Journal of mathematics and computer Science 11 (2014) 218 - 230

Contents list available at JMCS

Journal of Mathematics and Computer Science

Journal Homepage: www.tjmcs.com

Speech Steganography in Wavelet Domain Using Continuous Genetic Algorithm

Hojat Allah Moghadasi
Department of Information Communication Technology, Malek Ashtar University of Technology, Iran
h_moghadasi@mut.ac.ir

Mohammad Fakhredanesh
Department of Information Communication Technology, Malek Ashtar University of Technology, Iran
m-fakhredanesh@aut.ac.ir

Article history:
Received May 2014
Accepted June 2014
Available online June 2014

Abstract

In this paper, we present a new adaptive steganography method using Lifting Wavelet Transform (LWT). In this method, we first calculate the LWT of the sample of host and secret speech signal. Then wavelet coefficients of secret speech signal will be fitted effectively and efficiently in host signal wavelet coefficients using continuous genetic algorithm. We used indirect replacement technique in 5 bits host using a proposed formula. Due to the quantization error, there are some differences between the secret signal before steganography and the extracted signal after steganography. However, these differences have an appropriate Gaussian noise model. We compress these differences using Huffman lossless compression method. The compression rate of such differences approach to the entropy, which is derived from Shannon's first theorem. Huffman lossless compression method, cause to small noise. We these compressed differences sent along the stego signal. The experimental results show that the proposed model has a statistical transparency higher than Least Significant Bit (LSB), Frequency Masking (FM) and Efficient Wavelet Masking (EWM) algorithms in time domain and frequency domain.

Keywords: Lifting Wavelet Transform (LWT), Elitism, Incest prevention, Premature convergence, Cycle crossover, Crowding Factor (CF)

1. Introduction

Nowadays information transmission security is one of the most important issues that is proposed about confidential internet messages. Information security is an important filed which has been developed in the recent years. It has been combined with techniques like cryptography, watermarking and steganography. There are exist three parameters that show the quality of a work in the data hiding: transparency, capacity and robustness.

Transparency is the ability to avoid suspicion about the existence of a secret message. It may be statistical transparency, which is performed by statistical analysis of host and stego signals, or perceptual transparency, which is referred to human auditory system. Capacity indicates the amount of information hidden in the host signal. Robustness measures the vulnerability against intentional and unintentional attacks [1]. In the case of steganography, the most important feature is the transparency followed by the capacity while in watermarking robustness plays a more important role. In general, an information hiding system in different media like sound, image and video should assure the high capacity, and desired statistical and perceptual transparency before and after steganography. Some steganography methods use the wavelet domain for hiding message bits due to its advantages such as of high hiding rate of sending up to 200 kbps [2]. Cvejic et al. increase the steganography capacity in the frequency 44.1 kHz [3]. Delforouzi and Pooyan development of sound steganography transparency in the wavelet domain using a suitable hearing threshold [4], while Shahreza et al. proposed to avoid embedding in silence areas [5]. Adaption of stego signal to host signal with indirect replacement in the wavelet domain cause to increase the steganography capacity [1].

In this paper, we proposed to use a genetic algorithm and indirect replacement to steganography in the wavelet domain. In the next section, a review of some related works is presented. Section 3 describes the proposed method in two disjoint parts: data embedding and data extraction process. The experimental results and conclusions are provided in the sections 4 and 5 respectively.

2. Related works

In this section, some related works are mentioned. In some works the Frequency Masking (FM) mechanism are used for steganography in speech signals. In this method all samples of secret messages hide in the frequency components of host signal based on a threshold. In this procedure, a primary attenuation is necessary, and the threshold was fixed to -13 dB, then the maximum dynamic range of the secret message is assumed 2 bits below the host signal [6]. Another method is based on Efficient Wavelet Masking (EWM) where the wavelet coefficients are classified in significant and non-significant according to a threshold. Because a coefficient with low amplitude (non-significant) does not provide significant energy to the signal, it may be set to zero. This process is known as thresholding. The wavelet coefficients of secret message are reordered to be similar to the wavelet coefficients of the host signal. Practically if the non-silence duration of the secret message signal is less than or equal to the non-silence duration of the host signal, the secret message may be heard similar to the host signal. In this scheme, the secret message signal is 12 dB less than the host signal. In addition, a secret key is assigned to determine the first position [1]. Another domain that is used for audio steganography is frequency domain. Perceptual transparency is based on this fact that Human Auditory System (HAS) in high frequencies has less sensitivity. Therefore, the secret message can be secret in high frequencies without being suspect. Hence, the secret message spectrum is shifted to the highest frequencies of the host signal spectrum. In this scheme, the range of 18 kHz to 22 kHz of the host signal hides a speech signal with bandwidth of 4 kHz [7].

