Improving Hiding Information Process based on GA Technique with Secure Extraction Process

Lina Saeed Jajo* Susan .S. Ghazoul*

Received on: 12/07/2010
Accepted on: 15/12/2010

Abstract

In this paper we propose a new method of hiding information that produces a stego-image which is totally indistinguishable from the original image to extract the hiding message. GA is used as an efficient method to minimize the number of different bits between the cover image and the stego-image as minimum as possible by embedding the message in random locations of cover image, and then modifying the locations containing changed information in original image (cover) to improve stego-image quality. To satisfy excellent security we used a crypto-key which contains encrypted locations from hiding process. This key is used to extract the embedded message.

Keywords—Genetic Algorithms, Steganography, Data Hiding, Data Extraction, Random Locations.

تحسين عملية اخفاءالمعلومات المعتمدة على الخوارزمية الوراثية مع الأستخراج السري

الخلاصة

في ورقة العمل نقترح طريقة جديدة لإخفاء المعلومات لتوليد صورة تحوي معلومات مخفية بحيث يكون صعب التميز بين الصورتين لاستخلاص الرسالة المخفية . استخدمت الخوارزمية الوراثية كطريقة كفوءة لتقليل عدد البتات المختلفة بين الصورتين إلى اقل حد ممكن وذلك بتضمين الرسالة في مواقع عشوائية داخل صورة الغطاء ثم تحسين مواقع التي تغيرت معلوماتها لتحسين جودة الصورة التي تحوي معلومات مخفية. للتحقيق سرية ممتازة أستخدم المفتاح السري الذي يحوي المواقع المشفرة لعملية الاخفاء. يستخدم المفتاح في استخلاص الرسالة المخفية.

*Department of Control and Systems Engineering University of Technology Baghdad, Iraq.

1. Introduction

Steganography is the study of techniques for hiding the existence of a secret message in the presence of a cover file. The cover file is referred to as the carrier signal; the message is referred to as the payload signal or payload message. Steganography itself offers mechanisms for providing confidentiality and deniability; it should be noted that both requirements can also be satisfied solely through cryptographic means [1].

In recent years, many successful steganography methods have been proposed. Among all the methods, LSB (least significant bit) replacing method is widely used due to its simplicity and large capacity. In the LSB steganography, a secret message is converted into binary string. Then the least significant bit-plane is replaced by the binary string. The LSB embedding achieves good balance between the payload capacity and visual quality. However, the LSB replacing method flips one half of the least-significant bits. Thus the artifacts in the statistics of the image are easy to be detected [2]. Experimental results illustrate comparison between the steganography using GA technique and LSB method, and it is revealed that the proposed steganography based on GA exhibits excellent security and excellent image quality.

through the recent years many hiding researches are developed like: The "**Customized and Secure Image Steganography Through Random Numbers Logic"** by Sanjeev Manchanda, Mayank Dave, and S. B. Singh applied random numbers based methods and layout management schemes on least significant bit transformation for steganography and working upon steganalysis for the proposed methods and layout management schemes [3], **"A Secure Steganography Method based on GA"** by Shen Wang, Bian Yang and Xiamu Niu employs GA in modifing the pixel values of the steg-image, After embedding the secret message in LSB (least significant bit) of the cover image [4], and " **A Steganographic Approach by Using Session Based Stego-Key, Genetic Algorithm and Variable Bit Replacement Technique"** by Tanmay Bhattacharya, Sandeep Bhowmik, and S. R. Bhadra Chaudhuri, the secret image is firstly perturbed by Stego-Key and again perturbed by a genetically generated. In the next step the perturbed secret image is embedded within the Host image using a hash function [5].

2. Attacks and Robustness

There are two types of attacks to steganography and therefore there are two types of robustness. One type of attacks tries to reveal the hidden message and another type tries to destroy the hidden message. Substitution techniques are vulnerable against both types of attacks. The adversary who tries to reveal the hidden message must understand which bits are modified. Since substitution techniques usually modify the bits of lower layers in the samples LSBs, it is easy to reveal the hidden message if the low transparency causes suspicions.

Also, these attacks can be categorized in another way: Intentional attacks and unintentional attacks. Unintentional attacks like transition distortions could destroy the hidden message if it is embedded in the bits of lower layers in the samples LSBs [3]. In this paper we used GA as powerful and robust tool of hiding against the above two types of attacks.

