Voice Security Using Hybrid Algorithm

Alyaa Moufaq Abdul Majeed Haleem
AlyaaHaleem@uomosul.edu.iq
College of Computers Sciences and Mathematics
University of Mosul, Iraq

Received on: 04/11/2009 **Accepted on: 16/05/2010**

ABSTRACT

This research deals with constructing and implementing a new digital voice security Algorithm based on hiding large amount of data (sound file) in a 24 bits host color image (RGB image). The proposed method starts with speech compression to convert human speech into an efficiently encoded representation that can later be decoded to produce a close approximation of the original signal. The process of compression is achieved by first computing Discrete Wavelet Transform (DWT), truncating small-valued coefficients and then efficiently encoding them. The stream bits output from coder are encrypted using Linear Feedback Shift Register (LFSR) algorithm. These enciphered bits are then embedded into the image blocks. A binary key matrix and weight matrix are used as a secret key to protect the hidden information. The algorithm can hide as many as $(\log_2(M \times N \times 12+1))$ bits of data in the image by changing one bit in each block of size $(M \times N)$. High security algorithm was achieved using three layers to make it difficult to break by attacker. The algorithm has been implemented using MATLAB.

Keywords: DWT, LFSR, Voice, RGB, Correlation Coefficient.

أمنية الصوت باستخدام خوارزمية هجينة

علياء موفق عبد المجيد حليم

كلية علوم الحاسوب والرياضيات، جامعة الموصل

تاريخ الاستلام: 2009/11/04 **تاريخ القبول: 2010/05/16**

الملخص

يتناول البحث بناء وتطبيق نظام جديد لسرية الصوت الرقمي يعتمد على إخفاء كمية كبيرة من البيانات (ملفات صوت) داخل الصور الملونة ذات التمثيل 24 بت (RGB). الطريقة المقترحة تبدأ بضغط عينات الكلام لتحويل كلام الإنسان إلى التمثيل المرمز والتي يمكن فك ترميزها في وقت لاحق لاسترجاع القيمة التقريبية للإشارة الأصلية. عملية الكبس تمت باستخدام خوارزمية التحويل المويجي المتقطع وقطع المعاملات ذات القيم الصغيرة ومن ثم استخدام طريقة ترميز المعاملات التي قيمها مساوية للصفر. يتم بعد ذلك تشفير سلسلة bits الناتجة من العملية السابقة باستخدام خوارزمية LFSR. ثم تطمر bits المشفرة داخل كتل الصورة حيث يتم استخدام مصفوفة المفتاح الثنائية و مصفوفة الوزن كمفتاح سري لحماية المعلومات المخفية. تتمكن الخوارزمية من إخفاء $(\log_2(M \times N \times 12+1))$من bits في الصورة من خلال تغير bit واحد داخل كل كتلة من كتل الصورة التي بحجم M×N. تم تحقيق مستوى عالي من السرية باستخدام ثلاث مستويات لجعل عملية كسر النظام من قبل المهاجمين أكثر صعوبة. تم تنفيذ الخوارزمية باستخدام لغة ماتلاب.

الكلمات المفتاحية : التحويل المويجي المتقطع .الصوت . سجلات الازاحة الخطية . معاملات الارتباط

1. Introduction

Digital speech communication has been applied in many fields. But communication between two parties over long distances has always been subject to

interception. This led to the development of cryptography schemes. Cryptography schemes achieve security mainly through a process of making the speech unintelligible so that those who do not possess necessary keys cannot recover the speech. Though cryptography can hide the content of the speech, the existence of a cryptographic communication in progress cannot be hidden from a third party. If the third party discovers the cryptographic communication, he/she might be able to decipher the speech. It can be seen that latent danger exists in cryptography schemes. The need to avoid this led to the development of steganography schemes which compensate cryptography by hiding the existence of a secret communication[4]. The secure speech transmission system based on steganography by embedding a secret speech file in a cover medium has been increasingly gaining importance in the field of information technology[2].

Generally Steganography is the art and science of writing hidden messages in such a way that no one apart from the intended recipient knows of the existence of the message; this is in contrast to cryptography, where the existence of the message itself is not disguised, but the content is obscured. Cryptography hides the contents of a secret message from an attacker, whereas steganography even conceals the existence of this message. Therefore the definition of breaking the system is different. In cryptography, the system is broken when the attacker can read the secret message. Breaking a steganography system has two stages first the attacker can detect that steganography has been used. Additionally, he is able to read the embedded message [6][8].

