BUNDELKHAND INSTITUTE OF ENGINEERING AND TECHNOLOGY

Department of Information Technology

Session 2013-2014

Report

On

Steganalysis:Data Concealing & Encryption

Project Guide

Prof. Yashpal Singh

Team Members

Achala Bhati (1004313001)

Anvi Jain(1004313007)

Khyati(1004313017)

Megha Gupta(1004313021)

BUNDELKHAND INSTITUTE OF ENGINEERING AND TECHNOLOGY

Department of Information Technology

Certificate

Certified that this is a bona fide record of the Project Entitled "**Steganalysis:Data Concealing and Encryption**" Completed successfully by **Achala Bhati (1004313001), Anvi Jain(1004313007),Khyati(1004313017) and Megha Gupta(1004313021)** of the 7th semester, Information Technology, under the guidance of Respected faculties in the Partial fulfilment of the requirements to the award of Degree of Bachelor of Technology in Information Technology from "BUNDELKHAND INSTITUTE OF ENGINEERING & TECHNOLOGY" during the Academic year 2013-14.

Verified by:

……

Prof. Yashpal Singh

Acknowledgement

We are deeply indebted to our respected Head of the Department **Dr. Yashpal Singh** for guiding us.

Our sincerest thanks to all our teachers, seniors and colleagues whose help and guidance brought this project to successful completion.

Achala Bhati(1004313001)

Anvi Jain(1004313007)

Khyati (1004313017)

Megha Gupta(1004313021)

1. ABSTRACT

The rapid development of data transfer through internet made it easier to send the data accurate and faster to the destination. There are many transmission media to transfer the data to destination like e-mails; at the same time it is may be easier to modify and misuse the valuable information through hacking. So, in order to transfer the data securely to the destination without any modifications, there are many approaches like cryptography and steganography. This paper deals with the image steganography with encryption and compression for enhancing the security of the systemIt can be used to carry out hidden exchanges and hence can enhance individual privacy. Steganography aims at communicating the secret data in an appropriate multimedia carrier.

The aim of the project is to encrypt the data using 3-d substitution technique and then hide the data over an image using LSB steganography algorithms. LSB (Least Significant Bit) substitution is the process of adjusting the least significantbit pixels of the carrier image. This process helps to send the information to the authorised party without any potential risk. The proposed method will help to secure the content with in the image and encryption of data in the image will help to make the document much securer because even though if the unauthorised person succeeds in being able to hack the image, the person will not able to read the message. Finally we will compress the image using Huffman technique which will reduce the image size and enhance speed of transmission.

2. INTRODUCTION

2.1 Purpose

This project **Steganalysis-Data Concealing and Encryption** & aims at transferring data over the network by encrypting and hiding data in images. By doing so, it becomes easy to transfer confidential data over the internet. Steganalysis is the art and science of writing hidden messages in such a way that no one, apart from the sender and intended recipient, suspects the existence of the message, a form of security through obscurity.

The word steganalysis is of Greek origin and means "concealed writing" . The first recorded use of the term was in 1499 by Johannes Trithemius in his Steganographia, a treatise on cryptography and steganalysis disguised as a book on magic. Generally, messages will appear to be something else: images, articles, shopping lists, or some other cover text and, classically, the hidden message may be in invisible ink between the visible lines of a private letter.

The advantage of steganalysis over cryptography alone is that messages do not attract attention to themselves. Plainly visible encrypted messages—no matter how unbreakable—will arouse suspicion, and may in themselves be incriminating in countries where encryption is illegal. Therefore, whereas cryptography protects the contents of a message, steganalysis can be said to protect both messages and communicating parties.

As a simple example, a sender might start with an innocuous image file and adjust the colour of every 100th pixel to correspond to a letter in the alphabet, a change so subtle that someone not specifically looking for it is unlikely to notice it.

2.2 Scope :

Now a days data security over network is major issue. Generally, messages will appear to be something else: images, articles, shopping lists, or some other cover text and,

classically, the hidden message may be in invisible ink between the visible lines of a private letter..