3. GA Approach

The genetic algorithm (GA) is an optimization and search technique based on the principles of genetics and natural selection. A GA allows a population composed of many individuals to evolve under specified selection rules to a state that maximizes the “fitness” (i.e., minimizes the cost function). The method was developed by John Holland (1975). The genetic algorithm starts with no knowledge of the correct solution and depends entirely on responses from its environment and evolution operators such as reproduction, crossover and mutation to arrive at the best solution. By starting at several independent points and searching in parallel, the algorithm avoids local minima and converges to sub optimal solutions. In this way, GAs have been shown to be capable of locating high performance areas in complex domains without experiencing the difficulties associated with high dimensionality, as may occur with gradient decent techniques or methods that rely on derivative information [7, 8].

4. GA Steps Used in Hiding Process

The following steps of proposed GA are used to hide the massage after reading it:

1. Initially make random population of chromosomes, the length of each chromosome is equal to the secret data. The population size should be a large number because the length of chromosome is long. Each chromosome contains the locations in pixels of cover file in which the message is hidden.
2. Evaluate fitness function which is the objective function (amount of error defined as the number of difference bits between cover image and stego-image) for each chromosome in population as it calculates in section 5.
3. Repeat the following steps until new population has been created:
   a. Select a pair of parent chromosomes from the current population, the probability of selection being an increasing function of fitness. Selection is done "with replacement" meaning that the same string can be selected more than once to become a parent.
   b. With the crossover probability, crossover the pair at a randomly chosen multi-points to form two new strings. If no crossover takes place, form two new strings that are exact copies of their respective parents.
   c. Mutate the two new chromosomes at each locus with the mutation probability, and place the resulting strings in the new population.
4. Replace the current population with the new population.
5. If the optimal solution (minimum amount of error) is satisfied by fixing the value of error with a number of generations or a maximum generation number is reached then stops and store the best chromosome (key) that satisfies minimum amount of error. This key contains the best locations of pixels in cover file to hide the secret data. For more security Caesar encryption method is used to encrypt the best chromosome to get crypto key, else go to step 2.

5. Hiding Process with Improved GA Using LSB

It is the process of embedding the secret message in cover file. The simplest and the most widely used method of hiding is LSB. The LSB replacing method flips one half of the least-significant bits. Thus, the artifacts in the statistics of the image are easy to be detected by comparison

between the original image and stego-image, and may be getting many the amount of error (the number of difference bits between cover image and stego-image). For example the massage is the following 8 bits (10100101), and the hiding process using LSB (the hide bit is the bold bit) in following cover file:
11100110 11111101 00000010
11111111 01011101 00000101
11111110 00000000, as shown below:

1110011**1** 1111110**0** 0000001**1**
1111111**0** 0101110**0** 0000010**1**
1111111**0** 0000000**1**
The amount of error=6

The GA technique searches the best random location in pixels of the cover file to hide the secret data. The best location in each pixel is the location in pixel of cover file to hide a bit of secret data without changing the original information in pixel of cover file, in other words the information of both the massage and the best location of pixel in cover file coincide. In this method the amount of error is less than LSB and discovering the embedded message will be impossible because the attacks can not expect where the_massage data are hidden. For example the same massage used above (10100101) hides using GA (the hide bit it is the bold bit) in following cover file:
11100110 11111101 00000010
11111111 01011101 00000101
11111110 00000000, as shown below:

11**1**00110 111111**0**1 000000**1**0
1**0**111111 010111**0**1 0000010**1**
1111111**0** 0000000**1**
The amount of error =2.

From the above example we see the amount of error is decreased but the image quality will be decreased because of the changing density of byte (4) is greater than using LSB method. To solve this problem we proposed to modify the location of hiding bit from bit 7 to bit 1 that satisfies high image quality. We apply LSB method to the best key that results from GA. The LSB is applied to locations that contain changing information in cover file after hiding process using GA. This modification had been very effective on improving the image quality, as shown in table below.

Byte (4)	The value of the byte (4)	density
In cover file	“11111111”	255
After applying GA hiding process	“10111111”	191
After improving GA hiding process with LSB	“11111110”	254

The previous example illustrate that the changing in density of byte (4) in cover file after applying GA process (255-191=64) is greater than after improving GA hiding process with LSB (255-254=1). Figure (1) illustrates the essential steps of using secure improved hiding process based on GA using LSB.