Cryptographic techniques scramble a message so that if it is intercepted, it cannot be understood. This process is encryption and the encrypted message is sometimes referred to as cipher text. Steganography, in essence,"camouflages" a message to hide its existence and make it seem "invisible"thus concealing the fact that a message is being sent altogether. A cipher text message may draw suspicion while invisible message will not. [7][5]

In this research both sciences can be combined for better protection of information.

2. Hiding a message inside color images

Hiding information inside images is a popular technique nowadays. An image with a secret message inside may be easily spread over the World Wide Web or in newsgroups. The most widely used technique to hide data is the usage of the Least-significant bit (LSB). Although there are several disadvantages to this approach, the relative easiness to implement it, makes it a popular method. To hide a secret message inside an image, a proper cover image is needed. Because this method change the LSB bits of each pixel in the image, it is necessary to use a lossless compression format, otherwise the hidden information will get lost in the transformations of a lossy compression algorithm[6].

There are several disadvantages to a LSB strategy. The first is that it is easily recognizable by image analysis. The signature of the embedded text can be recognized and thus does not provide a safe cover for sensitive or copyright marks. Another disadvantage to embedding information inside other data is that lossless compression algorithms and formats such as jpeg will destroy the required structure to recover embedded information [9].

The other most common methods using a 24-bit color image, a bit of each of the red, green and blue color components can be used, so a total of 3 bits can be stored in each pixel. Thus, an 800*600 pixel image can contain a total amount of 1.440.000 bits

(180.000 bytes) of secret data. However, using this method may be lead to changes will be noticeable using statistical analysis against the different areas of the image and causes to distorted the image [7].

In this research to ameliorate the image hiding quality and hiding capacity our modification algorithm is capable of hiding large amount of data by changing a small number of bits in the original binary image. The modified method used 24-bit color image and partitioned it into blocks of size $(M \times N)$ and used the first 4 bit of the red, green and blue color components, so that the total size used is $(M \times N \times 12)$ bits. In this block size can conceal as many as $\left(\log_2(M \times N \times 12+1)\right)$ bits of data by changing only one bit of this block. This algorithm is more effective than the traditional methods (LSB), that can hide one bit by changing one bit in block.

3. Layers of Hybrid algorithm

New steganography algorithm using three layers of security has been constructed. These layers are developed to acquire high security. These layers work independently to provide unbreakable security as show in Figure (1).