3. Proposed method

In this section, proposed algorithm is described in two subsections: data embedding process and data extraction process.

3.1. Data embedding process

This process contains six steps: down sampling, lifting wavelet transform, continuous genetic algorithm, indirect LSB replacement, inverse lifting wavelet transform and Huffman lossless compression. The stages of data embedding process are shown in Fig. 1:

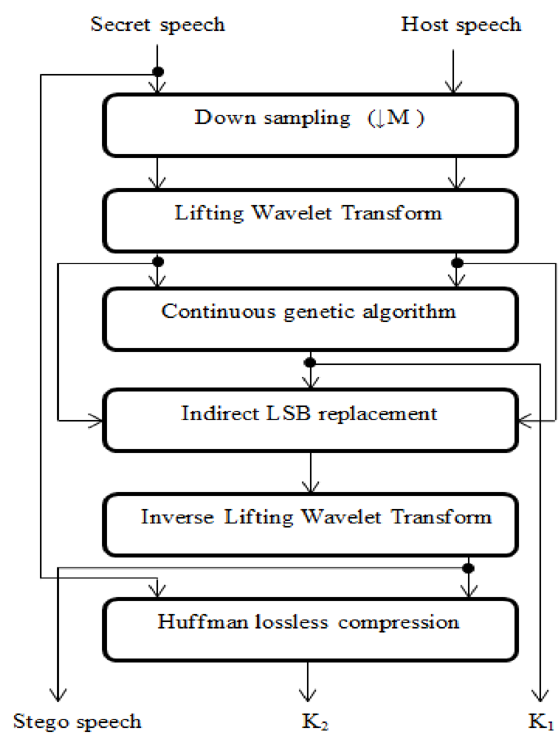

Fig. 1: Flowchart of the data embedding process

I. Down sampling

First, we proposed to implement a down sampling to have faster processing because the host and secret signal samples are very voluminous.

II. Lifting Wavelet Transform (LWT)

In this part of proposed algorithm, the LWT is applied to both host speech signal and the secret speech signal with the Haar filter in order to obtain the integer coefficients. Two important properties of the wavelet transform are the difference of the filters bank bandwidth, and using down sampling. These two properties are used in denoising and compression procedure. In each phase of information, transmission of dyadic wavelet transform audio signal is decomposed into two parts, detail coefficients (L) and approximation coefficients (H). Approximation coefficients contain low frequency components and detail coefficients contain high frequency components. We proposed to put the secret speech signal in approximation coefficients of host speech signal. Using windows with smaller length in higher frequencies cause to increase time accuracy of these frequencies. In addition, such structure is more adapted with human's ear cochlear.

III. Continuous genetic algorithm

The genetic algorithm (GA) is an optimization and search technique based on natural selection and the principles of genetics science. The genetic algorithm belongs to the family of evolutionary algorithms that was developed by John Holland in 1975 [8]. The GA starts with a group of chromosomes known as the population. Then, different chromosomes are composed together repeatedly and new generations are produced. The next generations are produced based on the principle of maximization a fitness function or minimization a cost function.

In this algorithm two main reproduction operators are used: the crossover and mutation. In crossover operator, some values of two or more chromosomes are exchanged and new chromosomes are created. The mutation operator, generates a new random chromosome by pseudo-random replacement of some values in a chromosome. At the end of each iteration a new population has been produced by applying a certain number of stochastic operators to the previous population. The chromosomes which are chosen to form new chromosomes are selected according to their fitness function. It means that next generations are usually more close to the optimum solution of the fitness function.

