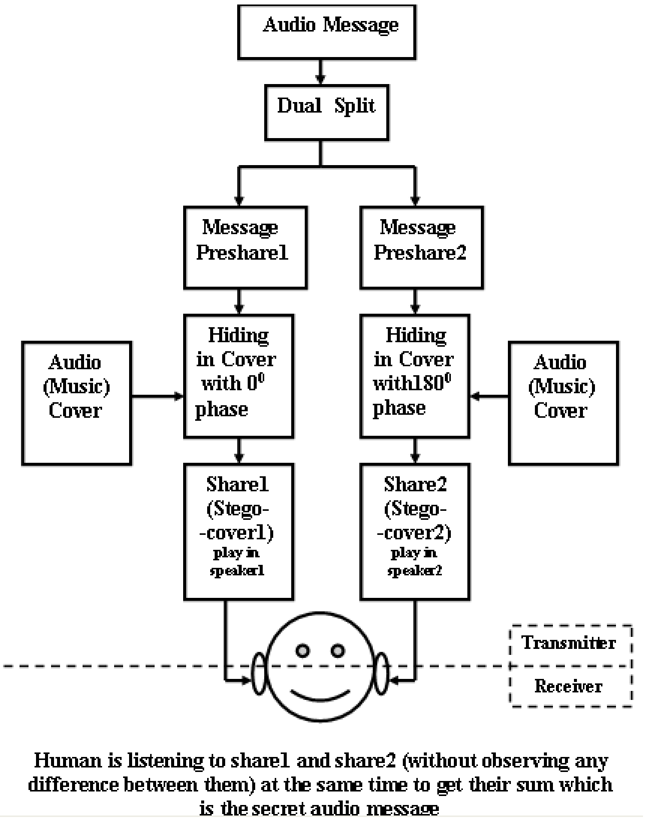
Audio Message
Dual Split
Message Preshare1
Message Preshare2
Audio (Music) Cover
Hiding in Cover with 0⁰ phase
Hiding in Cover with180⁰ phase
Audio (Music) Cover
Share1 (Stego- -cover1) play in speaker1
Share2 (Stego- -cover2) play in speaker2
Transmitter
Receiver
Human is listening to share1 and share2 (without observing any difference between them) at the same time to get their sum which is the secret audio message

Fig.(3) First primitive model

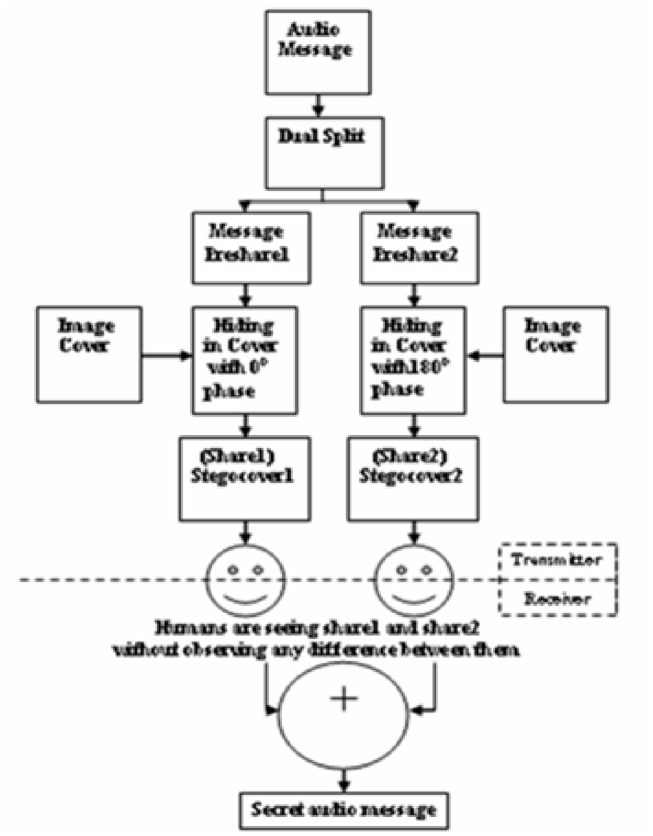

Fig.(4) Second primitive model

B. *Implementation of the Proposed Model*

Two covert layers are used with proposed model. Each layer with different (technique and cover) type as shown in fig.(5). Two sides are contained in the proposed model as described as follows.