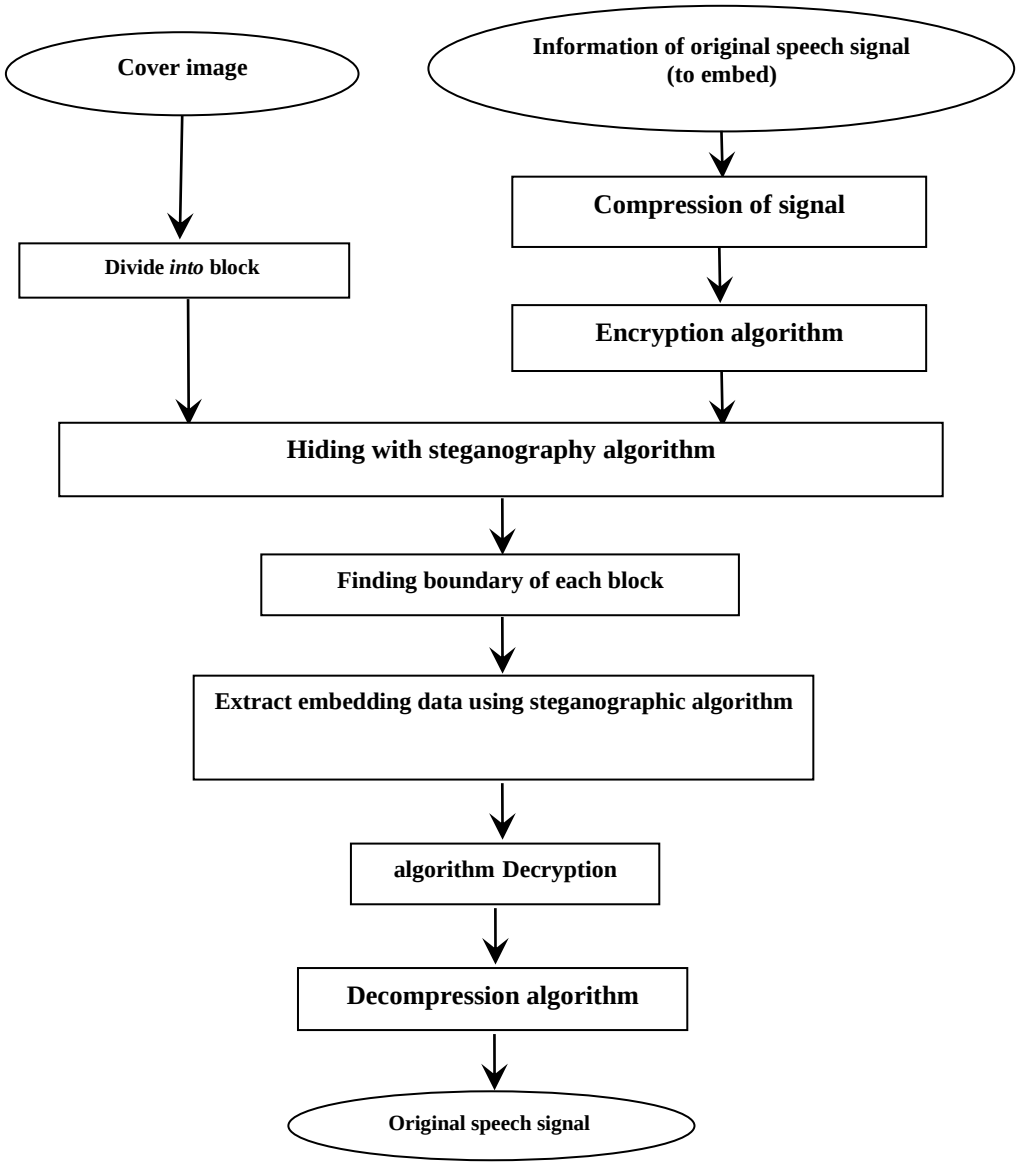


Figure 1: steganography with security layers

The compression mechanism is the first layer of security using Discrete Wavelet Transform (DWT) Technique and run length method. Then encryption the coding speech signal using stream cipher algorithm. these two layers are used before hiding input speech signal.

3.1 Compression Layer

The idea behind speech compression is to encode audio data to take up less storage space and less band width for transmission. To meet this goal we used Fast Wavelet Transform (FWT). Wavelets concentrate speech information (energy and perception) into a few neighboring coefficients [3]. Therefore as a result of taking the wavelet transform of a signal, many coefficients will either be zero or have negligible magnitudes. Data compression is then achieved by treating small valued coefficients as insignificant data and thus discarding them.

3.1.1 The Fast Wavelet Transform (FWT) Algorithm

The Discrete Wavelet Transform (DWT) coefficients can be computed by using Fast Wavelet Transform algorithm considers the following equation:

$$\emptyset(t) = \sum_k c(k)\emptyset(2t-k) \quad \ldots(1)$$

$$\boldsymbol{\varphi(t) = \sum_k (-1)^k c(1-k)\emptyset(2t-k)} \quad \ldots(2)$$

$$\boldsymbol{\sum_k c_k c_{k-2m} = 2\partial_{0,m}} \quad \ldots(3)$$

The first equation is known as the *twin-scale relation* (or the dilation equation) and defines the scaling function $\emptyset$. The next equation expresses the wavelet φ in terms of the scaling function $\emptyset$. The third equation is the condition required for the wavelet to be orthogonal to the scaling function and it translates.

The coefficients c(k) or $\{c_0, \ldots\ldots, c_{2N-1}\}$ in the above equations represent the impulse response coefficients for a low pass filter of length 2N, with a sum of 1 and a norm of $\frac{1}{\sqrt{2}}$.

The high pass filter is obtained from the low pass filter using the relationship $g_k = (-1)^k c(1-k)$, where k varies over the range (1-(2N-1)) to 1.

Equation (1) shows that the scaling function is essentially a low pass filter and is used to define the approximations. The wavelet function defined by equation (2) is a high pass filter and defines the details.

Given an input speech signal s of length N as shown in Figure(2), the DWT consists of $\log_2 N$ stages at most. The first step produces, starting from s, two sets of coefficients: approximation coefficients cA1, and detail coefficients cD1. These vectors are obtained by convolving s with the low-pass filter Lo_D for approximation, and with the High-pass filter Hi_D for detail, followed by dyadic decimation or down sampling by a factor of 2. As shown in the figure below[3][4].

