High Capacity Data Embedding Method in Image Steganography using Genetic Algorithm

Usha B.A
Assistant Professor
Dept of CSE, R.V.C.E
Bangalore - 560059

N.K Srinath, PhD
Prof and Dean PG Studies
Dept of CSE, R.V.C.E
Bangalore - 560059

Pulkita, Sarthak
Final Year BE
Dept of CSE, R.V.C.E
Bangalore - 560059

ABSTRACT
Steganography is the art of hiding data in a cover medium such that the existence of the data remains secretive and is not easily identifiable. It is known that cryptographic techniques help augment the image steganography process to a great degree. Genetic algorithms (GA) are used in image steganography due to '*adaptive heuristic search'* based on evolutionary biology. This provides us with various challenges such as choice of an encryption algorithm for encrypting the secret message, efficiently generating the search space for the GA by using image data, modelling an efficient fitness function for the GA to evolve into its solutions, and various challenges in handling the embedding of secret data, reading writing image data, producing stego image, and a process to decode the secret information back from the stego image.

Keywords
Cryptography, Genetic algorithm, Image steganography, LSB, Pixel Indexed table

1. INTRODUCTION
Steganography is the art of hiding the information in a cover medium such that it cannot be detected. It's an art and science of invisible communication. Steganography varies from cryptography as in where cryptography concentrates on keeping the substance of a message mystery and safe, steganography bargains on keeping the presence of a message secret.

Steganography and cryptography both are the approaches to shield data from undesirable gatherings yet neither innovation alone is culminated and can be bargained. Once the vicinity of hidden data is uncovered or even suspected, the motivation behind steganography is somewhat broken. [1][2][3]The quality of steganography can along these lines be amplified by consolidating it with cryptography.

Two sorts of technologies that are nearly related with steganography are watermarking [4] and fingerprinting [5]. These advancements are fundamentally utilized as a part of the assurance of licensed innovation; subsequently the calculations have distinctive necessities than steganography. In watermarking and fingerprinting the way that data is covered up inside the records may be open knowledge – now and then it may even be visible while in the steganography the mystery of the data is critical.

2. FUNDAMENTALS OF GENETIC ALGORITHM
Genetic Algorithms are modelled on evolutionary biology for search space optimization. Every generation comprises of a population that closely resemble the chromosome that we see in our DNA. Every individual speaks to a point in a pursuit space and a conceivable arrangement. The individuals in the population are then made to experience a process of development (evolution).

Genetic Algorithm is based on natural theory of evolution from Darwin's premise "the survival of the fittest". In this theory Sir Charles Darwin concludes that only the fittest individuals survive and reproduce to form a next generation.

The Thought of developmental figuring was presented in the 1960s by I. Rechenberg in his work "Development techniques" (Evolution strategies in unique). His thought was then grown by different scientists. Genetic Algorithm (GAs) were designed by John Holland and grew by him and his understudies and partners. This lead to Holland's book "Adaption in Natural and Artificial Systems" distributed in 1975.

In 1992 John Koza has utilized genetic algorithm to develop program to perform certain undertakings. He called his technique "Genetic programming" (GP). Less programs were utilized, in light of the fact that programs in this dialect can communicated as a "parse tree", which is the object the GA [6] deals with.

The GA keeps up a populace of n chromosomes (solution) with related wellness values. Folks are chosen to mate, on the premise of their wellness, creating posterity by means of a regenerative arrangement. Thus profoundly fit arrangements are given more chances to recreate, so posterity acquires qualities from every guardian. As folks mate and produce posterity, room must be made for the fresh introductions since the populace is kept at a static size. People in the populace pass on and are supplanted by the new arrangements, inevitably making another era once every single mating opportunities in the old populace have been depleted. Thusly it is trusted that over progressive eras better arrangements will flourish while the minimum fit arrangements vanish.

New eras of solution are delivered containing, overall, more great qualities than a normal solution in a past generation. Every progressive era will contain more great `partial solution' than past eras. In the long run, once the populace has met and is not delivering posterity recognizably unique in relation to those in past generation, the algorithm itself is said to have united to a solution of answers for the current issue.

3. PROPOSED METHOD
The study for the Steganography [7][8] project starts from identifying and defining the software components of the architecture to put together a viable framework for steganography using genetic algorithms. After considering the different architectural patters, we chose to build the Steganography project using a component based architecture.