The advantage of steganalysis over cryptography alone is that messages do not attract attention to themselves. Plainly visible encrypted messages—no matter how unbreakable—will arouse suspicion, and may in themselves be incriminating in countries where encryption is illegal .

Stenography is major provide these services.

1. Network security
2. Secure transmission of bank passwords
3. Transmission of Confidential information.
4. Data concealing

2.3 Requirement Specification

Software Specification

SENDER'S SIDE	
Operating System	Windows
Programming Language	Java enterprise edition
IDE	Net Beans 7.4

Hardware Specification

SENDER'S SIDE	
Intel p4 or equivalent	2 GB

RECEIVER'S SIDE	
Intel p4 or equivalent	2 GB

2.4 Description of Module

Sender Side

Encryption-Encryption is the process of encoding messages (or information) in such a way that eavesdroppers or hackers cannot read it, but that authorized parties can.[1]:374 In an encryption scheme, the message or information (referred to as plaintext) is encrypted using an encryption algorithm, turning it into an unreadable cipher text (ibid.). This is usually done with the use of an encryption key, which specifies how the message is to be encoded. Here we are using the method of 3D encryption.

Steganalysis- Steganalysis includes the concealment of information within computer files. In digital steganalysis, electronic communications may include steganographic coding inside of a transport layer, such as a document file, image file, program or protocol. Media files are ideal for steganographic transmission because of their large size.

Compression- Image compression may be lossy or lossless. Lossless compression is preferred for archival purposes and often for medical imaging, technical drawings, clip art, or comics. Lossy compression methods, especially when used at low bit rates, introduce compression artefacts. Lossy methods are especially suitable for natural images such as photographs in applications where minor (sometimes imperceptible) loss of fidelity is acceptable to achieve a substantial reduction in bit rate.

Receiver Side

Decompression- Image decompression may be lossy or lossless. Lossless decompression is preferred for archival purposes and often for medical imaging, technical drawings, clip art, or comics. Lossy decompression methods, especially when used at low bit rates, introduce decompression artefacts. Lossy methods are especially suitable for natural images such as photographs in applications where minor (sometimes imperceptible) loss of fidelity is acceptable to achieve a substantial reduction in bit rate.

Data Extraction- It is the process of extracting the data back from the image with the help of suitable key and algorithm. The data is hidden during the process of stegnography for secure transmission has to be converted back into the original text for further processing

Decryption-Decryption is the process of decoding messages (or information) in such a way that only authorised user can read it. In a decryption scheme, the message or information (referred to as plaintext) is decrypted using an decryption algorithm, turning it into an readable plain text (ibid.). This is usually done with the use of decryption key, which specifies how the message is to be decoded. Here we are using the method of 3D decryption.

3.BRIEF LITERATURE SURVEY

2.1 Information Security

In general, security denotes "the quality or state of being secure to be free from danger". Security is classified into different layers depending on the type of content intended to be secured:

Physical security: Defines the required issues that are needed to protect the physical data or objects from unauthorized intrusion. Personal security: It is defined as the security of the individuals who are officially authorized to access information about the company and its operations.

Operational security: It mainly relies on the protection of the information of a particular operation of the chain of activities.

Communication "s security: The communication" s security encompasses the security issues regarding the organisation's communication media, technology and content.

Network security: The network security is responsible for safeguarding the information regarding the „networking components connections" and contents.

Information security: Information security is the protection of information and the systems and hardware that use, store, and transmit that information. Information security can be defined as measures adopted to prevent the unauthorized use or modification of use of data or capabilities.

The main objective of the project is to propose the method and critically discuss the properties which help to transmit the data or information over a network without any modifications. The critical characteristics of information are

1. Availability
2. Accuracy
3. Authenticity
4. Confidentiality
5. Integrity

2.2 Security attacks

The data is transmitted from source to destination which is known as its normal flow as shown in the figure.