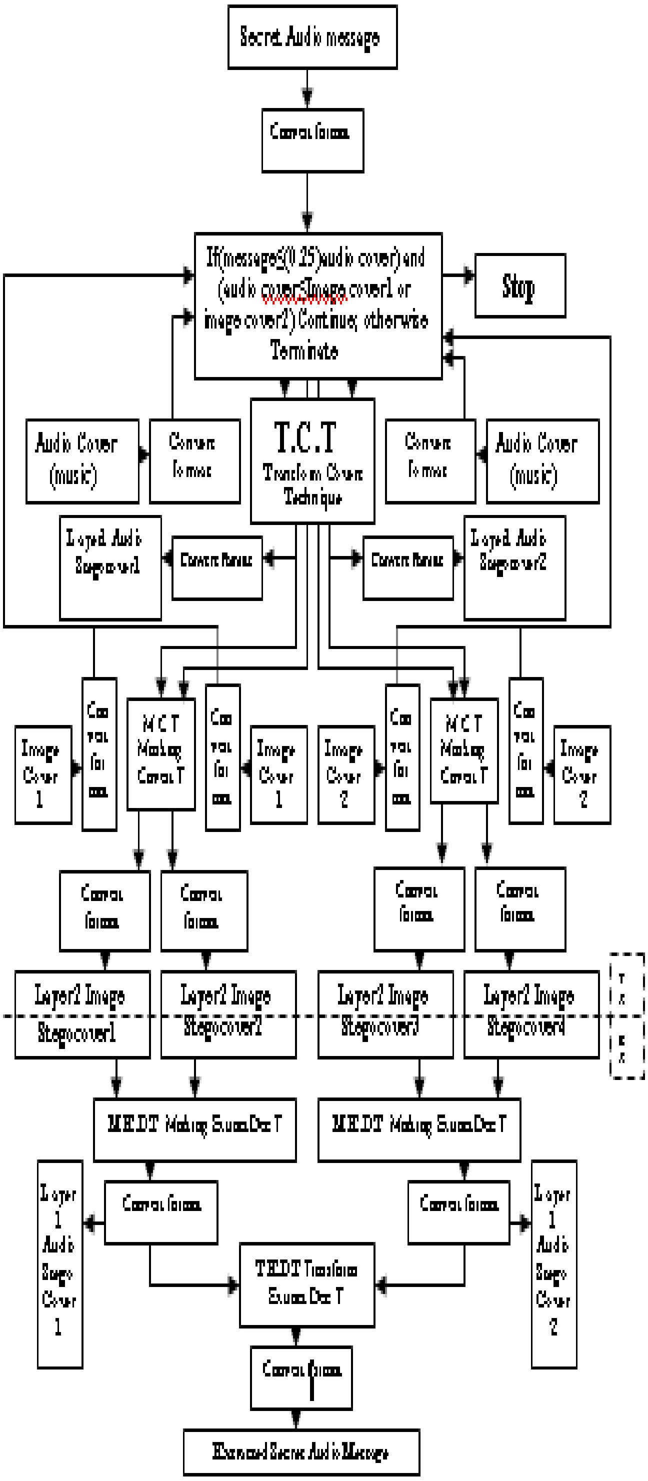

Fig.(5) Block diagram of Multilayer New Covert Audio Cryptographic Model

1) Transmitter Side

On transmitter side of this model, there are two layers. First of all, comparison process must be done to check if the secret audio message size is smaller than or equal to (0.25) of the audio cover size; Also to check if the audio cover is smaller than or equal to image cover; otherwise, termination is accomplished. At layer1, the secret audio message is secretly shared into two preshares. These two preshares are wavelet transform hidden in audio cover with (0^0 and 180^0) phase shifts, in order to produce layer1 stegocover1 and stegocover2. This process is dependent on Transform Covert Technique (T.C.T).

T.C.T:

The transform covert technique can be shown in fig.(6). The main elements of T.C.T can be described as follows.

