



Jaypee University of Information Technology
Solan (H.P.)
LEARNING RESOURCE CENTER

Acc. Num.

Call Num:

General Guidelines:

- ◆ Library books should be used with great care.
- ◆ Tearing, folding, cutting of library books or making any marks on them is not permitted and shall lead to disciplinary action.
- ◆ Any defect noticed at the time of borrowing books must be brought to the library staff immediately. Otherwise the borrower may be required to replace the book by a new copy.
- ◆ The loss of LRC book(s) must be immediately brought to the notice of the Librarian in writing.

Learning Resource Centre-JUIT



SP03161

DATA HIDING IN AN IMAGE

By

ADITY SHARMA - 031013

ANOO AGARWAL - 031114



MAY-2007

Submitted in partial fulfillment
of
the Degree of Bachelor of Technology

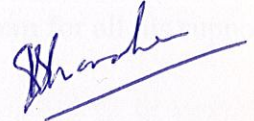
DEPARTMENT OF ELECTRONICS AND
COMMUNICATION
JAYPEE UNIVERSITY OF INFORMATION
TECHNOLOGY - WAKNAGHAT

CERTIFICATE

This is to certify that the work entitled, “**DATA HIDING IN AN IMAGE**” submitted by **ADITY SHARMA AND ANOO AGARWAL** in partial fulfillment for the award of degree of Bachelor of Technology in Electronics and Communication Engineering of Jaypee University of Information Technology has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.



(Prof. Vinay Kumar)




(Dr. Sunil V. Bhooshan)


ACKNOWLEDGEMENT

The success of any project depends largely on the encouragement and guidelines of many others. Therefore we take this opportunity to express our sincere gratitude to the people who have been instrumental in the successful completion of the project. We would like to express our sincere appreciation and gratitude to our guide **Prof. Vinay Kumar** without whose able guidance, tremendous support and continuous motivation the project would not have been carried to perfection. We sincerely thank him for spending all his valuable time and energies during the execution of project.

We take this opportunity to express our sincere gratitude to the Head of Department (Electronics and Communication Engineering) **Dr. Sunil V. Bhooshan** for all his support and valuable inputs.

The successful compilation of final year project depends on the knowledge and attitude inculcated in the total length of course. So we want to express our sincere gratitude to all the faculties who taught us during the four years of B.Tech.


(Adity Sharma)


(Anoo Agarwal)

CONTENTS

List of figures	1
List of abbreviations	2
Abstract	3
1 Introduction	4
1.1 Steganography	4
1.2 Types of steganography	5
1.3 Properties of hiding schemes	6
2 Brief resume of the existing steganography techniques	8
2.1 Historical steganography techniques	8
2.2 Modern steganography techniques.....	9
2.2.1 Substitution techniques	9
2.2.2 Transform Domain techniques	9
2.2.3 Spread Spectrum techniques.....	10
2.2.4 Statistical techniques.....	10
2.2.5 Distortion techniques.....	10
2.2.6 Cover Generation techniques.....	10
2.2.7 Masking and Filtering.....	11
2.3 Applications	11
2.4 Advantages and Disadvantages	12
3 Developed technique	14
3.1 Proposed algorithm.....	14
3.2 Hiding principle.....	14
3.2.1 Encryption algorithm.....	14
3.3 Retrieval principle	15
3.3.1 Decryption algorithm.....	15
3.4 Advantages of developed technique.....	15
3.5 Block diagram for the developed technique.....	16

LIST OF CONTENTS

4 Experimental results 17
4.1 Results for BMP image 17
4.2 Results for GIF image..... 18
4.3 Results for TIFF image 19
4.4 Histogram plots..... 21
5 Conclusion 23
6 Bibliography 24

LIST OF FIGURES

- Figure 1.1 Types of steganography.
- Figure 1.2 Steganography with encryption.
- Figure 1.3 Comparison of secret communication techniques.
- Figure 1.4 The magic triangle.
- Figure 3.1 Block diagram for the developed technique.
- Figure 4.1.1 Original BMP image
- Figure 4.1.2 Modified BMP image.
- Figure 4.2.1 Original GIF image.
- Figure 4.2.2 Modified GIF image.
- Figure 4.3.1 Original TIFF image.
- Figure 4.3.2 Modified TIFF image.
- Figure 4.4.1 Histogram of BMP image without data.
- Figure 4.4.2 Histogram of BMP image with embedded data.
- Figure 4.4.3 Histogram of error.