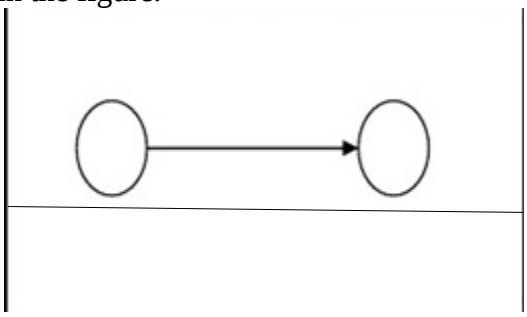

Figure 1: Normal data flow

But the hackers might hack the network in order to access ormodify the original data. These types of attacks are formally known as securityattacks. A hacker can disrupt this normal flow by implementing the different types oftechniques over the data and network in following ways. They are:

1. Interruption
2. Interception
3. Modification
4. Fabrication
5. Source info destination information

There are different types of approaches for preventing the security attacks. The mostuseful approaches are

1. Cryptography
2. Steganography
3. Digital watermarking

2.3CRYPTOGRAPHY

The word cryptography is derived from two Greek words which mean “secret writing”. Cryptography is the process of scrambling the original text by rearranging and substituting the original text, arranging it in a seemingly unreadable format for others. Cryptography is an effective way to protect the information that is transmitting through the network communication path. Cryptology is the science that deals about cryptography and cryptanalysis. Cryptography is the approach of sending the messages secretly and

securely to the destination. Cryptanalysis is the method of obtaining the embedded messages into original texts.

In general, cryptography is transferring data from source to destination by altering it through a secret code. The cryptosystems uses a plaintext as an input and generatea cipher text using encryption algorithm taking secret key as input. The important elements in cryptosystems are