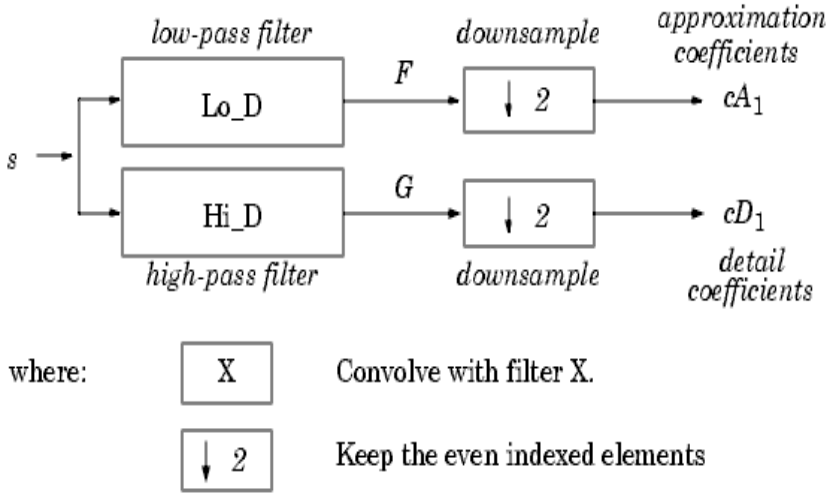


Figure 2: Filtering operation of the DWT

The length of each filter is equal to 2N. If n = length (s), the signals F and G are of length (N+ 2n – 1), and the coefficients cA1 and cD1 are of length $floor\left(\frac{n-1}{2}\right) + N$. The next step splits the approximation coefficients cA1 in two parts using the same scheme, replacing s by cA1 and producing cA2 and cD2, and so on. So the wavelet decomposition of the signal s analyzed at level j has the structure shown in Figure (3).

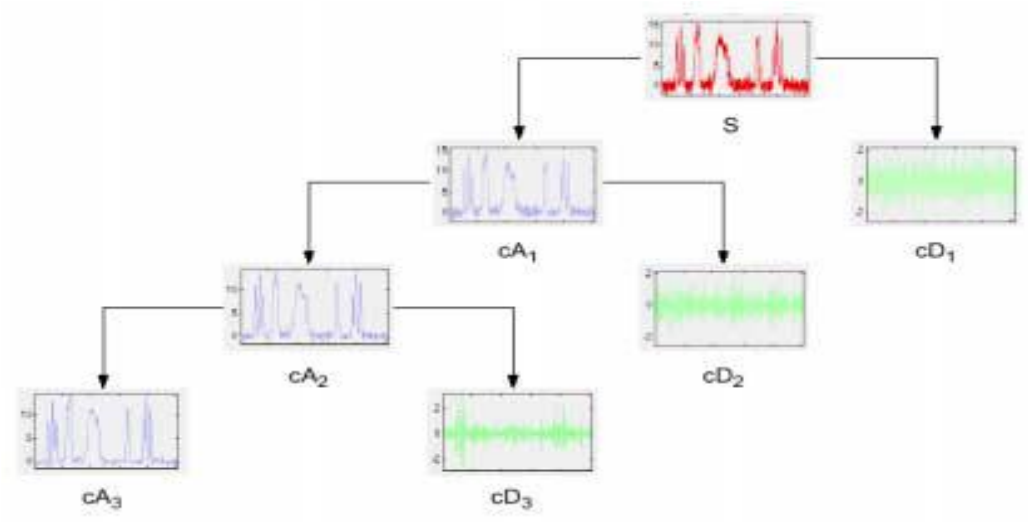


Figure 3: 3-level Decomposition of Signal S

After calculating the wavelet transform of the speech signal, compression is achieved by **first truncation wavelet coefficients** below a threshold. An experiment conducted on a male spoken sentence, shows that most of the coefficients have small magnitudes. More than 90% of the wavelet coefficients have less than 5% of the maximum value.This means that most of the speech energy is in the high-valued coefficients [3]. Thus the small valued coefficients can be truncated or zeroed. **Secondly, encode** consecutive zero valued coefficients with two bytes using Run length encoding method. One byte to indicate a sequence of zeros in the wavelet transforms vector and the second byte representing the number of consecutive zeros. Figure (4) show the flowchart of compression algorithm.