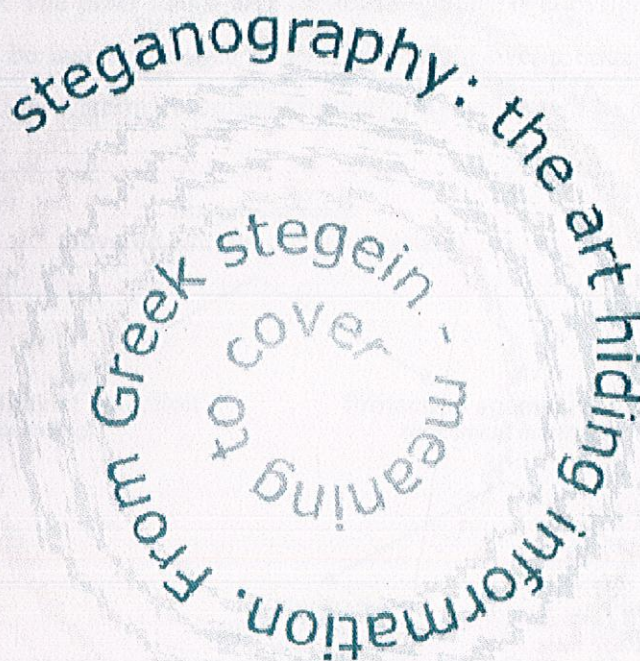
LIST OF ABBREVIATIONS

1. A/D Analog to Digital.
2. D/A Digital to Analog.
3. BMP Bitmap.
4. JPEG Joint Photographic Expert's Group.
5. LSB Least Significant Bit.
6. DCT Discrete Cosine Transform.
7. SNR Signal to Noise Ratio.
8. GIF Graphic Interchange Format
9. TIFF Tagged Image File format

ABSTRACT

In this project, data hiding in digital image is proposed to embed high volume of data and facilitate its secret and secure communication by using cryptography techniques without taking any file format into consideration. An effective algorithm for this process has been presented, which hides the information in an image without producing any noticeable distortions. The coding is done using file handling in C. The software developed has been tested on different image formats and yielded satisfactory results.

1.1 Steganography



steganography: the art hiding information.
From Greek stegan - meaning to cover

Pertaining to the requirement of privacy and security, there has been a lot of progress in the field of data hiding and numerous new algorithms are being developed enhancing the science of hiding data in recent years.

Steganography simply takes one piece of information and hides it within another. Steganography can be viewed as akin to cryptography. Both have been used throughout recorded history as means to protect information. At times these two technologies seem to converge while the objectives of the two differ. Cryptographic techniques "scramble" messages so if intercepted, the messages cannot be understood. Steganography, in an essence, "camouflages" a message to hide its existence and make it seem "invisible" thus concealing the fact that a message is being sent

altogether. An encrypted message may draw suspicion while an invisible message will not.

1.2 Types of steganography

Figure 1 shows how information hiding can be broken down into different areas. Steganography can be used to hide a message intended for later retrieval by a specific individual or group. In this case the aim is to prevent the message being detected by any other party. The other major area of steganography is copyright marking, where the message to be inserted is used to assert copyright over a document. This can be further divided into watermarking and fingerprinting which will be discussed later.

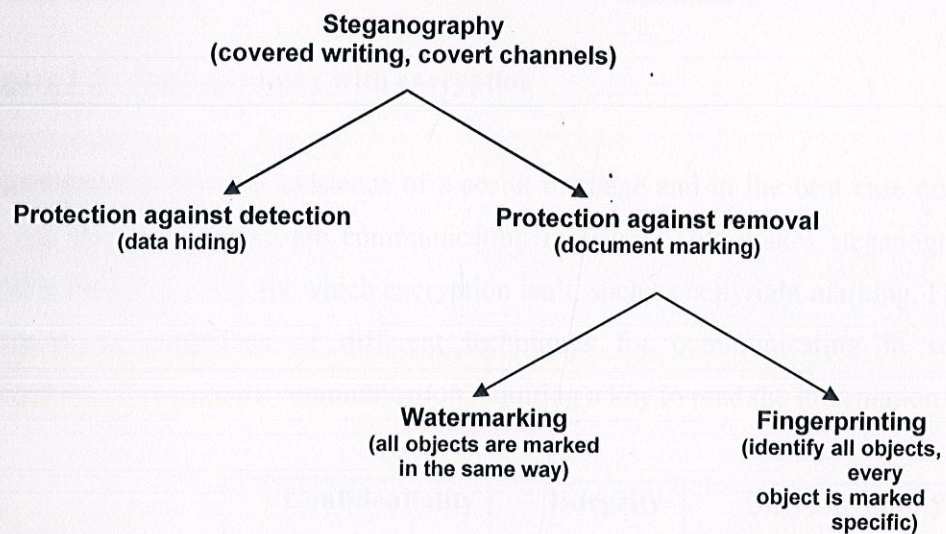


Figure 1.1 : Types of steganography.