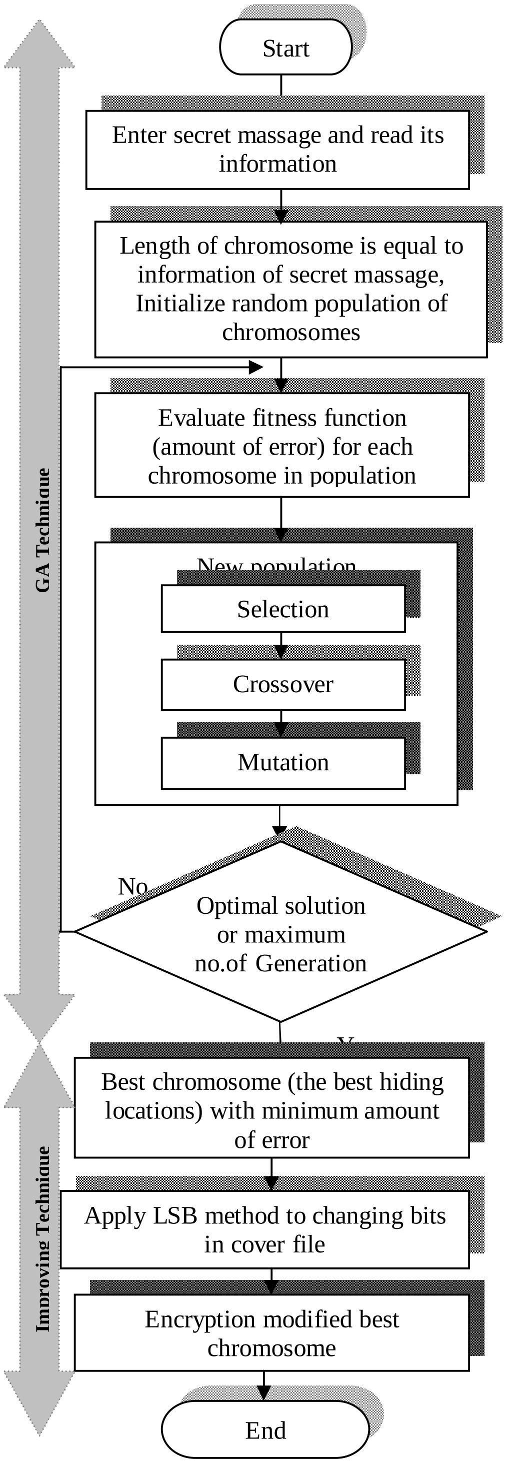

GA Technique
Improving Technique
Start
Enter secret massage and read its information
Length of chromosome is equal to information of secret massage, Initialize random population of chromosomes
Evaluate fitness function (amount of error) for each chromosome in population
New population
Selection
Crossover
Mutation
No
Optimal solution or maximum no.of Generation
Y
Best chromosome (the best hiding locations) with minimum amount of error
Apply LSB method to changing bits in cover file
Encryption modified best chromosome
End

Figure (1) Improving a secure hiding process based on Genetic Algorithms technique using LSB method.

6. Extraction Process

For excellent security we used crypto key to extract the secret massage from cover file. The crypto key contains the encrypted location of embedding bits. Here the extraction process is used to test getting the correct message which is used in hiding process and to improve the impossibility of extracting the hiding massage without the crypto key.

Figure (2) shows the proposed steps that are used in hiding process (based on GA) and extraction process.

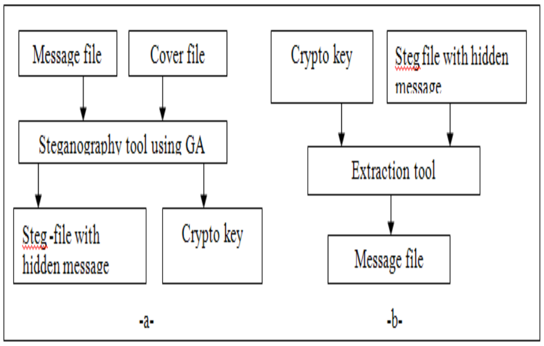

Message file
Cover file
Steganography tool using GA
Steg -file with hidden message
Crypto key
-a-
Crypto key
Steg file with hidden message
Extraction tool
Message file
-b-

Figure (2) Proposed steps, a) for hiding process based on GA, and b) extraction process.