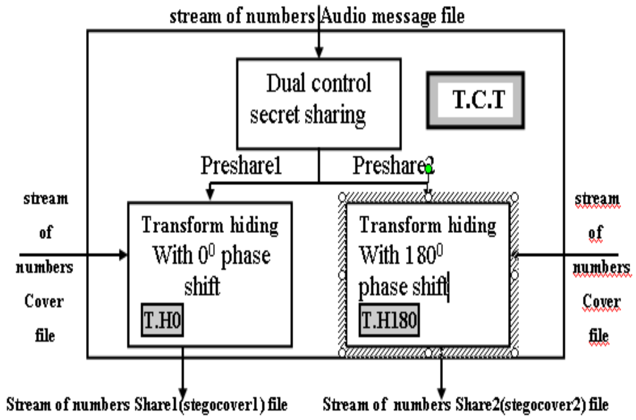
stream of numbers Audio message file
Dual control secret sharing
T.C.T
Preshare1
Preshare2
stream of numbers Cover file
Transform hiding With 0⁰ phase shift
T.H0
Transform hiding With 180⁰ phase shift
T.H180
stream of numbers Cover file
Stream of numbers Share1(stegocover1) file
Stream of numbers Share2(stegocover2) file

Fig.(6) Block diagram of Transform Covert Technique (T.C.T)

Wavelet transform is the first step in (T.C.T). This technique is based on normalized wavelet transform, which is applied to the values of audio cover. Depending on equations (1 and 2):

$$\Psi_{sb}(t) = (1/|s|^{0.5}) \, \Psi((t-b)/s) \qquad ……(1)$$

$$F(s, b) = \int f(t) \, \Psi((t-b)/s) \, dt \qquad ……(2)$$

Ψ (t) is the mother wavelet, b represents a time shift, and s is a scaling factor used with t, time. f is analogue signal, F is transformed signal.

Replacing cover coefficients with secret audio message is the second step in (T.C.T). After employing wavelet transform on cover file, the high energy elements are clustered at certain positions with each transformed window (block). Therefore, some coefficients with low energy can be discarded from each block without distorting the reconstructed stegocover. The principal idea of this technique is done by discarding low energy coefficients using Zonal Sampling method [8], which depends on discarding the elements that have small variances and keeping the elements that have large variances. The number of discarded coefficients depends on the type of transform used and the discarding method.

Inverse wavelet transform is the third step in (T.C.T) as shown in figure (2). This technique is based on normalized inverse wavelet transform, which is applied to the wavelet transformed (audio or image) cover with scaled secret audio message coefficients depending on equation (3).

$$f(t) = \iint F(s,b) \, \Psi((t-b)/s) \, db ds \qquad ……(3)$$

At layer2, each layer1 stegocover is considered as secret audio message that needs to be covered. Therefore, the layer1 stegocover1and stegocover2 are secretly divided into two preshares. These preshares are masked by image cover with

(0^0 and 180^0) phase shifts, in order to produce layer2 stegocover1, stegocover2, stegocover3, and stegocover4 using Masking Covert Technique (M.C.T).

M.C.T.:

The principle of this technique is Temporal Masking. Masking hiding is, usually, audio watermarking technique [9]. However, masking is used as a steganographic technique in this paper. The masking technique is implemented by quieting the layer1 stegocover1 in comparison with image cover file using equation (4).

$$\text{Stegocover}=k\times m+(1-k)\times c$$

$$k=\text{masking factor } (0.01 \rightarrow 1)$$

$$c=\text{Image cover}$$

$$m=\text{secret object}=\text{stegocover} \qquad \ldots(4)$$

2) Receiver Side

At receiver side, there are also two layers. At layer1, the image stegocover1, stegocover2, stegocover3, and stegocover4 are converted and mod-added in order to decrypt and extract the values of audio stegocover1 and stegocover2 using Masking Decrypting Extracting Technique (M.D.E.T).

M.D.E.T:

This technique is shown in fig. (7). The main (simple) process in (M.E.D.T) is mod-addition technique, in which the cover values in stegocover1 and stegocover2 cancel each other by mod-addition process, while preshare1 values and preshare2 values reconstruct stream of values for secret audio message.

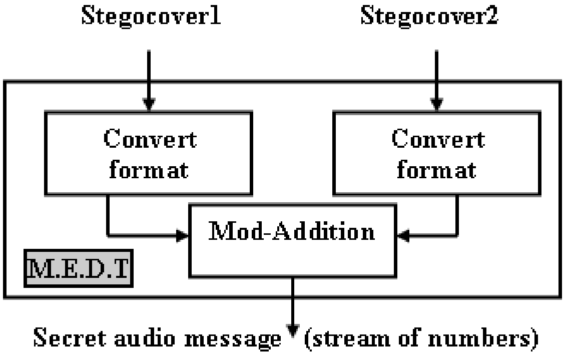

Fig.(7) Block diagram of Masking Decrypting Extracting Technique (M.D.E.T)

At layer2, the layer1 audio stegocover1 and stegocover2 are (converted, wavelet transformed, and added) in order to decrypt and extract stream of values of secret audio message using Transform Decrypting Extracting Technique (T.D.E.T). This technique is shown in fig.(8).