Problem solving strategy

Our solving strategy based on genetic algorithm is described with Pseudo code and is shown :

Procedure genetic algorithm

begin
chromosome=random permutation of size of cover
set $counter_i$
R=total search space
$R_1, R_2, \ldots, R_n \leftarrow R$ *// partitioning into subspace*
for every R_i *do {*
for GA=1 to Generation do
for i=1 to number of population do
for j=1 to size of cover do
$Z_i(j) \leftarrow ceil(C*c(Chr_i(j))*m(j)^{-1})$
end $_j$
if size of $Z_i(j) \leq 5bit$ *then*
$counter_i \leftarrow counter_i + 1$
end if
end $_i$
fitness evaluation {
alculate $counter_j$

}
calculate best chromosome
selection for mating pool
cycle crossover
dynamic & swap mutation
elitism selection
crowding replacement
incest prevention
end $_{GA}$
k_1=best chromosome

The structure of genetic algorithm

First the search space is partition into several disjoint sub-spaces. On the other hand, we assume that R is total space, we have:

$$R = R_1 \cup R_2 \cup \ldots \cup R_n \quad (1)$$

where R is all search spaces and R_i's, i=1,2,…,8 are sub spaces. For each R_i, the population and chromosomes are defined as follows:

$$\text{population} = \{Chr_1,\ldots,Chr_t\} \quad , \quad t\leq 16 \quad (2)$$

$$Chr_j = [a_1, a_2, \ldots,a_m] \quad , \quad m \leq \text{length}(R_i) \quad (3)$$

$$a_x \in \{1,2,\ldots,\text{length}(R_i)\} \quad , \quad i=1,2,\ldots,8 \quad (4)$$

The following step is performed k times for indirect embedding:

$$Z_j(k) = \left\lceil \frac{C * c\left(chr_j(k)\right)}{m(k)} \right\rceil \text{ , K=1,2,…, length}(R_i) \text{ , } j \leq 16 \quad (5)$$

where $\lceil x \rceil$ is least integer greater than or equal to x, C is positive integer constant, c and m are the wavelet coefficients of host speech and secret speech respectively. In this paper, C is proposed to be 2. We called the lower bound of each $Z_j(k)$ as d_{min} and their upper bound as d_{max}. In chr_j for every specific a_x ,x=1,2,…, length(R_i) the number of $Z_j(k)$'s which are in the set $\{d_{min}, d_{min+1}, \ldots, d_{max}\}$ are computed and assigned to $counter_j$:

$$\text{If } Z_j(k) \in \{d_{min},\ldots,d_{max}\} \text{ then } counter_j \leftarrow counter_j+1 \quad (6)$$

In this method d_{min} = 2 and d_{max} = 31. Now, the fitness function F for each j is defined as $counter_j$:

$$F(j) = counter_j \quad , \quad j=1,2,\ldots,10 \quad (7)$$

The better chromosomes of each generation are transmitted to next generation directly based on the elitism approach. Of course, it should not transmit many of the better chromosomes directly, because it makes the dominant specie and implies a premature convergence. For this reason, just the best chromosome is transmitted. According to the researches accomplished by Haupt size of the population for continuous genetic algorithms is preferred to be less than 16 [9]. Worse chromosomes are not removed certainly, because all properties of these chromosomes are not weak and the deletion of them can loss of diversity in next generation. For this reason, the weakest chromosome is maintained. Since the roulette wheel selection, have high selection pressure, the direct application of its cause loss of diversity and premature convergences. Hence, it is suitable to use the fitness scaling methods for adjustment in selection pressure. In this paper, we are using Boltzmann's selection as follows:

$$P_i = \frac{\text{Fitness value}}{\sum_{i=1}^{10}(\text{Fitness value}_i)} \qquad (8)$$

$$Q_i = \exp\left(\frac{(P_i - 1) * \text{generation}}{(2 * \max\{\text{generation}\})}\right) \qquad (9)$$

After computing Q_i values, parent chromosomes are chosen from roulette wheel. The generation is the number of iterations. In fact by combining the Boltzmann's selection and roulette wheel at the beginning of running genetic algorithm, fitness of weak and strong chromosomes are closed together and so the selection pressure is keeping down. While the next generations produced, the distance of weak and strong chromosomes fitness become far from each other. Therefore, the selection pressure will be increased slowly. The cycle crossover method (CX) and single point are applied for crossover operator. A limitation of this crossover approach is that the number of each Allele (value of each gene) is derived from only one parent. This crossover approach is not depends on the location of cut and a similar cycle, spans a group of algebraic permutations for the children. Since in permutation problems great mutation rate is not suitable [10] the mutation rate is set to be low, and is obtained dynamically by the following equation:

$$P_m(t) = \left(2 + \frac{(\text{length}(R_i) - 2) * t}{(\text{generation} - 1)}\right)^{-1} \qquad (10)$$

Where t is the current generation counter. This mutation is defined as a swap type, which two points of a chromosome is chosen randomly and their locations will be changed.