7. Experimental Results and Discussion:

We used different samples to test the proposed hiding technique. Compression between the LSB method results and proposed improved GA method results are used as testing experimental process.

(Sample1): Hide text1 (short text) in Lena.bmp image (size 92x92 pixels).

- Applying LSB method and the results are:

MSE= $5.265*10^{-2}$

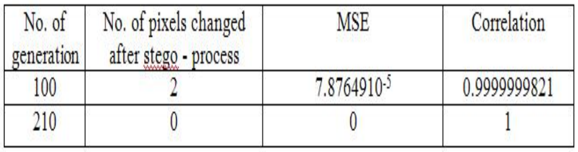
Number of pixels changed after stego – process (amount of error) = 56

- Applied the improved GA method ,and the GA specifications and results are:

Population size=400

Pc=0.9
Pm=0.01

The tested results are shown in table below, and the cover image and stego-images shown in figure (3).



No. of generation	No. of pixels changed after stego - process	MSE	Correlation
100	2	7.8764910^{-5}	0.9999999821
210	0	0	1

Cover image
Stego-image with 100 generations
Stego-image with 210 generations

Figure (3) Lena.bmp before and after hiding process.

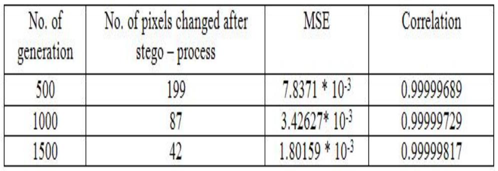
(Sample2): Hide text2 (medium text) in Lena.bmp image (size 92x92 pixels).

- Applying LSB method and the results are:

MSE= 2.0951*10 $^{-2}$
Number of pixels changed after stego-process (amount of error) = 532

- Applied the improved GA method, and the GA specification are:

Population size=1000
Pc=0.9
Pm=0.01
The tested results are shown in table below, and the cover image and stego-image shown in figure (4).



No. of generation	No. of pixels changed after stego – process	MSE	Correlation
500	199	$7.8371 * 10^{-3}$	0.99999689
1000	87	$3.42627* 10^{-3}$	0.99999729
1500	42	$1.80159 * 10^{-3}$	0.99999817

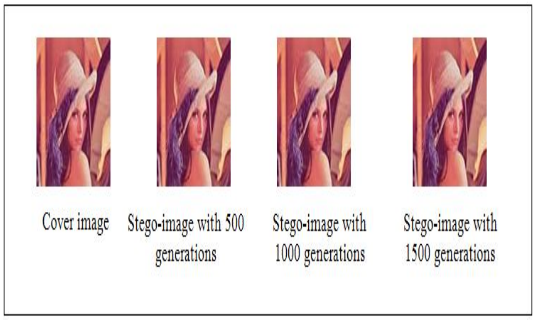

Cover image
Stego-image with 500 generations
Stego-image with 1000 generations
Stego-image with 1500 generations

Figure (4) Lena.bmp before and after hiding process.

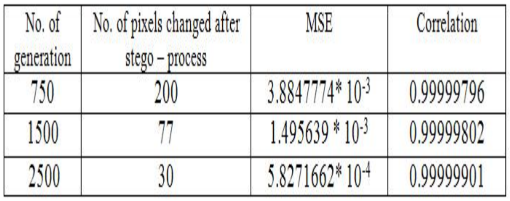
(Sample3): Hide text3 (medium text) in Lena.jpg image (size 131x 131 pixels).

- Applying LSB method and the results are:

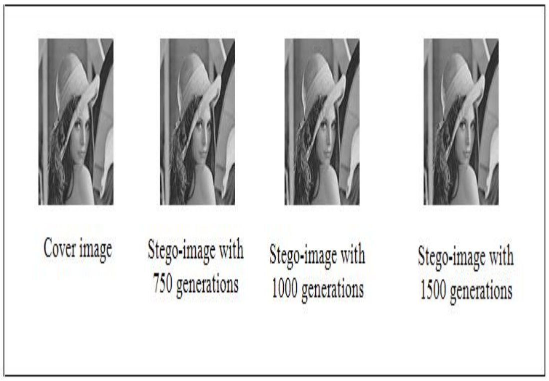
MSE= 2.5192*10 $^{-2}$
Number of pixels changed after stego – process (amount of error) =1297
Correlation=0.99999464

- Appling the improved GA method, and the GA specification are:

Population size=1000
Pc=0.9
Pm=0.01
The tested results are shown in table below, and the cover image and stego-image shown in figure (5).