We identified that the project could be de-composed into modules and to develop each module in isolation, independent of the other module would very much suit the design, development and testing of the steganography project is shown in figure 3.1.
The key software components / modules that are identified as follows:

1. User interface components
2. Image handler component
3. Cryptography component
4. LSB insertion component
5. GA component
6. Image testing component
7. Results and reporting component

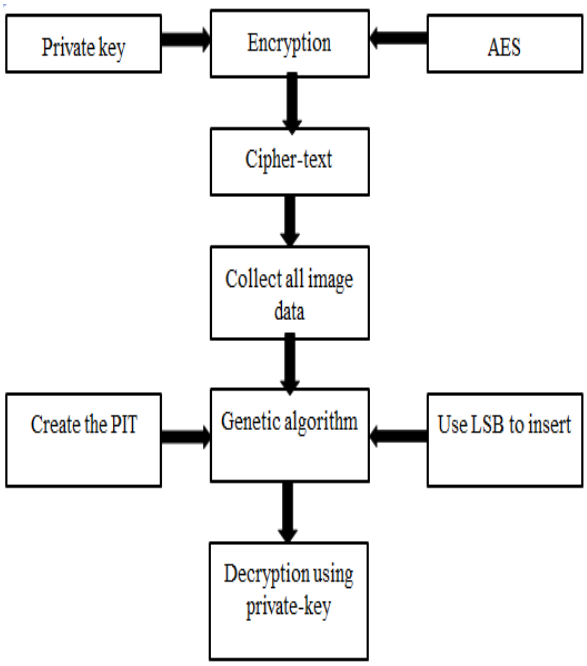

Figure 3.1.Proposed System Model

3.1 Algorithm Description

GAs is in light of a relationship with the genetic structure and conduct of chromosomes inside a population of individuals utilizing the accompanying establishments:

Individuals in a population vie for assets and mates.

Those individual best in every "rival" will deliver more posterity than those individual that perform ineffectively.

The outlines of the Genetic Algorithm are:-

1. [start]Generate random population of n chromosome (suitable solution for the problem).
2. [Fitness] Evaluate the fitness f(x) of each chromosome x in the population.
3. [New population] Create new population by repeating following steps until new population is completed.

-[Selection] Select two guardian chromosomes from a populace as per their wellness (the better fitness function, the greater opportunity to be chosen)

-[Crossover] With a hybrid likelihood traverse the folks to frame another posterity (children). In the event that no crossover was performed, posterity is a careful duplicate of parents.

-[Mutation] With a mutation there is probability of transform new posterity at every locus (position in chromosome).

-[Accepting] Place new posterity in another populace.

4. [Replace] Utilize new created populace for a further run of algorithm.
5. [Test] In the event that the end condition is fulfilled, stop, and return the best solution in current population.
6. [Loop]Go to step 2.

4. EXPERIMENTAL RESULTS AND ANALYSIS

This section list the result of the project and the inference to be made from the testing result. The evalution metrics have been listed and the results have been accordingly quantified as shown in table 4.1, 4.2 and 4.3.

Metrics help gaging the advancement, quality and wellbeing of a software test exertion. Measurements can likewise be utilized to assess past execution, current status and conceive future patterns. Successful measurements are straightforward, objective, quantifiable, and significant and have effectively open hidden information.

For making a basic and important quantifiable result, our assessment measurements are on the lines of MSE, PSNR and Execution. PSNR is utilized just in light of the fact that it is ordered under the distinction mutilation metrics.

Measuring software for performance should involve an investigation through enough to check and verify every stage of its operations and to identify the bottlenecks for performance. Once these bottlenecks are identified, suitable measures has to be taken to overcome the underlying issues which are giving way to the bottlenecks, such that the systems overall performance is improved. This process has to be re-iterated whenever the following components change

- Lines of code increases significantly: This is giving us an idea that more unit tests should be performed to deduce conclusive results of performance and correctness.
- Unit test failures: These are conclusive evidence that some of the modules are not operating as desired. Fixing these will immediately ensure high performance.
- Number of bugs
- Actual time to finish the encoding and decoding as the input size of secret message increases. This is conclusive of measuring time complexity with big O notation.