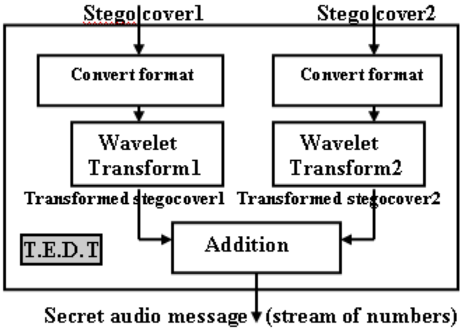

Fig.(8) Block diagram of Transform Extracting Decrypting Technique (T.E.D.T)

IV. RESULTS

The digital representation of recorded audio samples is Windows Audio Video (WAV) format, with following attributes: 8-bit, sampling rate (11.025 KHz), stereo. The test sample of secret audio message is with (7 sec) length and (141 KB) size. The test sample of audio cover is with (76 sec) length and (1.58 MB) size. The digital representation of image cover sample is Tag Image File (TIF) format, with true color, 8-bit. The image cover1 size is (2.2 MB with dimension 1129×667) and The image cover2 size is (2.1 MB with dimension 735×985).

Results:

1) Fidelity of layer1 audio stegocover1:
   Square Mean Square Error Ratio=0.0036%.
   Mean Square Signal to Noise Ratio in dB=46.558.
2) Fidelity of layer1 audio stegocover2:
   Square Mean Square Error Ratio=0.1782%.
   Mean Square Signal to Noise Ratio in dB=12.711.
3) Fidelity of layer2 image stegocover1:
   Square Mean Square Error Ratio=0.0218%.
   Mean Square Signal to Noise Ratio in dB=30.0269.
4) Fidelity of layer2 image stegocover2:
   Square Mean Square Error Ratio=0.0425%.
   Mean Square Signal to Noise Ratio in dB=24.0081.
5) Fidelity of layer2 image stegocover3:
   Square Mean Square Error Ratio=0.0249%.
   Mean Square Signal to Noise Ratio in dB=29.5176.
6) Fidelity of layer2 image stegocover4:
   Square Mean Square Error Ratio=0.0491%.
   Mean Square Signal to Noise Ratio in dB=23.3579.
7) Fidelity of extracted audio cover1:
   Square Mean Error Ratio=0.0812%.
   Mean Square Signal to Noise Ratio in dB=19.428.
8) Fidelity of extracted audio cover2:
   Square Mean Error Ratio=0.1796%.
   Mean Square Signal to Noise Ratio in dB=12.6158.
9) Fidelity of extracted audio message:
   Square Mean Error Ratio=0.4917%.
   Mean Square Signal to Noise Ratio in dB=25.04789.

The comparison between the original image cover and layer2 image stegocovers is shown in fig.s (9) and (10). The difference between original music cover and audio stegocovers, and extracted audio stegocovers is shown in fig. (11). However, the difference between the original secret speech message and the extracted secret speech message is shown in figure (12).

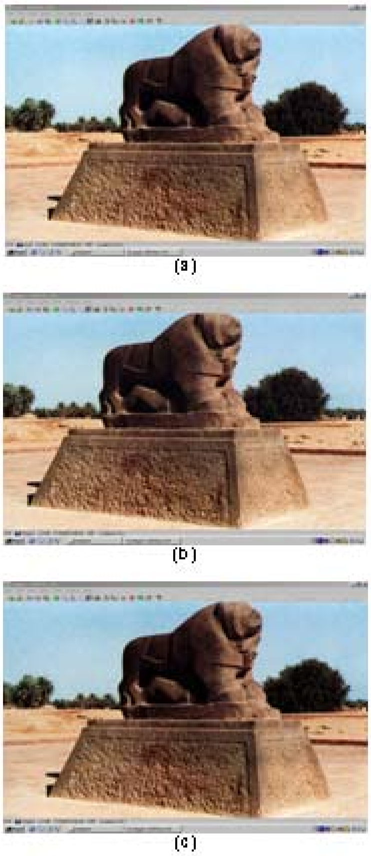

Fig.(9) Comparison among Images of (a)Original cover1 (b) Stegocover1 (c) Stegocover2