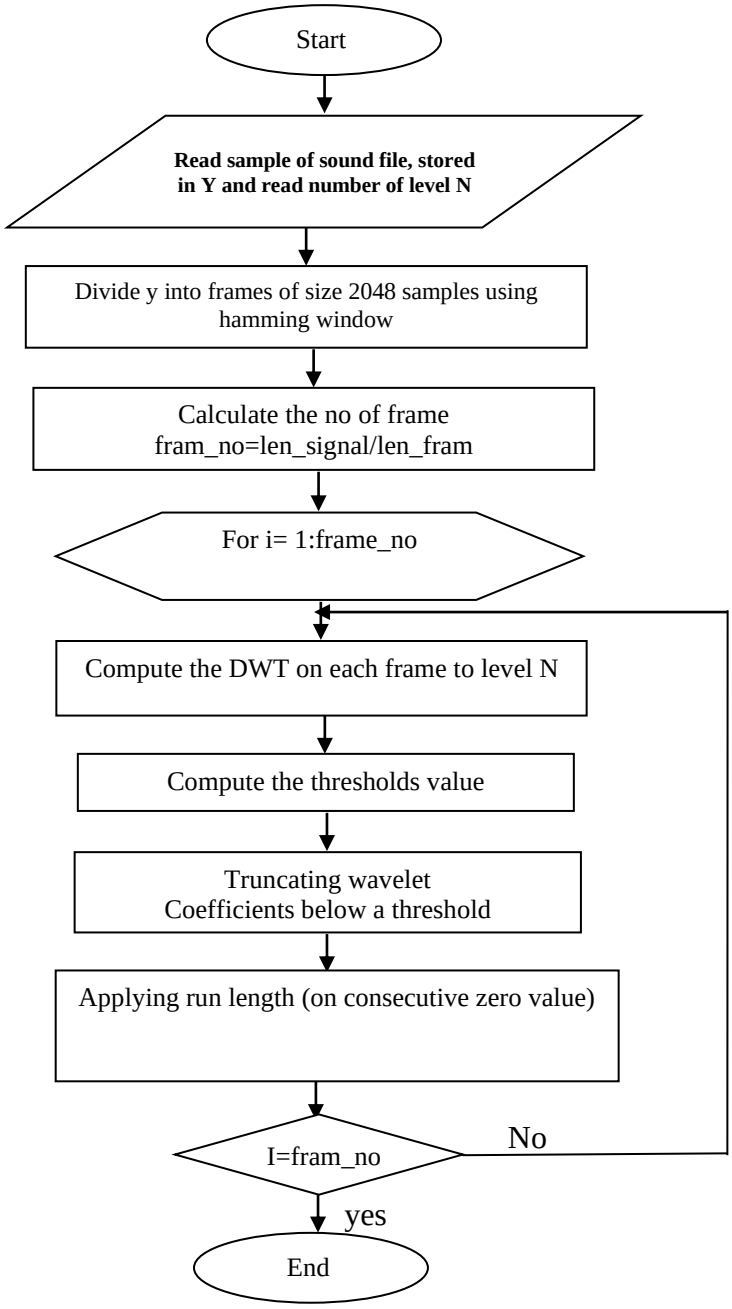


Figure 4: Flow chart of compression Speech

3.2 Encryption Layer

The Linear Feedback Shift Register (LFSR) has been one of the most popular encryption techniques widely used in speech communication. LFSR is suitable for speech because speech is continuous streaming data. They encrypt individual character (usually binary digits) of a plaintext message one at a time, using an encryption transformation which varies with time. Stream cipher which used LFSR is algorithm that encrypts plaintext one bit at a time[10]. Key stream generator generates outputs stream of bits k_1, k_2, ..., k_n. Cipher text is obtained by XOR this key stream bits with plain text bits p_1, p_2, ..., p_n.

Generally , the length of the sequence before repetition occurs depends upon two things, the feedback taps and the initial state. An LFSR of any given size m (number of registers) is capable of producing every possible state during the period $N=2^m-1$, but will do so only if proper feedback taps, or terms, have been chosen. Such a sequence is called a *maximal length sequence*, *maximal sequence*, or less commonly, *maximum length sequence*. It is often abbreviated as *m-sequence*. In certain industries m-sequences are referred to as a *pseudonoise* (PN) or *pseudorandom* sequences.

3.2.1 M-Sequence Properties

The Properties of m-sequences include the following:

1. An m-bit register produces an m-sequence of period 2^m-1.
2. An m-sequence contains exactly $2^{(m-1)}$ ones and $2^{(m-1)}-1$ zeros.
3. The modulo-2 sum of an m-sequence and another phase of the same sequence yields yet a third phase of the sequence.

3a. (A corollary of 3.) Each node of an m-sequence generator runs through some phase of the sequence. (While this is obvious with a Fibonacci LFSR, it may not be with a Galois LFSR.)

4. A sliding window of length m, passed along an m-sequence for 2^m-1 positions, will span every possible m-bit number, except all zeros, once and only once.
5. Define a run of length r to be a sequence of r consecutive identical numbers, bracketed by non-equal numbers. Then in any m-sequence there are:
   1 run of ones of length m.
   1 run of zeros of length m-1.
   1 run of ones and 1 run of zeros, each of length m-2.
   2 runs of ones and 2 runs of zeros, each of length m-3.
   4 runs of ones and 4 runs of zeros, each of length m-4.

3.2.2 Algorithm of Linear Feedback Shift Register

Step1: Input coefficient (Ci), initial state (Si) (randomly value) and plain text (P). A linear feedback shift register of length L (length of initial value) consists of L stages numbered 0, 1…. L – 1.