IV. Indirect LSB replacement

After choosing the best chromosome (K_1), indirect replacement is computed as follows:

$$Z_{final}(k) = \left\lceil \frac{C * c\left(k_1(k)\right)}{m(k)} \right\rceil, \quad K = 1, 2, \ldots, \text{length}(R_i) \qquad (11)$$

Where Z_{final} is the indirect representative of secret speech signal. These coefficients are embedded in the 5 Least Significant Bits (LSB) of host speech signal.

V. Inverse Lifting Wavelet Transform (ILWT)

In this block, the ILWT of the stego signal is performed and sent to the receiver.

VI. Huffman lossless compression

Difference between the secret signal and the extracted signal due to quantization, cause to decreasing the received signal quality. This difference signal, which is called K_2 key, can be modeled by a Gaussian noise.

3.2. Data extraction process

This process contains three steps: lifting wavelet transform, speech signal retrieval and inverse lifting wavelet transform. The stages of data extraction process are shown in Fig. 2:

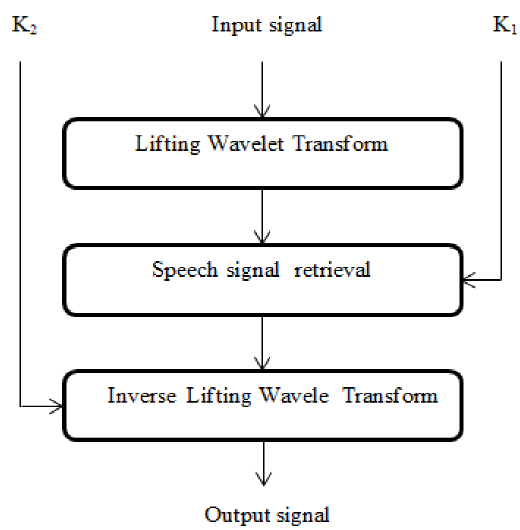

Fig. 2: Flowchart of the data extraction process

I. Lifting Wavelet Transform

In this step we also use the LWT with Haar filter same as the LWT part in the data embedding process.

II. Speech signal retrieval

In this part, we apply k_1 key on wavelet coefficients. The secret speech signal G_2, which is the lifting wavelet, transform of input signal is obtained as follows:

$$G_2(k) = \left\lceil \frac{C*G_1(K_1(k))}{\mathrm{dec}(m_b)} \right\rceil \quad (12)$$

In equation (12), G_1 contains the approximation coefficients of stego signal (input signal) in wavelet domain. m_b is the five least significant bits of wavelet transform coefficients of stego signal that have the information related to the secret signal and *dec* is decimal format.

III. Inverse Lifting Wavelet Transform

In this part, the inverse lifting wavelet transform of G_2 signal is performed and then K_2 key is added for increasing the sound quality of secret speech.

4. Experimental results

In this paper, a combination of Elitism replacement and crowding replacement together with incest prevention is proposed. As long as the population evolves, its members become more similar. To satisfy incest prevention it is tried to mate the parents, which are more different, this causes more evolution. This method computes the Euclidean distance between chromosomes and the best chromosome. If this difference is greater than the threshold level, we put it in the parents list [11]. In the proposed method, the threshold level is considered 20000. If at the end of each evaluation, there is no change in the parents list, the threshold level decrease 1000 units. For keeping the population diversity, we use crowding replacement method. Crowding Factor (CF) is considered 3 and the number of population is chosen randomly. Then the chromosome with the highest level of similarity is replaced by the chosen chromosome. The results of Fig. 3 show that incest prevention cannot promote fitness function lonely. However, combination of it with crowding replacement will result more favorable.