1. Plain text (input)

2. Encryption algorithm

3. Secret key

4. Cipher text

5. Decryption algorithm

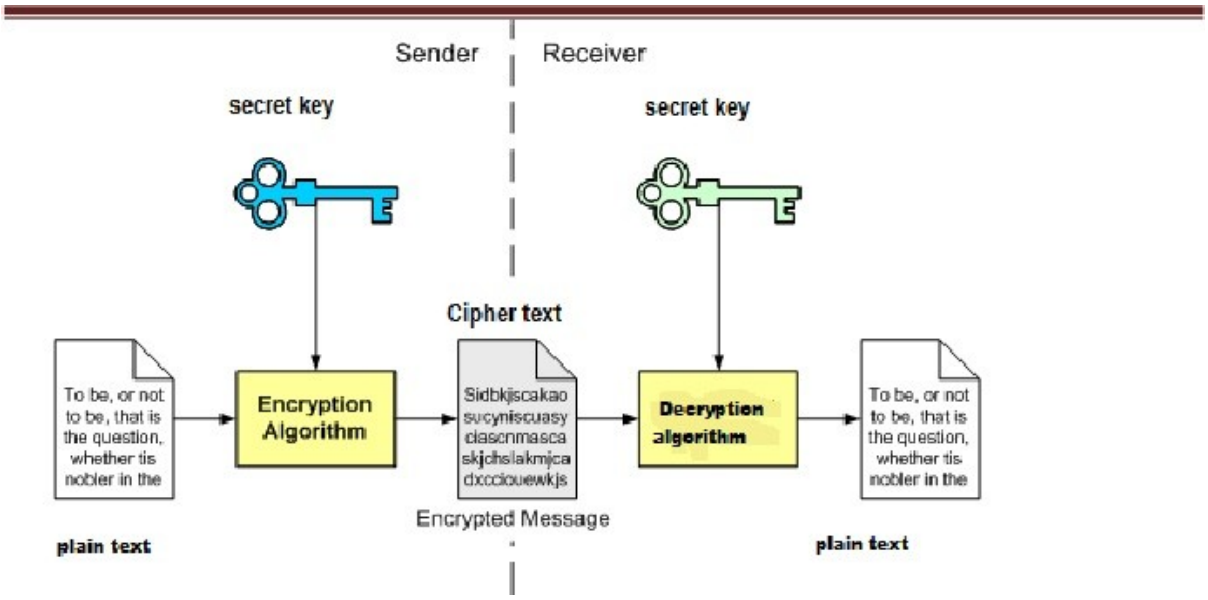

Figure 2: General Symmetric Cryptographic System

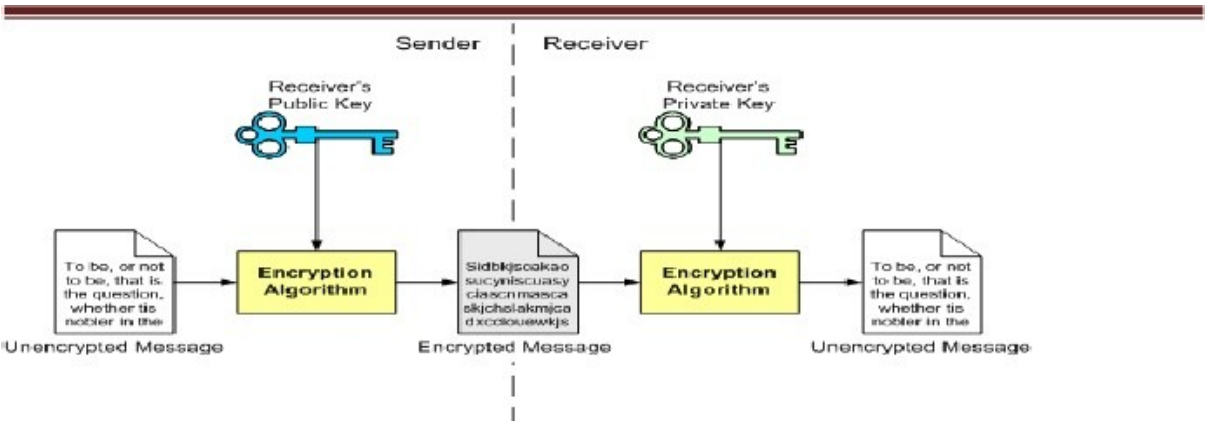

Figure 3: General Asymmetric Cryptographic System

2.3 STEGANOGRAPHY

Steganography in Greek means „covered writing". Steganography is the process of hiding the one information into other sources of information like text, image or audio file, so that it is not visible to the natural view. There are varieties of steganographic techniques available to hide the data depending upon the carriers we use.

Steganography and cryptography both are used for the purpose of sending the data securely. The same approach is followed in Steganography as in cryptography like encryption, decryption and secret key. In steganography the message is kept secret without any changes but in cryptography the original content of the message is differed in different stages like encryption and decryption

.Steganography supports different types of digital formats that are used for hiding the data. These files are known as carriers. Depending upon the redundancy of the object the suitable formats are used. „Redundancy" is the process of providing better accuracy for the object that is used for display by the bits of object.

The different types of steganographic techniques that is available are

1. Pure steganography
2. Public key steganography
3. Secret key steganography

Pure steganography

Pure steganography is the process of embedding the data into the object without using any private keys. This type of steganography entirely depends upon the secrecy. This type of steganography uses a cover image in which data is to be embedded, personal information to be transmitted, and encryption decryption algorithms to embed the message into image. This type of steganography can"t provide the better security

because it is easy for extracting the message if the unauthorised person knows the embedding method. It has one advantage that it reduces the difficulty in key sharing.

Secret Key steganography

Secret key steganography is another process of steganography which uses the same procedure other than using secure keys. It uses the individual key for embedding the data into the object which is similar to symmetric key. For decryption it uses the same key which is used for encryption. This type of steganography provides better security compared to pure steganography. The main problem of using this type of steganographic system is sharing the secret key. If the attacker knows the key it will be easier to decrypt and access original information.