Step2: Perform AND operation between coefficient and initial value.

Step3: Applying XOR operation between the bits of the result from step2.

Function $=S_0C_0 \oplus S_1C_1$

Step4: The first bit from result ($S_0C_0 \oplus S_1C_1$) puts in the sequence and shift the initial value by one.

Step5: The max length of sequence is (2^m-1) when m is length of coefficient vector.

Step6: Applying M-Sequence condition on sequence generate from previous steps
if M – sequence then go to step 7
else go to step 1

Step7: Convert the samples of P to binary. Repeat the M-sequence until become equal to length of plain text in binary.

Step8: Use XOR operation between plain text (samples) in binary and sequence.

3.3 Steganography Layer

In this layer used the cipher speech signal produced from previous two algorithms (compression and encryption) and embeds the bits of this signal into selected image.

3.3.1 sound hiding Algorithm

Step1: Read cover image of type **(RGB)** and save in **X.**
Step2: Read the bits of cipher speech signal (com_enc_sig) to be embed.
Step3: Convert each component byte of **(R G B)** of each pixel to binary.
Step4: Divided X into blocks of size M×N & from each pixel of block take only the low nibble bits of **(RGB)** bytes.
Step5: Generate key matrix (secret key shared by the sender and the receiver). The elements of key are randomly select of binary value and of size **m×n×12**.
Step6: read the number of bits (no_bit) to be embedded in each block of X. The value of no_bit should be test as following:

If $2^{no\_bits} - 1 \leq m \times n \times 12$ ***then***
Go to step7
Else
Enter another value of no-bit

Step7: Find the maximum size of message that accepted by the cover image X.
Let the variable total_bits represent the total bits to be embedded.

If total_bits $\leq$ ***no_block×no_bits then***
Go to steps of embedded the message (8)
Else
Select another file of speech message

Step8: Generate a weight matrix shared by the sender and receiver.

$$\{[W_{i,j}], i = 1,2,..m\ , j = 1,2,,n\} = \{1,2,3,\dots,2^{no_{bit}} - 1\}$$

Step9: perform the XOR operation.[11]

$$[FIK_k]_{i,j} = [X_k]_{i,j} \oplus key_{i,j}$$

Step10: perform the component wise multiplication operation. [11]

$$[ST_k]_{i,j} = [FIK_k]_{i,j} \otimes W_{i,j}$$

Step11: let the variable $Packt_{index}$ represent the bits embedded into block. [11]

If $\left(SUM([ST_k]_{i,j})\right) \bmod 2^{no\_bit} = Packt_k$ ***then***
The $[X_k]_{i,j}$ ***does not need to change***
Else
bit in $[X_k]_{i,j}$ ***should be modify***

The steps to modify the block:

If $[FIK_k]_{i,j}$***=0 then complementing*** $[X_k]_{i,j}$***will increase the modular sum by*** $W_{i,j}$***.***

If $[FIK_k]_{i,j}$***=1 then complementing*** $[X_k]_{i,j}$ ***will decrease the modular sum by*** $W_{i,j}$***.***

3.3.2 Sound Recovery Algorithm

Step1: Divided X(Stego_image) into blocks of size (M×N).
Step2: Calculate the number of embedded bits in each block from weight matrix.

Step 3: For each block do the step from 3 to 5

$$[FIK_k]_{i,j} = [X_k]_{i,j} \oplus key_{i,j}$$

Step4: Perform the component wise multiplication operation.

$$[ST_k]_{i,j} = [FIK_k]_{i,j} \otimes W_{i,j}$$

Step5: Find the bits embedded in this block by

$$\text{message\_bit}_k = \left(\text{sum} \left(\left[ST_k \right]_{i,j} \right) \right) \bmod 2^{\text{no\_bit}}$$

4. Measure for image and recovered sound quality

Image quality after hide message and Sound quality of recovered sound are usually judged by objective measures such as Correlation Co-efficient. Correlation is a measure of the strength of relationship between random variables. The population correlation between two variables X and Y is defined as:

$$\rho(X, Y) = \text{Covariance}(X, Y) / \{\text{Variance}(X) \times \text{Variance}(Y)\}^{1/2} \quad ...(4)$$

ρ is called the Product Moment Correlation Coefficient or simply the Correlation Coefficient. It is a number that summarizes the direction and closeness of linear relations between two variables. The correlation coefficient can take values between -1 through 0 to +1. The sign (+ or -) of the correlation defines the direction of the relationship. The table (1) shows guidelines for describing the strength and direction of a correlation.