Steganography and encryption are both used to ensure data confidentiality. However the main difference between them is that with encryption anybody can see that both parties are communicating in secret.

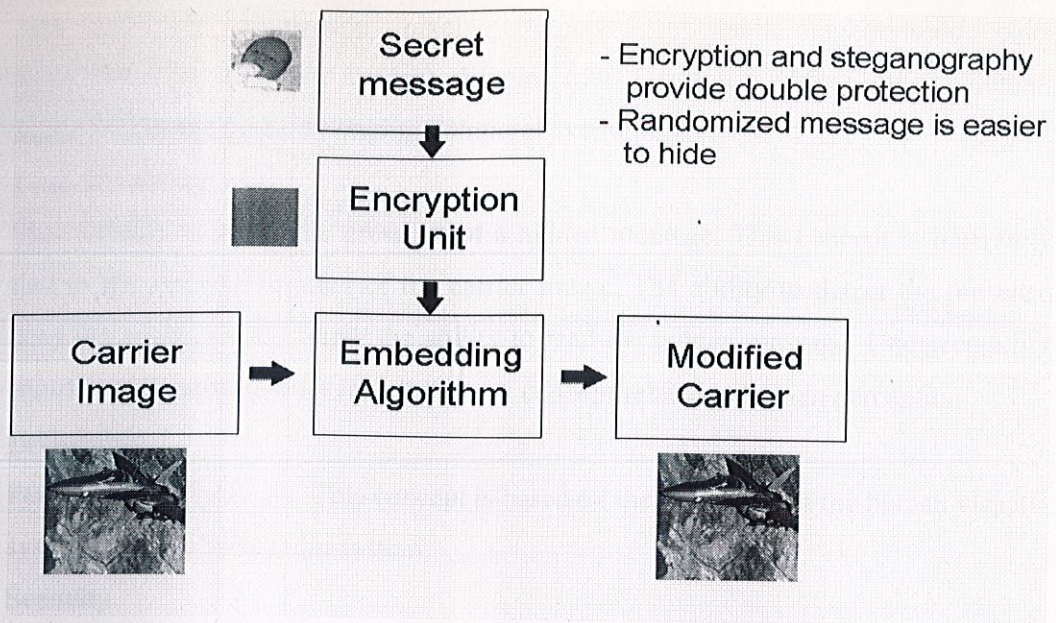


Figure 1.2 : Steganography with encryption

Steganography hides the existence of a secret message and in the best case nobody can see that both parties are communicating in secret. This makes steganography suitable for some tasks for which encryption isn't, such as copyright marking. Figure 2 shows a comparison of different techniques for communicating in secret. Encryption allows secure communication requiring a key to read the information.

	Confidentiality	Integrity	Unremovability
Encryption	Yes	No	Yes
Digital Signatures	No	Yes	No
Steganography	Yes / No	Yes / No	Yes

Figure 1.3 : Comparison of secret communication techniques.

1.3 Properties of hiding schemes

Robustness

The ability to extract hidden information after common image processing operations: linear and nonlinear filters, lossy compression, contrast adjustment, recoloring, resampling, scaling, rotation, noise adding, cropping, printing / copying / scanning,

D/A and A/D conversion, pixel permutation in small neighborhood, color quantization (as in palette images), skipping rows / columns, adding rows / columns, frame swapping, frame averaging (temporal averaging), etc.

Undetectability

Impossibility to prove the presence of a hidden message. This concept is inherently tied to the statistical model of the carrier image. The ability to detect the presence does not automatically imply the ability to read the hidden message. Undetectability should not be mistaken for invisibility – a concept related to human perception.