Public Key Steganography

Public key steganography uses two types of keys: one for encryption and another for decryption. The key used for encryption is a private key and for decryption, it is a „public key" and is stored in a public database. For encryption and decryption of text messages using the secret keys steganographic system uses algorithms known as steganographic algorithms.

The mostly used algorithms for embedding data into images are
1. LSB (Least Significant Bit) Algorithm
2. JSteg Algorithm
3.F5 Algorithm

2.4 LEAST SIGNIFICANT BIT

LSB (Least Significant Bit) substitution is the process of adjusting the least significant bit pixels of the carrier image. It is a simple approach for embedding message into the image. The Least Significant Bit insertion varies according to number of bits in an image. For an 8 bit image, the least significant bit i.e., the 8^{th} bit of each byte of the image is changed to the bit of secret message. For 24 bit image, the colours of each component like RGB (red, green and blue) are changed. LSB is effective in using BMP images since the compression in BMP is lossless. But for hiding the secret message inside an image of BMP file using LSB algorithm it requires a large image which is used as a cover. LSB substitution is also possible for GIF formats, but the problem with the GIF image is whenever the least significant bit is changed thewhole colour palette will be changed. The problem can be avoided by only using thegray scale

GIF images since the gray scale image contains 256 shades and thechanges will be done gradually so that it will be very hard to detect. For JPEG, thedirect substitution of steganographic techniques is not possible since it will use lossy compression. So it uses LSB substitution for embedding the data into images. There are many approaches available for hiding the data within an image: one of the simple least significant bit submission approaches is „Optimum Pixel Adjustment Procedure".

The simple algorithm for OPA explains the procedure of hiding the sample text in an image.

Step1: A few least significant bits (LSB) are substituted with in data to be hidden.

Step2: The pixels are arranged in a manner of placing the hidden bits before the pixel of each cover image to minimize the errors.

Step3: Let n LSBs be substituted in each pixel.

Step4: Let d= decimal value of the pixel after the substitution.d1 = decimal value of last n bits of the pixel.d2 = decimal value of n bits hidden in that pixel.

Step5: If (d1~d2)<=(2^n)/2then no adjustment is made in that pixel. Else

Step6: If(d1<d2)d = d-2^n.If(d1>d2)d = d + 2^n. This d is converted to binary and written back to pixel.

This method of substitution is simple and easy to retrieve the data and the image quality better so that it provides good security

3.NEED AND SIGNIFICANCE

- **Secure Transmission of Sensitive Data**

Confidential data needs to be transmitted over the network securely. For this cryptography was used but there are certain limitation of using cryptography as cipher text can be easily seen by attackers and cryptanalysis can be performed over it to recover original text. Steganography overcomes this limitation bt hiding it over the image so attacker does not recognise if some transmission is going on.

- **Size of Payload**

For fast and easy transmission we need that the size of payload should be small. For achieving this feature in our project we compress the image which reduces its size and it can be transmitted easily over the network.

- **Robust against malicious and unintentional attacks**

Attackers are not able to recognise that some transacaction is carried out and even if they detect the image they cannot extract the data out of it.Moreover to enhance the security data is in cipher text format.

- **Use in modern Parameter**

Steganography is used by some of the modern printers, including HP and Xerox brand color laser printers. Tiny yellow dots are added to each page. The dots are barely visible and contain encoded printer serial numbers, as well as date and time stamps

3. OBJECTIVE

In this project the primarily concentration is on the data security issues when sending the data over the network using steganographic techniques. The main objectives of the project are

- Requirement of this steganography system is that the hider message carried by stego-media should not be sensible to human beings.
- The other goal of steganography is to avoid drawing suspicion to the existence of a hidden message.
- It also uses compression so that easy and fast transmission of data is possible.
- The compression used is lossless so that no data is lost and image does not get distorted.

- This approach of information hiding technique has recently become important in a number of application areas.