Table 4.1: Performance Table

Sl.No	Secret Message Size (bytes)	Image Dimension	Encoding Time (seconds)	Decoding Time (seconds)
1	26	64 x 64	3.02	0.32
2	415	64 x 64	6.15	0.73
3	1024	64 x 64	29.12	2.5
4	26	128 x 128	3.08	0.37
5	415	128 x 128	15.41	2.1
6	1024	128 x 128	200.02	8.05
7	26	512 x 512	60.02	5
8	415	512 x 512	285.13	12.02
9	1024	512 x 512	600.23	20.05

Table 4.2: Stego analysis table

Sl. No.	Secret Message Size (bytes)	Image Dimension	MSE	PSNR (db)
1	26	64 x 64	0	Infinity
2	415	64 x 64	1	48.13
3	1024	64 x 64	10	38.13
4	26	128 x 128	0	Infinity
5	415	128 x 128	1	43.55
6	1024	128 x 128	0	Infinity
7	26	512 x 512	0	Infinity
8	415	512 x 512	10	46.32
9	1024	512 x 512	0	Infinity

Table 4.3: Pixel Index Table Readings V/s PSNR

Sl. No	Image Size	Total Solutions Required (as per the secret text)	Exact Solutions Found By GA	Percentage of GA solution to Required Solution	Maximum Delta Recorded in PIT (Absolute)	PSNR of Original v/s Stego Image
1	64 x 64	31	31	100.00%	0	Infinity
2	64 x 64	200	183	92.00%	21	46.4
3	64 x 64	415	380	92.00%	40	42.3

From the above table, we conclude that whenever there is a Maximum Delta of zero in the Pixel Index Table, The GA has managed to find an exact solution and MSE will be 0, Hence PSNR will be infinity. One Such instance is shown below for Sl.No 1

(Actual Readings from the PIT File)

[X, Y, Delta Red, Delta Green, Delta Blue]

1, 44,-1000, 0,-1000
54, 59, 0,-1000,-1000
34, 41,-1000, 0,-1000
56, 63,-1000,-1000, 0
7, 20,-1000, 0,-1000
49, 23,-1000, 0,-1000
6, 22, 0,-1000,-1000
5, 32,-1000,-1000, 0
52, 39, 0,-1000,-1000
27, 0,-1000, 0,-1000
53, 26, 0,-1000,-1000
2, 14,-1000, 0,-1000
59, 52, 0,-1000,-1000
23, 26,-1000, 0,-1000
35, 44, 0,-1000,-1000
45, 17,-1000, 0,-1000
28, 46,-1000, 0,-1000
13, 18, 0,-1000,-1000
13, 23,-1000, 0,-1000
4, 15, 0,-1000,-1000
6, 3, 0,-1000,-1000
6, 21, 0,-1000,-1000
7, 38, 0,-1000,-1000
28, 40,-1000, 0,-1000
2, 17, 0,-1000,-1000
58, 36,-1000,-1000, 0
48, 49, 0,-1000,-1000
13, 26,-1000, 0,-1000
26, 39,-1000, 0,-1000
28, 60, 0,-1000,-1000
4, 11, 0,-1000,-1000
13, 15,-1000, 0,-1000

It is clearly evident that the GA has managed to find the best solution and that the maximum difference of any pixel is zero compared to the original image. This results in totally identical stego Image with a PSNR of Infinity.

In all three circumstances, a 64 x 64 bit bmp image was used as the cover medium. GA's solution is closer to 100% when the required solution is small and tends to get to 100% when required as shown in Figure 4.1 and 4.2.

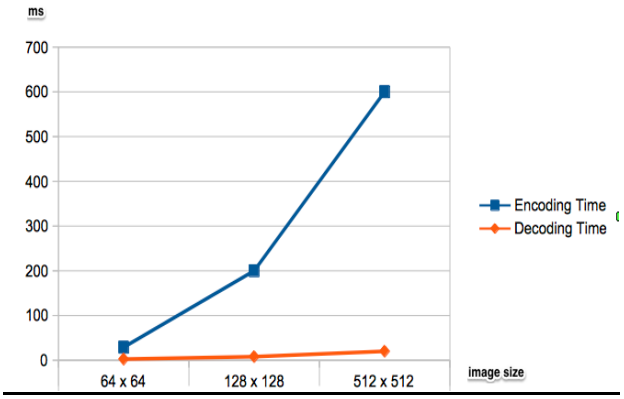

Figure 4.1: Graph showing the Encoding Time and Decoding Time in (milliseconds) for different image sizes