Correlation coefficient (P)	**Strength**	**Direction**
P = 1	**Perfect linear relationship**	**positive**
1 > P >= 0.9	Very strong linear relationship	positive
0.9 > P >= 0.7	Strong linear relationship	positive
0.7 > P >= 0.5	Moderate linear relationship	positive
0.5 > P >= 0.3	Weak linear relationship	positive
0.3 > P > 0	Very weak linear relationship	positive
P = 0	**No relationship**	**No direction**
0 > P > -0.3	Very weak linear relationship	Negative
-0.3 >= P > -0.5	Weak linear relationship	Negative
-0.5 >= P > -0.7	Moderate linear relationship	Negative
-0.7 >= P > -0.9	Strong linear relationship	Negative
-0.9 >= P > -1	Very strong linear relationship	Negative
P = -1	**Perfect linear relationship**	**Negative**

Table (1) : strength and direction of a correlation

5. Experimental Result

The program of the proposed method was tested on the speech message signal of size 37.1 Kb with sampling rate 8 KHz and a 8 bits sample size. Figure (5) shows a sample speech message that to be hidden inside image before and after applying compression algorithm. The size of message after compression is 14.2 Kb and the compression ratio is 38.2.

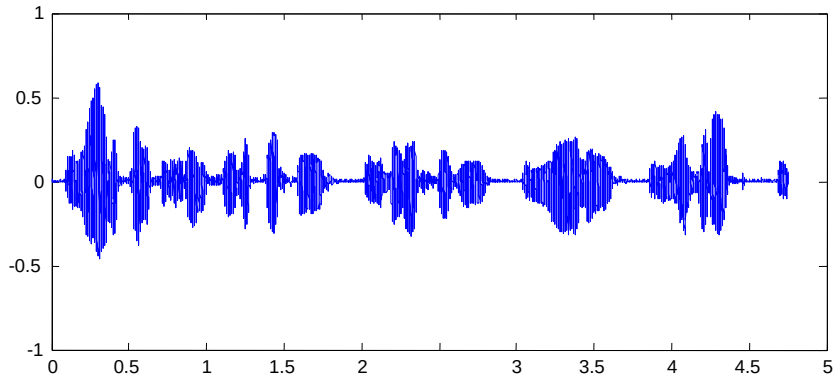


Figure (5.a) SAMPLES OF ORIGINAL SPEECH MESSAGE

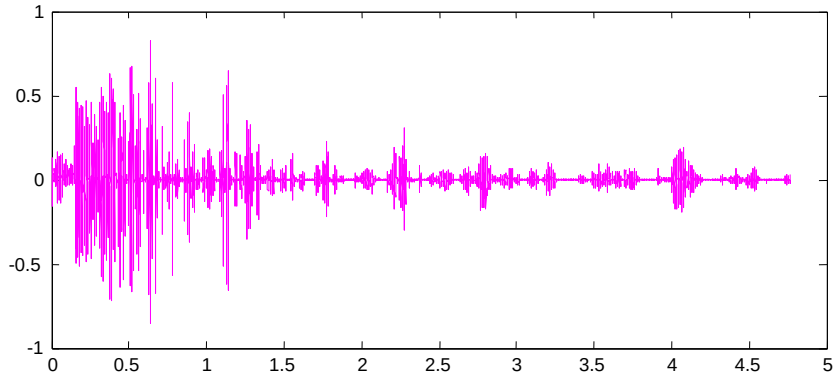


Figure (5.b) DWT COEFFICIENTS of MESSAGE AT LEVEL 5

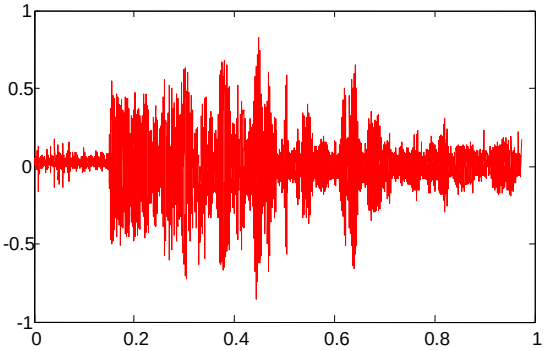
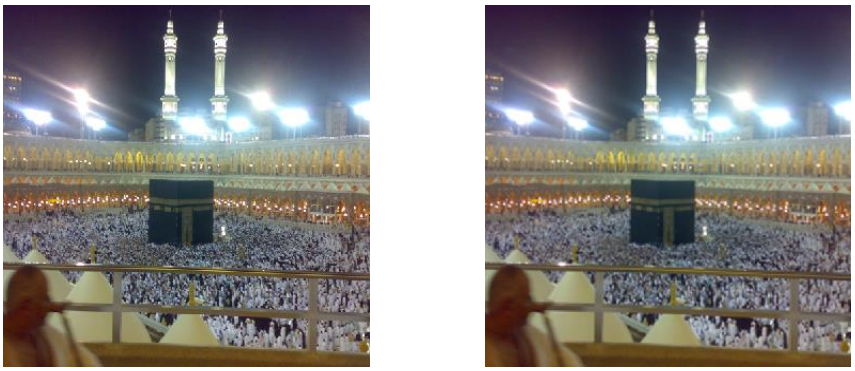