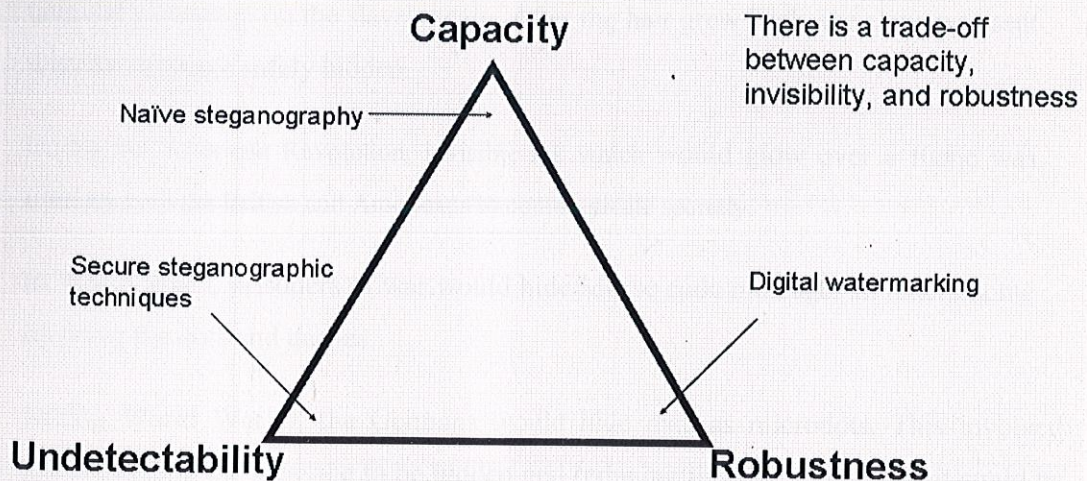
Invisibility

Perceptual transparency. This concept is based on the properties of the human visual system or the human audio system.

Security

The embedded information cannot be removed beyond reliable detection by targeted attacks based on a full knowledge of the embedding algorithm and the detector (except a secret key), and the knowledge of at least one carrier with hidden message.

The "Magic" Triangle



Additional factors: • Complexity of embedding / extraction
• Security

Figure 1.4 : The magic triangle

BRIEF RESUME OF EXISTING STEGANOGRAPHY TECHNIQUES

2.1 Historical Steganographic Techniques

Early steganography was messy. Before phones, before mail, before horses, messages were sent on foot. If you wanted to hide a message, you had two choices: have the messenger memorize it, or hide it on the messenger. In fact, the Chinese wrote messages on silk and encased them in balls of wax. The wax ball, "la wan," could then be hidden in the messenger.

Although the term steganography was only coined at the end of the 15th century, the use of steganography dates back several millennia. In ancient times, messages were hidden on the back of wax writing tables, written on the stomachs of rabbits, or tattooed on the scalp of slaves. Invisible ink has been in use for centuries-for fun by children and students and for serious espionage by spies and terrorists.

Herodotus, an entertaining but less than reliable Greek historian, reports a more ingenious method. Histaeus shaved the head of his most trusted slave, then tattooed a message on the slave's scalp. After the hair grew back, the slave was sent with the message safely hidden.

During the American Revolution, invisible ink which would glow over a flame was used by both the British and Americans to communicate secretly.

In World War I, prisoners of war would hide Morse code messages in letters home by using the dots and dashes.

During World War II, the Germans would hide data as microdots. This involved photographing the message to be hidden and reducing the size so that that it could be used as a period within another document.

Today, thanks to modern technology, steganography is used on text, images, sound, signals, and more.

2.2 Modern steganographic techniques

2.2.1 Substitution Techniques:

The least significant bit insertion method is probably the most well known image steganography technique. It is a common, simple approach to embedding information in a graphical image file. Unfortunately, it is extremely vulnerable to attacks, such as image manipulation. A simple conversion from a GIF or BMP format to a lossy compression format such as JPEG can destroy the hidden information in the image. When applying LSB techniques to each byte of a 24-bit image, three bits can be encoded into each pixel. (As each pixel is represented by three bytes.) Any changes in the pixel bits will be indiscernible to the human eye. For example, the letter A can be hidden in three pixels. Assume the original three pixels are represented by the three 24-bit words below:

```
(00100111 11101001 11001000 ) ( 00100111 11001000 11101001 ) ( 11001000  
00100111 11101001 )
```

The binary value for the letter A is (10000011). Inserting the binary value of A into the three pixels, starting from the top left byte, would result in:

```
(00100111 11101000 11001000 ) ( 00100110 11001000 11101000 ) ( 11001000  
00100111 11101001 )
```

The emphasized bits are the only bits that actually changed. The main advantage of LSB insertion is that data can be hidden in the least and second to least bits and still the human eye would be unable to notice it. When using LSB techniques on 8-bit images, more care needs to be taken, as 8-bit formats are not as forgiving to data changes as 24-bit formats are. Care needs to be taken in the selection of the cover image, so that changes to the data will not be visible in the stego-image.