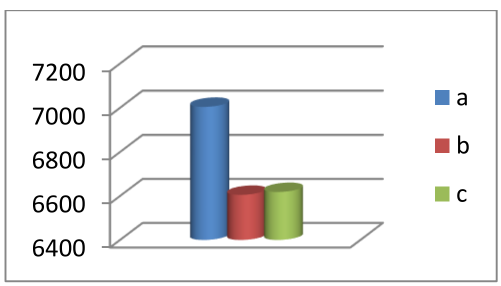

Fig. 3: (a) With incest prevention and crowding replacement, (b) With incest prevention and without crowding replacement, (c) Without incest prevention and crowding replacement

In this paper, 2 seconds are assigned to both host speech signal samples and secret speech signal samples, the frequency sampling is 44/1 KHz and 16 bits are assigned to each sample. In this method, down sampling with scale 6, is proposed ($\downarrow M=6$) and the secret message is attenuated by -6 dB in relation to the host signal. In Fig. 4 and Fig. 5, the host signal and stego signal and the differences between them are shown in time domain and wavelet domain.

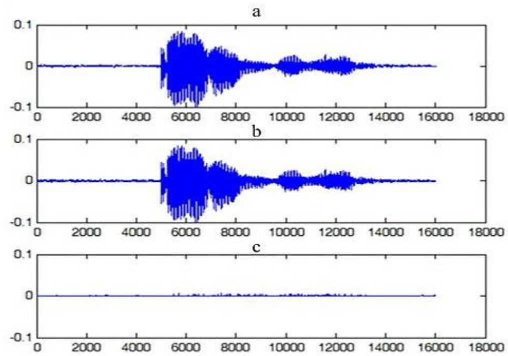

Fig. 4: (a) Host signal, (b) Stego signal, (c) Difference between host signal and stego signal (time domain)

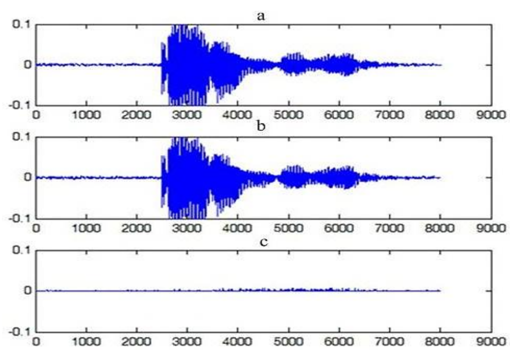

Fig. 5: (a) Host signal, (b) Stego signal, (c) Difference between host signal and stego signal (wavelet domain)

In this work, we apply two steganalysis methods, which are steganalysis test in time domain [13] and steganalysis test in frequency domain [12], and compute four formulas as follow:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} \quad (13)$$

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{(\mu-1)} \quad (14)$$

$$sk = \frac{\sum_{i=1}^{N}(x_i-\mu)^3}{(N-1)\sigma^3} \quad (15)$$

$$k = \frac{\sum_{i=1}^{N}(x_i-\mu)^4}{(N-1)\sigma^4} \quad (16)$$

where μ is the mean, σ^2 is the variance, sk is the skewness, k is the kurtosis, x is the input signal of the steganalysis test (x[n], X(f)) and N is the size of x. We compute the analysis in time domain and frequency domain, according to the bellow relations [12,13]:

$$X[n]=\log_{10}(|v[n]|+1) \quad (17)$$

$X_a[n]=[v[2]-v[1] \ldots v[n]-v[n-1]]$

$X_b[n]=[X_a[2]-X_a[1] \ldots X[n]_a-X_a[n-1]]$

$X(f)=abs(fft\{ X_b[n]\})$ (18)

In the above relations v[n] is the input speech signal. Results of the proposed method are compared to the Least Significant Bit (LSB) [3], Frequency Masking (FM) [6], Efficient Wavelet Masking (EWM) [2] and shown in table 1 and table 2. The obtained numbers are given as percent of difference between the host signal and the stego signal for above parameters.