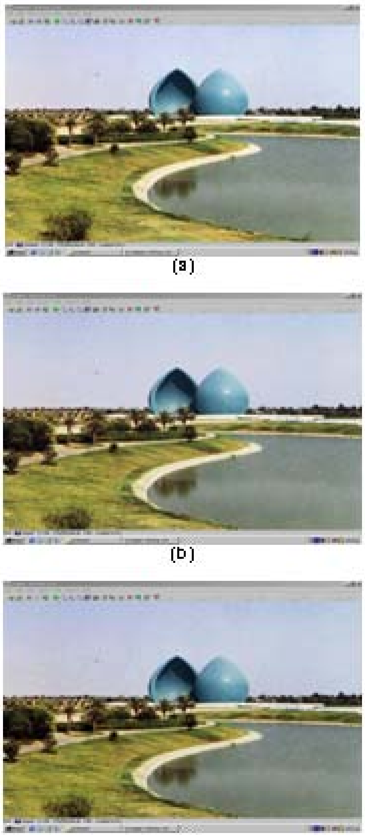

Fig.(10) Comparison among Images of (a)Original cover2 (b) Stegocover3 (c) Stegocover4

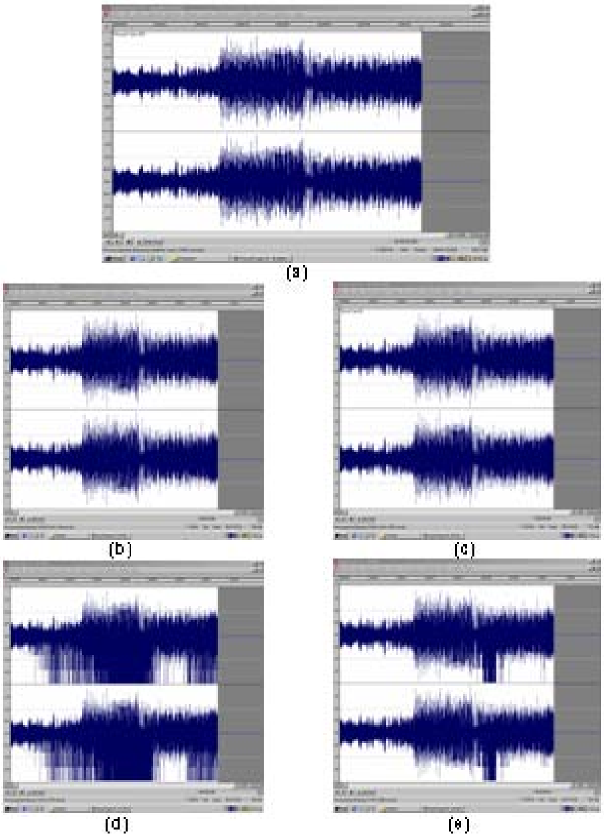

Fig.(11) Comparison among left right waveforms of (a)Original music cover (b) Stegocover1 (c) Stegocover2 (d) Extracted stegocover1 (e) Extracted stegocover2

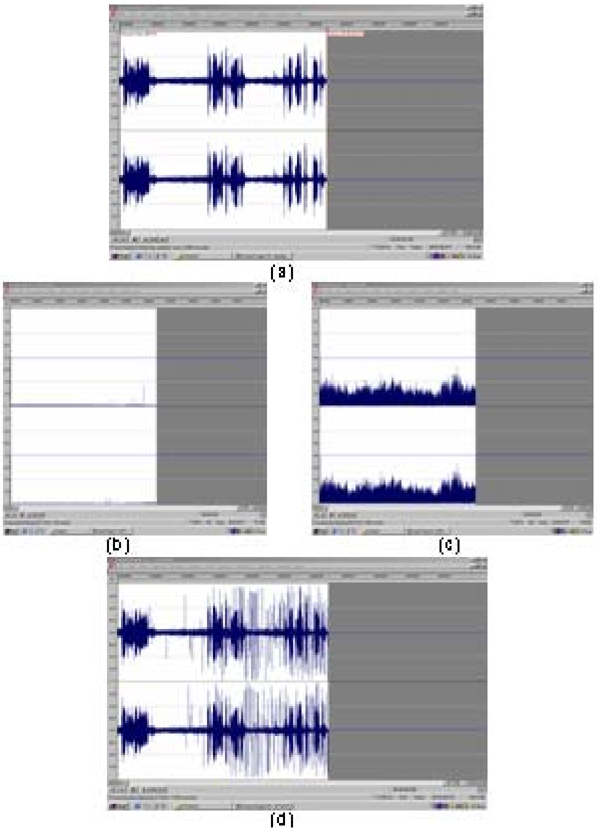

Fig.(12) Comparison among left right waveforms of (a)Original secret speech message (b) Hidden preshare1 in layer1stegocover1 (c) Hidden preshare2 in layer1stegocover2 (d) Extracted secret speech message