No. of generation	No. of pixels changed after stego – process	MSE	Correlation
750	200	$3.8847774*10^{-3}$	0.99999796
1500	77	$1.495639*10^{-3}$	0.99999802
2500	30	$5.8271662*10^{-4}$	0.99999901



Cover image

Stego-image with 750 generations

Stego-image with 1000 generations

Stego-image with 1500 generations

Figure (5) Lena.jpg before and after hiding process.

(Sample4): Hide text4 (short text) in CAT.bmp image (size 70x64 pixels)..

- Applying LSB method and the results are:

MSE= 0.02157738

Number of pixels changed after stego – process (amount of error) = 290

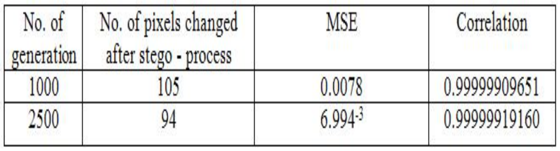
- Appling the improved GA method, and the GA specifications are:

Population size=120

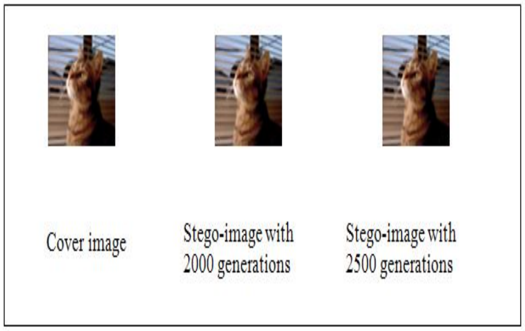
Pc=0.9

Pm=0.01

The tested results are shown in table below, and the cover image and stego-images shown in figure (6).



No. of generation	No. of pixels changed after stego - process	MSE	Correlation
1000	105	0.0078	0.99999909651
2500	94	6.994^{-3}	0.99999919160



Cover image

Stego-image with 2000 generations

Stego-image with 2500 generations

Figure (6) CAT.bmp before and after hiding process.

(Sample5): Hide text5 (medium text) in Mountain.bmp image (size 250x188 pixels).

- Applying LSB method and the results are:

MSE= $2.3*10^{-3}$

Number of pixels changed after stego – process (amount of error) = 325

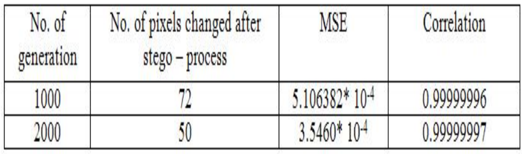
- Applied the improved GA method , and the GA specification are:

Population size=120

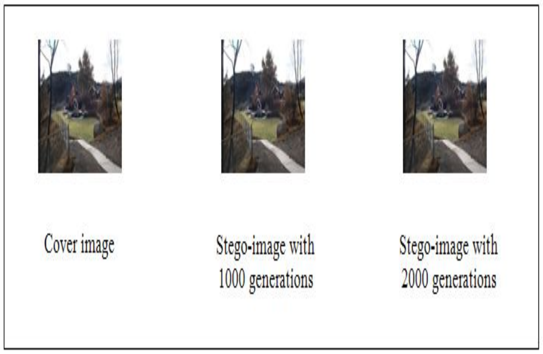
Pc=0.9

Pm=0.01

The tested results are shown in table below, and the cover image and stego-image shown in figure (7).



No. of generation	No. of pixels changed after stego – process	MSE	Correlation
1000	72	$5.106382* 10^{-4}$	0.99999996
2000	50	$3.5460* 10^{-4}$	0.99999997



Cover image

Stego-image with 1000 generations

Stego-image with 2000 generations

Figure (7) Mountain.bmp before and after hiding process.

MSE and correlation measures are used as testing measures to compare between the proposed and LSB hiding methods.

The (weighted) mean squared error between the cover image and the stego-image can be used as one of the measures to assess the relative perceptibility of the embedded message. Imperceptibility takes advantage of human psycho visual redundancy, which is very difficult to quantify.