Figure (5.c) COMPRESSED SPEECH SIGNAL

Figure (5) Show Speech message

Figure (6) shows the 512×512 RGB image used to hide speech message before and after embedding. objective test used for measuring the quality of image after hide message correlation coefficient was calculated between the original image and stego-image according to the formula (3), the result of the calculation be equal to (0.998), a linear relationship is very strong according to the schedule (1). And the quality of the recovered speech measured using correlation coefficient was calculated between the original message and recovered message the result of the calculation be equal to (0.9838), a linear relationship is very strong according to the schedule (1) as shown in figure(7).



(6.a) Original image (6.b) stego image

Figure (6) Show the image

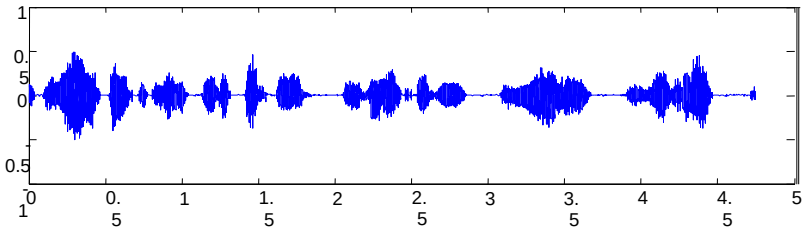


Figure (7) Speech Signal After Recovery

5. Conclusion

Through the implementation of this system, we conclude the followings:

1. The data-hiding ratio is improved by weight matrix used in the steganography algorithm.
2. we can improve the performance of steganography by applying compression, and reduced the size of data to increase the speed of encryption process that followed compression.
3. We used three layers of security to secure data by obscuring the context in which it was transferred. And the use of secret key matrix and weight matrix in steganography algorithm appended extra layer of security.
4. The steganography presented here is capable of hiding any type of files that can be present in bits stream.

REFERENCES

[1] A.J.Menezes, P.C. Van Oorschot,and S.A.Vanstone, (1997) ,"Handbook of Applied Cryptography".

[2] Christian Krätzer, Jana Dittmann, Thomas Vogel, Reyk Hillert, (2007),"Design and Evaluation of Steganography for Voice-over-IP".

[3] Donaled G.Childers, (2000),"Speech Processing and Synthesis Tool Boxes", United States of America.

[4] DU Cheng-tou, YAN Di-qun,(2008), "A Secure Speech Transmission System Based on Steganography" , Faculty of Information Science and Technology, Ningbo University, China, Journal of Ningbo University (NSEE) Dec., Vol.21 No.4

[5] H. Al-Barhmtoshy, E. Osman and M. Ezzat, (2004), "A Novel Security Model Combining Cryptography and Steganography", King Abdul-Aziz University, Computer Science Dept., 80203 Jeddah, Saudi Ar*abia.*

[6] Mr. P.D.Khandait, Mrs. S.P.Khandait, (2004), "LSB Technique for Secure Data Communication".

[7] Nameer N.EL-Emam,(2007), "Hiding a Large Amount of Data with High Security Using Stegangoraphy Algorithm", Faculty of information Technology, Philadelphia University, Jordan, Journal of Computer Science 3(4):223-232.

[8] René Rosenbaum. Heidrun Schumann,"A steganographic framework for reference colour based encoding and covers image selection".

[9] Scott Bishop, (2004), "Steganographic Techniques Using Digital Images", California State University, Hayward, August 18, CS 6520,Cryptography & Data Security

[10] Tin Lai Win, and Nant Christina Kyaw, (2008), "Speech Encryption and Decryption Using Linear Feedback Shift Register (LFSR)", Proceedings of world Academy of Science, Engineering and Technology, Volume 36 December ISSN 2070-3740.

[11] Umamaaheswarl, Manoj Kumar, A.Shanmugam, (2004),"A Novel Steganography Scheme for RGB images with high image quality and Data hiding capacity" ,Department of Elctronics and communication Engineering PSG College of Technology, Coimbatore, INDIA, Academic Open internet jornal Volume 12.