Table 1: Four first statistics moments in the time domain

Method	Mean (%)	Variance (%)	Skewness (%)	Kurtosis (%)
LSB	1.110	0.543	0.377	0.352
FM	4.200	2.370	6.590	2.790
EWM	0.204	0.137	0.231	0.128
Proposed method	0.018	0.094	0.016	0.008

Table 2: Four first statistics moments in the frequency domain

Method	Mean (%)	Variance (%)	Skewness (%)	Kurtosis (%)
LSB	2.580	2.870	9.440	5.690
FM	12.64	3.340	1.510	2.600
EWM	0.571	0.922	3.310	1.520
Proposed method	0.017	0.018	2.060	2.990

Tables 1 and 2 show four first statistics value for the proposed method and three related works in the time and frequency domain respectively. Table 1 show that four first statistics of the proposed method, is much lower than other methods. This is due to using the genetic algorithm to find the best location for embedding message bits. In Table 2, two first statistics moment of the proposed method are much lower than the other methods. Only skewness and kurtosis values are slightly higher than the corresponding values of other methods.

The Fig. 6 shows the number of selections for steganography before and after using the genetic algorithm in 8 subsections for 500 generation. In this method, the size of population is considered 10 the maximum capacity is in first subsection with 11.5% and minimum capacity is in fifth subsection with 5.5%, as shown the Fig. 6. Using the genetic algorithm suitable locations increase about 10% in average.

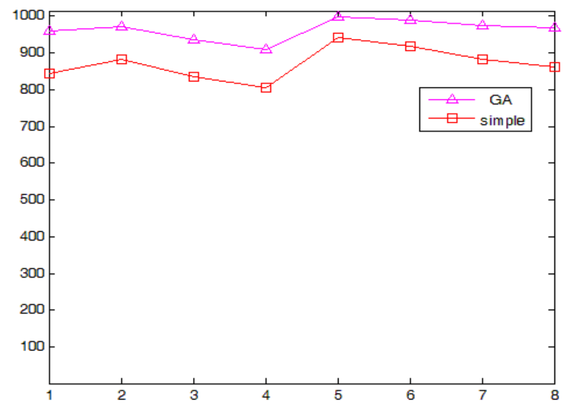

Fig. 6: Number of selection before and after genetic algorithm

The experiments show that the difference between before and after the extraction of secret speech signal can be modeled with a channel noise. The histogram of the values in the interval [-1,1] is plotted in Fig. 7. This Figure shows that a Gaussian distribution can estimate the histogram.

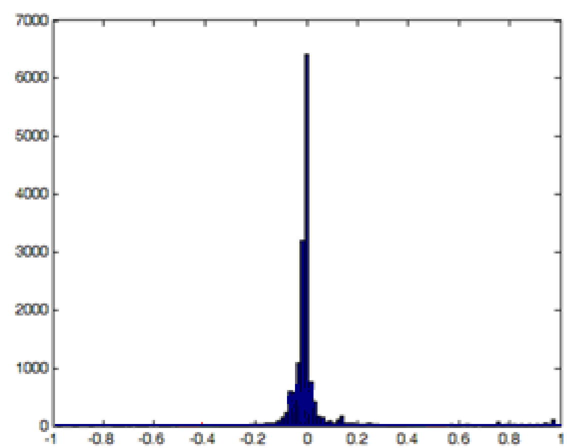

Fig. 7: Modeled noise with Gaussian distribution

According to Shannon's first theorem (lossless source coding theorem) source entropy, is the essential limit on the number of bits required for the source recovery target. We use Huffman coding for lossless compression of channel-modeled noise. In this kind of encoding for each symbol some bits are considered proportional with probability inverse. These calculations show that the proposed compression rate is much closed to the Shannon's limit i.e. the entropy. The result of compression is shown in table 3. Source code entropy (H), the average length of Huffman code ($\bar{n}$), redundancy of source code (ρ), source redundancy with Huffman code (ρ'), source efficiency (e), and source efficiency with Huffman encoding (e') from the following relations are calculated (s is the number of source symbols and p_i is the related probability).

$$H=-\sum_{i=1}^{s} p_i log_2 p_i \qquad (19)$$

$$\bar{n} = \sum_{i=1}^{S} n_i p_i \qquad (20)$$

$$e = \frac{H}{log_2^S} \quad , \quad \rho = 1 - e \qquad (21)$$

$$e' = \frac{H}{\bar{n}} \quad , \quad \rho' = 1 - e' \qquad (22)$$

Table 3: Results of Huffman encoding

Quantitative evaluation	values
H	9.3131
$\bar{n}$	9.3381
P	15%
ρ'	0.27%
e	85%
e'	99.73%

Table results, show the performance of Huffman encoding in data compression with high efficiency. In fact, the rate of 9.3 bit/symbol is reach to sending noise signal, rather than the rate of 16 bit/symbol.