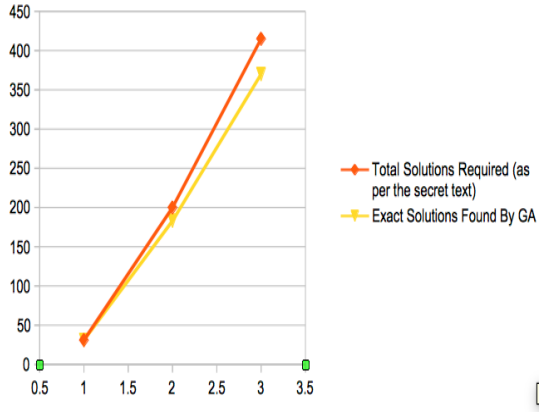

Figure 4.2 Graph showing the growth of GA solutions found vs Total solutions required.

5. CONCLUSION AND FUTURE WORK

As per our analysis, we conclude that we have successfully achieved high capacity in hiding data in the bmp image. The results obtained were conclusive enough for us to fit a text document, image into a cover image and at the same time achieve very less distortion of the Stego image. However, we have noticed that with a very high image, the performance decreases even when the input size is less. This boils down to the issue of optimizing the ranking algorithm and fitness functions for high throughput. With this, we conclude that the project established a good technique to achieve high capacity and less distortion of the Stego image. Following are points for further enhancements

- To Improving the granularity in the GA Researchers can consider a different encoding mechanism for the GA, a binary encoding mechanism will achieve higher granularity in terms of defining the problem and evolving towards it., although it can be noted that this might present us with the problem of performance, so we need to re-consider the time complexities of our fitness function and our ranking selection methods.
- Hiding the PIT as part of the Stego Image

  If we can accomplish the task of hiding the pit data inside the Stego image, we will not be required to transmit the PIT to the receiver in the future and the decoder module will just have the dependency of the stego image & not the stego image and the pit together.
- Consider using different symmetric encryption algorithms

  Given the key length constraint of Java's AES encryption, we can try out different symmetric encryption algorithms in the future.
- Use multi-threaded code in UI to display dynamically updating logs like the console application.

6. ACKNOWLEDGMENT

It is our privilege to acknowledge thanking all the department personals and sponsors who gave us an opportunity to present a paper at this level. We wish to place our deep sense of gratitude to all reference papers authors for their beneficial papers, books and websites etc.

7. REFERENCES

[1] G.Prema, S.Natarajan - "Steganography using Genetic Algorithm along with Visual Cryptography for Wireless Network Application", International Conference On Information communication and embedded systems.Shivakasi , India. Vol 5 pp 727-730, 2013

[2] Ahmed,A, Agarwal,N, Banerjee S - "Image Steganography By Closest Pixel-pair Mapping", International Conference on Advances in Computing, Communications and Informatics, India. pp 200-221, Sept 2014.

[3] Prasenjit Das, NirmalaKar - "A DNA Based Image Steganography using 2D Chaotic Map", International Conference On Electronics and Communication Systems ,Agartala, India. pp 1-5, 2014.

[4] Yafeng Z, - "A Study of influence between digital watermarking and steganography", International Conference on Wavlet Analysis and Pattern Recogition, IEEE,Vol 7,pp 331-339 2013.

[5] Whitelam, C , Osia,N ,Bourlai,T - "Securing multimodal biometric data through watermarking and steganography.", International Conference On Technologies For Homeland Security, IEEE, Vol 13, pp 22-28,May 2013.

[6] Jyoti, Sabir,"Genetic Algorithm Based Image Steganography for Enhancement of Concealing Capacity and Security", Internal Journal Image, Graphics and Signal Processing, Vol 7, Issue 18-25, pp 305-309, 2013

[7] Shen Wang, Bian Yang and Xiamu Niu - "A Secure Steganography Method Based On Genetic Algorithm", Journal Of Information Hiding and Multimedia Signal Processing, Harbin, China. Vol 27. pp 171-182, January 2010.

[8] Artz, D., "Digital Steganography: Hiding Data within Data", IEEE Internet Computing Journal, Vol 13, pp 58-68, June 2001.

[9] Provos, N. & Honeyman, P., "Hide and Seek: An introduction to steganography", IEEE Security and Privacy Journal, pp 1613-1626, March 2003.