2.2.2 Transform Domain Techniques:

Embed secret message in a transform space (e.g. frequency domain) of cover.

Example: Steganography in the Discrete Cosine Transform (DCT) domain .Split the cover image into 8*8 blocks. Each block is used to encode one message bit. Blocks are chosen in a pseudorandom manner. The relative size of two predefined DCT coefficients is modulated using the message bit. The two coefficients are chosen from middle frequencies (trade off between robustness and imperceptibility).

2.2.3 Spread Spectrum Techniques:

Adopt ideas from spread spectrum communication where a signal is transmitted in a bandwidth in excess of the minimum necessary to send the information. In other words, the message is spread over a wide frequency bandwidth. The SNR in every frequency band is small (difficult to detect). Even if parts of the message are removed from several bands, enough information is present in other bands to recover the message. Thus, it is difficult to remove the message completely without entirely destroying the cover (robustness).

2.2.4 Statistical Techniques:

Encode information by changing several statistical properties of a cover. The cover is split into blocks. Each block is used to hide one message bit. If the message bit is "1" then the cover block is modified, otherwise the cover block is not modified. Difficult to apply in many cases, since a good test must be found which allows distinction between modified and unmodified cover blocks.

2.2.5 Distortion Techniques:

Store information by signal distortion. The encoder applies a sequence of modifications to the cover. This sequence corresponds to the secret message. The decoder measures the differences between the original cover and the distorted cover to detect the sequence of modifications and consequently recover the secret message. Not useful in many applications since the decoder must have access to the original cover. Example: vary the distance between consecutive lines or words to transmit secret information.

2.2.6 Cover Generation Techniques:

Encode information in the way a cover is generated Example: Automated Generation of English Text. Use a large dictionary of words categorized by different types, and a style source which describes how words of different types can be used to form a meaningful sentence. Transform message bits into sentences by selecting words out of the dictionary which conforms to a sentence structure given in the style source.

2.2.7 Masking and filtering

Masking and filtering techniques hide information by marking an image in a manner similar to paper watermarks. Because watermarking techniques are more integrated into the image, they may be applied without fear of image destruction from lossy compression. By covering, or masking a faint but perceptible signal with another to make the first non-perceptible, we exploit the fact that the human visual system cannot detect slight changes in certain temporal domains of the image. Technically, watermarking is not a steganographic form. Strictly, steganography conceals data in the image; watermarking extends the image information and becomes an attribute of the cover image, providing license, ownership or copyright details. Masking techniques are more suitable for use in lossy JPEG images than LSB insertion because of their relative immunity to image operations such as compression and cropping.

2.3 Applications

1. Most of the newer applications use steganography like a watermark, to protect a copyright on information. Photo collections, sold, on CD, often have hidden messages in the photos which allow detection of unauthorized use. The same technique applied to DVDs is even more effective, since the industry builds DVD recorders to detect and disallow copying of protected DVDs.
2. Even biological data, stored on DNA, may be a candidate for hidden messages, as biotech companies seek to prevent unauthorized use of their genetically engineered material. The technology is already in place for this: three New York researchers successfully hid a secret message in a DNA sequence and sent it across the country. Sound like science fiction? A secret message in DNA provided *Star Trek's* explanation for the dubious fact that all aliens seem to be humans in prosthetic makeup!
3. Unobtrusive communication-Military and intelligence agencies, Criminal.
4. Plausible deniability- Fair voting, personal privacy, Limitation of liability.

5. Anonymous communication -Vote privately, Make political claims, Access censored material, Preserve free speech.

2.4 Advantages and Disadvantages:

Steganography is beneficial for securely storing sensitive data, such as hiding system passwords or keys within other files. However, it can also pose serious problems because it's difficult to detect. Network surveillance and monitoring systems will not flag messages or files that contain steganographic data. Therefore, if someone attempted to steal confidential data, they could conceal it within another file and send it in an innocent looking email.

The advantage of steganography is that it can be used to secretly transmit messages without the fact of the transmission being discovered. Often, using encryption might identify the sender or receiver as somebody with something to hide.

Further, steganography can be used to tag notes to online images (like post-it notes attached to paper files).

However, steganography has a number of disadvantages as well. Unlike encryption, it generally requires a lot of overhead to hide a relatively few bits of information. However, there are ways around this. Also, once a steganographic system is discovered, it is rendered useless. This problem, too, can be overcome if the hidden data depends on some sort of key for its insertion and extraction.