For colored images:

$$MSE = \frac{1}{m*n}\sum_{i=1}^{m}\sum_{j=1}^{n}\left\|C(i,j)-S(i,j)\right\|^2 \quad .(1)$$

where *m* and *n* are the number of rows and number of columns, respectively, of the cover image, C (*i*, *j*) is the pixel value from the cover image, S (*i*, *j*) is the pixel value from the stego-image [9].

The similarity test is the correlation between the cover-image and stego-image. Correlation is one of the best known methods that evaluates the degree of closeness between two functions. This measure can be used to determine the extent to which the original image and the stego-image are close to each other, even after embedding data.When the stego-image is perceptually similar to the original cover-image; then the correlations equals one [9]. Pearson Correlation Coefficient (Corr) is given by;

$$corr = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(S(i,j)-\overline{S})*(C(i,j)-\overline{C})}{\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}(S(i,j)-\overline{S})^2*\sum_{i=1}^{m}\sum_{j=1}^{n}(C(i,j)-\overline{C})^2}} \quad ..(2)$$

..(2)

Where $\overline{S} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}S(i,j)}{m*n}$ and

$$\overline{C} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}C(i,j)}{m*n}$$

S: stego-image. C: cover-image.

From the results we find that the amount of error decreases and subsequently MSE and correlation measures are decreased with the increasing of both the number of generation and population size then we expect that the amount of error may reache "0", as shown in sample (1) or minimum possible value through increasing the number of generations with excellent image quality.

8. Conclusion

This paper proposes a scheme to hide secret data in image file. The hiding process is based on the idea which improves the hiding process based on GA technique using LSB method so that the secret data is hidden in the best random locations using GA. The best locations are the maximum ability to embed secret data without changing the pixels of the cover image which means minimum amount of error but at the expense of the quality of image. The modification of the best locations by applying LSB method to the location of flipping bit is to satisfy excellent image quality as shown in tested samples used in this paper.

The used samples illustrate the compression between the two hiding methods: the proposed scheme and LSB by using two image tested measures MSE and correlation, and we find from the results the two image measures are improved with increasing both population size and number of generations.

The proposed scheme satisfies best security in extracting secret data and provides high efficiency against attacks which try to discover the hidden message because of using crypto key, so it is impossible to discover the secret data without knowing the crypto key that contains the encryption of the best improved hiding locations in cover image.

As a future work we propose using GA as steganalysis technique to discover the hiding message when it is impossible to get the crypto key.

References

[1] Marin Alvaro, Sapiro Guillermo and Seroussi Gadiel, ***"Is Image Steganography Natural?"***, *IEEE Transactions On Image Processing*, Vol.14, No. 12, December, 2005.

[2] A. Westfeld and A. Pfitzmann. ***"Attacks on steganographic systems"***. 1768 springer 2000, ISBN3-540-67182-X.

[3] Sanjeev Manchanda, Mayank Dave, and S. B. Singh, ***"Customized and Secure Image Steganography Through Random Numbers Logic"***, Signal Processing: An International Journal, Volume 1, 2007.

[4] Shen Wang, Bian Yang and Xiamu Niu, ***"A Secure Steganography Method based on GA"***, Journal of Information Hiding and Multimedia Signal Processing, Volume 1, Number 1, January 2010, Harbin, China.

[5] Tanmay Bhattacharya, Sandeep Bhowmik, and S. R. Bhadra Chaudhuri, " **A Steganographic Approach by Using Session Based Stego-Key, Genetic Algorithm and Variable Bit Replacement Technique"**.

[6] ***A Genetic-Algorithm-Based Approach for Audio Steganography"***, Mazdak Zamani 1, Azizah A. Manaf 2, Rabiah B. Ahmad 3, Akram M. Zeki 4, and Shahidan Abdullah.

[7] Haupt R. L. and Haupt S. E, ***"PRACTICAL GENETIC ALGORITHMS"***, Second edition, John Wiley & Sons, Inc , 2004.

[8] Goldberg, David E. ***"Genetic Algorithms in Search, Optimization and Machine Learning"***, Addison-Wesley Pub., 1989.

[9] A. Abraham, M. Paprzycki and Venkatraman. S, "**Significance of Steganography on Data Security**", Dept. of Computer Science & Engineering, University of Madras, INDIA, Dept. of Computer Science, Oklahoma State University, USA, Proceedings of the International Conference on Information Technology, 2004 IEEE.