5. Conclusions

In this paper, a sound steganography method is proposed in the wavelet domain. Due to using genetic algorithm and appropriate designing of its parameters, suitable locations are chosen for steganography. The four first statistics moments in the proposed method, are developed comparing to three methods, LSB, FM and EWM. The advantage of the method is the independency to host signal. The similarity between stego signal and host signal increases, because we use indirect LSB replacement. Due to the quantization error between the secret signal before and after extraction, there is a difference which sent by Huffman lossless compression with the stego signal.

This method can be used in all speech steganography (speech in speech) approaches. However, there are some ideas for future works to improve proposed method in this paper. Some of the most challenges of using genetic algorithms are parameters values, mutation and crossover types, and chromosome selection and replacement approaches utilizing in the algorithm. Since the novelty of this paper is utilizing genetic algorithm in the audio steganography, there was not enough time to test other aspects of this topic. However, suitable results of the proposed algorithm, is implied to improve them in future works.

In addition, genetic algorithm used in the proposed algorithm, can be replaced by other meta-heuristic approaches such as the *ant colony optimization* and *simulated annealing*. The meta-heuristic approaches can be useful where the search space is very huge and the complete search of it is impossible in the acceptable time. Other meta-heuristic approaches may have some better results in comparison with the genetic algorithm.

6. References

[1] D. M. Ballesteros L and J. M. Moreno A, "Highly transparent steganography model of speech signals using Efficient Wavelet Masking," *Expert Systems with Applications,* vol. 39, pp. 9141-9149, 2012.

[2] F. Djebbar et al., "A view on latest audio steganography techniques", in Innovations in Information Technology (IIT), 2011 International Conference on, 2011, pp. 409-414.

[3] N. Cvejic and T. Seppanen, "A wavelet domain LSB insertion algorithm for high capacity audio steganography", in Digital Signal Processing Workshop, 2002 and the 2nd Signal Processing Education Workshop. Proceedings of 2002 IEEE 10th, 2002, pp. 53-55.

[4] A. Delforouzi, Mohammad Pooyan. "Adaptive digital audio steganography based on integer wavelet transform." Circuits, Systems & Signal Processing 27, no. 2 (2008): 247-259.

[5] T. Manzuri et al., "High capacity error free wavelet domain speech steganography." In Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on, pp. 1729-1732. , 2008.

[6] F. Djebbar, H. Hamam, K. Abed-Meraim, and D. Guerchi, "Controlled distortion for high capacity data-in-speech spectrum steganography," in Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on, 2010, pp. 212-215.

[7] D. E. Skopin et al., "Advanced algorithms in audio steganography for hiding human speech signal", in Advanced Computer Control (ICACC), 2010 2nd International Conference on, 2010, pp. 29-32.

[8] J. H. Holland, Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence: U Michigan Press, 1975.

[9] R. L. Haupt, "Optimum population size and mutation rate for a simple real genetic algorithm that optimizes array factors", in Antennas and Propagation Society International Symposium, 2000. IEEE, 2000, pp. 1034-1037.

[10] D. Whitley, T. Starkweather, and D. Shaner, The traveling salesman and sequence scheduling: Quality solutions using genetic edge recombination: Colorado State University, Department of Computer Science, 1991.

[11] L. J. Eshelman, R. A. Caruana, and J. D. Schaffer, "Biases in the crossover landscape", in Proceedings of the third international conference on Genetic algorithms, 1989, pp. 10-19.

[12] Q. Liu, A. H. Sung, and M. Qiao, "Temporal derivative-based spectrum and mel-cepstrum audio steganalysis", Information Forensics and Security, IEEE Transactions on, vol. 4, pp. 359-368, 2009.

[13] Y.-C. Qi, L. Ye, and C. Liu, "Wavelet domain audio steganalysis for multiplicative embedding model", in Wavelet Analysis and Pattern Recognition, 2009. ICWAPR 2009. International Conference on, 2009, pp. 429-432.