Another limitation is due to the size of the medium being used to hide the data. In order for steganography to be useful the message should be hidden without any major changes to the object it is being embedded in. This leaves limited room to embed a message without noticeably changing the original object. This is most obvious in compressed files where many of the obvious candidates for embedding data are lost. What is left is likely to be the most perceptually significant portions of the file and although hiding data is still possible it may be difficult to avoid changing the file.

Although many of the uses of steganography are perfectly legal, it can be abused by certain groups. The potential exists for terrorist groups to communicate using these

techniques to hide their messages and rumours persist that Al-Qaeda have used it to communicate. Also of concern is that these techniques may be used by paedophiles to hide pornographic images within seemingly innocuous material.

As a result the need for detection of steganographic data has become an important issue for law enforcement agencies. Attempting to detect the use of steganography is called steganalysis and can be either passive, where the presence of the hidden data is detected, or active, where an attempt is made to retrieve the hidden data.

DEVELOPED TECHNIQUE

3.1 Proposed algorithm

The proposed algorithm is based on appending the data at the end of the image using file handling. Data hiding is used in conjunction with cryptography so that the information is doubly protected, first it is encrypted and then hidden so that an adversary has to first find the information and then has to decrypt it.

3.2 Hiding principle

The text to be hidden is first encrypted using a secret key that is shared between sender and receiver. This cipher text is then appended at the end of the image file after a special string (marker) indicating the start of the encrypted hidden text.

3.2.1 Encryption Algorithms

1. The data to be hidden is first incremented by the value of the secret key.
eg. If the alphabet is a and key is 6 then it will be stored as g.
2. The data to be hidden is stored in an array. The secret key is then added to each character. Since key will be an integer and data can be an integer, character or some special character, we first have to take the integer equivalent of the data, perform the increment operation and then store it back in the character array. The value of the data will be increased by the key. For e.g. If the character is "a" and key is 7 the encrypted data will be h. The data in the array is then rearranged i.e. we will first take all the even position elements and then the odd position elements. Then these will be appended at the end of the file.

NOTE: It is preferred to use the key greater than 26 so that the data hidden moves out of the range of alphabets and could not be detected when image is opened in notepad or word pad.

Data to be hidden: THIS IS OUR FINAL PROJECT

Data appended at the end of the image file: TI SORFNLPOETHSI U IA RJC

(if the key entered is 0) and if the key entered is 34 all the characters will be shifted by 34.

3.3 Retrieval principle

The retrieval process uses the knowledge of the secret key and the marker. The image file contents are read until marker is encountered, which marks the beginning of the encrypted hidden data. This data is decrypted using the priority knowledge of the key and is written into a text file. Absence of hidden data is signaled by using an indicator which returns false value when the marker could not be found in the image. This is useful in the way that if someone does not have the knowledge of the maker he/she cannot retrieve the data.

3.3.1 Decryption algorithms

1. The encrypted data obtained from the image is decremented by the value of the secret key shared between sender and receiver to obtain the required hidden message.
2. The encrypted data hidden in the image is stored in an array. We will then use the shared key to get back the original jumbled up characters. These characters are then rearranged. There will be two counters. First will be initialized to 0 and second to the half of the length of the data plus 1. We will keep on reading data until the file is exhausted with the help of these two counters and obtain the required information. A special care has to be taken for files containing even and odd number of data items.

Data before rearranging: TI SORFNLPOETHSI U IA RJC

Counter1 = 0

Counter2 = (25/2)+1 = 13

First we will write the 0th element then 13th element then 1st then 14th and so on.

Retrieved data will be : THIS IS OUR FINAL PROJECT

3.4 Advantages of developed technique:

1. Supports all image file formats.
2. Relatively large volume of data can be embedded in comparison to other data hiding techniques.
3. Image distortion is negligible.
4. Information is protected since it is first encrypted and then hidden.
5. Marker provides one more level of security.