V. CONCLUSION

New covert cryptography is mixing between cryptography and information hiding with simple hardware (decryptor\extractor). However, the proposed model may be used successfully as new covert audio cryptographic system, by software means. The developed model uses the transform and masking techniques in two sequenced layers as a new idea. The proposed model has been successfully simulated between two nodes at the Internet, with understandable extracted audio message. Two important factors that must carefully be selected, masking factor at masking layer and scaling factor at transform layer. These factors are considered as a control factors. The secrecy of the developed model is very high.

REFERENCES

[1] N. Cvejic, “Algorithms for Audio Watermarking and Steganography”, Academic Dissertation, the Faculty of Technology, University of Oulu , Finland , 2004, Abstract.

[2] H.J. Beker, "Analogue Speech Security System", Proceedings of workshop on Cryptography, Lecture Notes in Computer Science 149, Springer-Verlag, Burg Feuerstein, Germany, March 29 – April 2, 1982, PP:130-146 .

[3] A.P. Fabien, J. A. Ross and G. K. Markus "Information Hiding – A Survey", Proceedings of the IEEE, Special Issue on Multimedia, 87(7), July 1999, PP:1062-1078.

[4] F. N. Johnson, D. Zoran and J. Sushil "Information Hiding: Steganography and Watermarking - Attacks and Countermeasures", Kluwer Academic Publishers, Advances in Information Security, 2001, ch. 1.

[5] Tri Van Le, "Covert Cryptography", MS.C Thesis, The University of Wisconsin_Milwaukee, August 1999, Abstract .

[6] Frost & Sullivan. (April 2010). “Audio Steganography”. Hi-Tech Security Solutions, The Journal for Security, Operations & Risk Management [On Line] , Available: http://www.technews.co.za.

[7] M. Naor, and A. Shamir, "Visual Cryptography", Proceedings of International Conference EUROCRYPT'94, Springer-Verlag, 1995, PP:1-12 .

[8] Y. Desmedt, H. Shuang and J. Q. Jean "Audio and Optical Cryptography", Proceedings of International Conference on the Theory and Application of Cryptology and Information Security, China , October 1998, PP:392-404.

[9] J.J. Quisquater, Y. Desmedt and L. V. Tri "Non Binary Audio Cryptography", Proceedings of 3rd International Workshop of Information Hiding, Dresden, Germany, 1999, PP: 478-489.

[10] Y. Desmedt and L. V. Tri "Moire Cryptography", Proceedings of the 7th ACM Conference on Computer and Communications Security, Athens, Greece, 2000, PP: 116-124.

[11] Ziyad T. M. Al-Ta’i, “Simulation of new covert audio cryptographic model”, Proceedings of the 3rd International Conference on Machine Learning and Computing (ICMLC), Singapore, February 26-28, 2011, VIP-ICMLC-C00386-001, to be published..

[12] H. Shuang, Y. Desmedt and J.J. Quisquater "Cerebral Cryptography", Proceedings of 2nd International workshop of Information Hiding (IH'98), Lecture Notes on Computer Science, Volume 1525/1998, USA, April 14-17, 1998, PP: 62-72.

Ziyad T. M. Al-Ta’i was born in Baghdad (1964). Received BS.C degree in Electrical Engineering from University of Baghdad (1987) and MS.C degree in Computer Science from University of Technology – Baghdad (1995), and received Ph.D. in Computer Science from University of Technology – Baghdad (2002). Field of specialization is Network Security and Information Hiding.

He is Assistant Professor (2007) at University of Diyala–Iraq as Head of Department Computer Science. He has published (12) journal national papers, (5) papers in national conferences, (1) journal international paper, and (1) paper in international conference.

Assis. Prof. Al-Ta’i is a senior member of IACSIT , Member of Iraqi Engineers Union , Member of Iraqi Teachers Union, and Member of Iraqi Association for Information Technology.