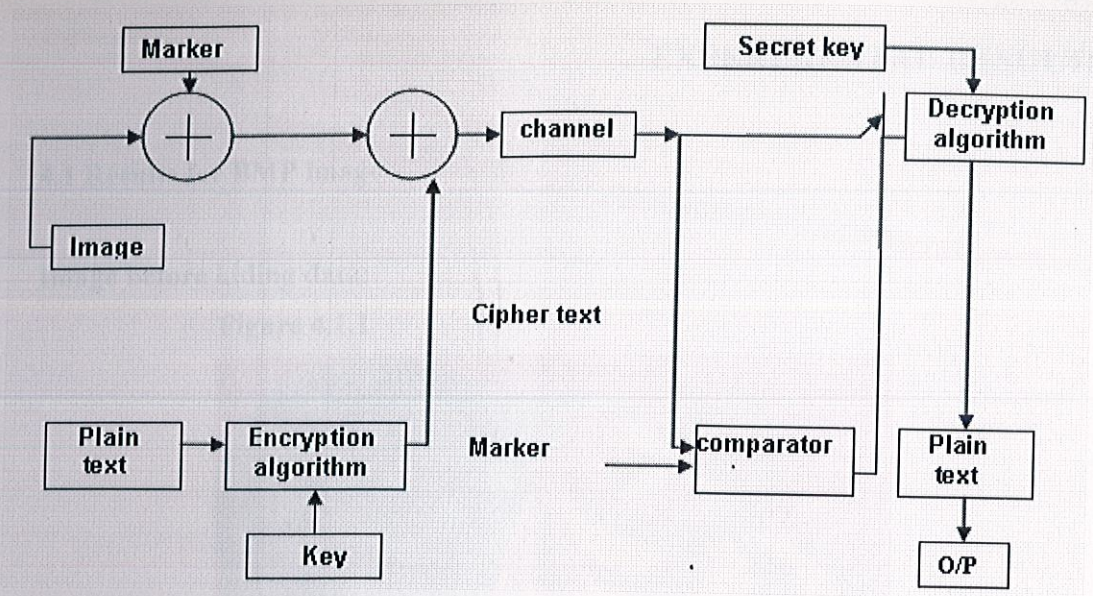


Figure 3.1 Block Diagram for the developed technique.

EXPERIMENTAL RESULTS

4.1 Results for BMP image

Image before hiding data:

Figure 4.1.1



Image with hidden data:

Figure 4.1.2



4.2 Results for GIF image

Image before hiding data:

Figure 4.2.1



Image with hidden data:

Figure 4.2.2



4.3 Results for TIFF image

Image before hiding data:

Figure 4.3.1



Image with hidden data:

Figure 4.3.2



Can you find what's the difference in the pairs of the images shown?

You might argue that both the images are same, but there is a Difference.

The difference is that **second image of each format contains complete C program of the developed technique.**

4.4 Histogram Plots

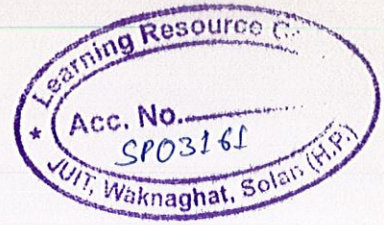


Figure 4.4.1 Histogram of BMP image without data:

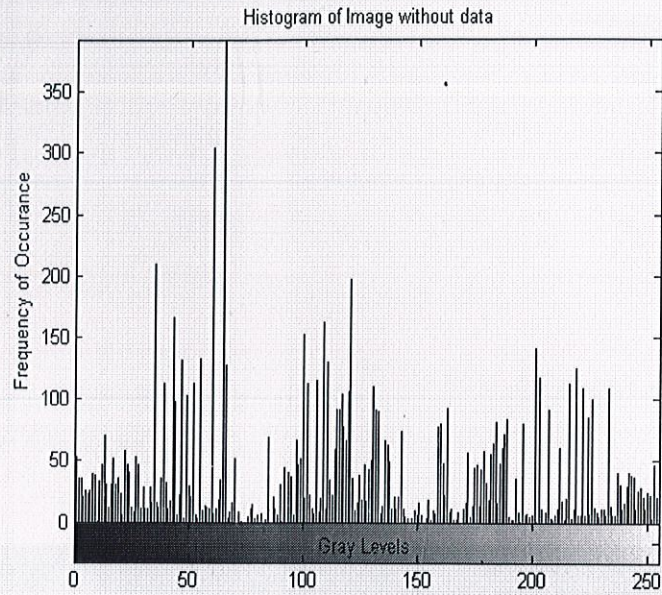


Figure 4.4.2 Histogram of BMP image with embedded data

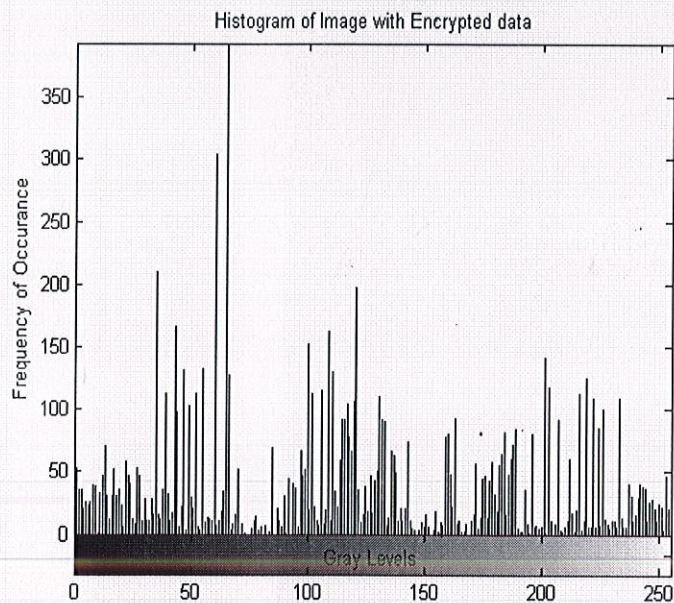
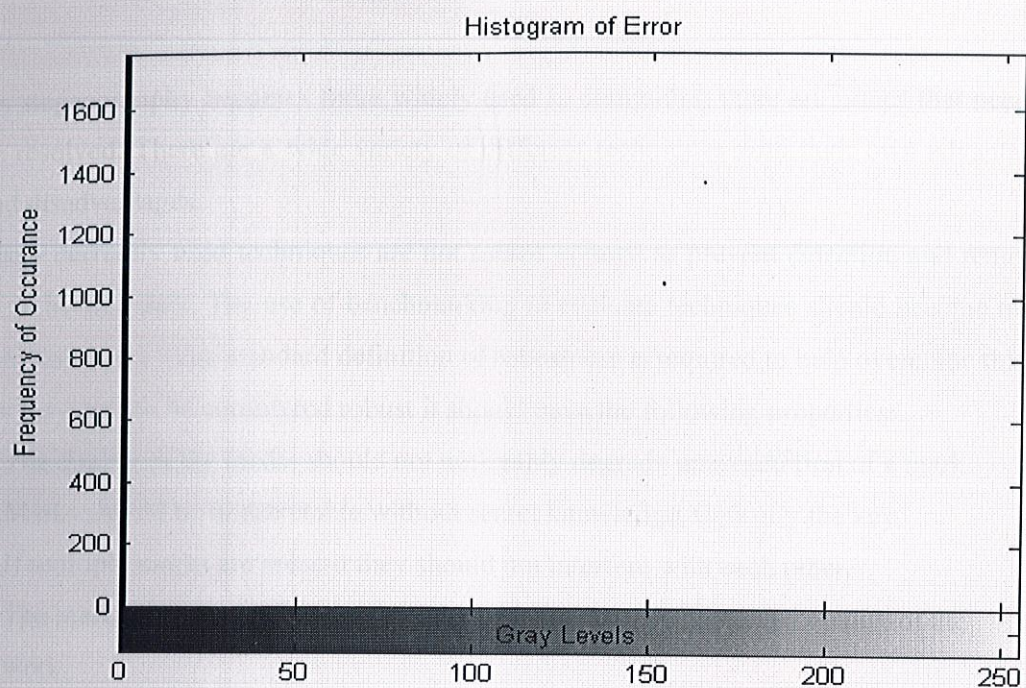


Figure 4.4.3 Histogram of error:



5

CONCLUSION

As steganography becomes more widely used in computing there are issues that need to be resolved. There are a wide variety of different techniques with their own advantages and disadvantages.

Many currently used techniques are not robust enough to prevent detection and removal of embedded data. The use of benchmarking to evaluate techniques should become more common and a more standard definition of robustness is required to help overcome this.

For a system to be considered robust it should have the following properties:

- 1.The quality of the media should not noticeably degrade upon addition of a mark.
- 2.Marks should be undetectable without secret knowledge, typically the key.
- 3.If multiple marks are present they should not interfere with each other.
- 4.The marks should survive attacks that don't degrade the perceived quality of the work.

As attacks are found that work against existing techniques, it is likely that new techniques will be developed that overcome these deficiencies. The continuing use of digital media will drive development of new techniques and standards for watermarking are likely to be developed.

Meanwhile techniques used by law enforcement authorities to detect embedded material will improve as they continue to try and prevent the misuse of steganography.

BIBLIOGRAPHY

1. Network Security – Bible by Dr. Eric Lole, Dr. Ronald Krutz & James W. Conley.
2. Deitel & Deitel, C programming.
3. www.cs.bham.ac.uk/~mdr/teaching/modules03/security/students/SS5/Steganography.pdf
4. [www.totse.com/An introduction to steganography](http://www.totse.com/An%20introduction%20to%20steganography)
5. www.google.com
6. www.wikipedia.com