4. PROPOSED METHODOLOGY DURING PROJECT

Steganography in Greek means „covered writing". Steganography is the process of hiding the one information into other sources of information like text, image or audiofile, so that it is not visible to the natural view. There are varieties of steganographictechniques available to hide the data depending upon the carriers we use.

We have encrypted the file and then use stegnography to hide data with in image. After hiding data in image we compress the file before sending it to other hand. Here are main methodology we use for data hiding using stegnography.

6.1 Encryption

The word cryptography is derived from two Greek words which mean "secret writing". Cryptography is the process of scrambling the original text by rearranging andsubstituting the original text, arranging it in a seemingly unreadable format for others.Cryptography is an effective way to protect the information that is transmittingthrough the network communication paths. We have use 3D substitution technique for encryption.

3D substitution

In cryptography, a **substitution cipher** is a method of encryption by which units of plaintext are replaced with ciphertext according to a regular system; the "units" may be single letters (the most common), pairs of letters, triplets of letters, mixtures of the above, and so forth. The receiver deciphers the text by performing an inverse substitution.

Substitution ciphers can be compared with transposition ciphers. In a transposition cipher, the units of the plaintext are rearranged in a different and usually quite complex order, but the units themselves are left unchanged. By contrast, in a substitution cipher, the units of the plaintext are retained in the same sequence in the ciphertext, but the units themselves are altered.

In this we have given a sequence of alphabet as a key and we can replace the plain text using modulus function. We can use more than one key and for different result of modulus we can use different sequence of 26 alphabets as a key. So using key we can get different letter for different occurrence of same letter in plain text.

6.2 Stegnography:

Steganography and cryptography both are used for the purpose of sending the datasecurely. The same approach is followed in Steganography as in cryptography likeencryption, decryption and secret key. In steganography the message is kept secretwithout any changes but in cryptography the original content of the message isdiffered in different stages like encryption and decryption.

Steganography supports different types of digital formats that are used for hiding thedata. These files are known as carriers. Depending upon the redundancy of the object the suitable formats are used. „Redundancy" is the process of providing betteraccuracy for the object that is used for display by the bits of object.

The different types of steganographic techniques that is available are

1. Pure steganography
2. Public key steganography
3. Secret key steganography

For encryption and decryption of text messages using the secret keyssteganographic system uses algorithms known as steganographic algorithms. Themostly used algorithms for embedding data into images are

LSB algorithm

LSB (Least Significant Bit) substitution is the process of adjusting the least significant bit pixels of the carrier image. It is a simple approach for embedding message into the image. The Least Significant Bit insertion varies according to number of bits in an image. For an 8 bit image, the least significant bit i.e., the 8 bit of each byte of the image is changed to the bit of secret message. For 24 bit image, the colours of each component like RGB (red, green and blue) are changed. LSB is effective in using BMP images since the compression in BMP is lossless. But for hiding the secret message inside an image of BMP

file using LSB algorithm it requires a large image which is used as a cover. LSB substitution is also possible for GIF formats, but the problem with the GIF image is whenever the least significant bit is changed the whole colour palette will be changed. The problem can be avoided by only using the gray scale GIF images since the gray scale image contains 256 shades and the changes will be done gradually so that it will be very hard to detect. For JPEG, the direct substitution of steganographic techniques is not possible since it will use lossy compression. So it uses LSB substitution for embedding the data into images. There are many approaches available for hiding the data within an image: one of the simple least significant bit submission approaches

is „Optimum Pixel Adjustment Procedure". The simple algorithm for OPA explains the procedure of hiding the

sample text in an image.

Step1: A few least significant bits (LSB) are substituted with in data to be hidden.

Step2: The pixels are arranged in a manner of placing the hidden bits before the pixel of each cover image to minimize the errors.

Step3: Let n LSBs be substituted in each pixel.

Step4: Let d= decimal value of the pixel after the substitution. d1 = decimal value of last n bits of the pixel. d2 = decimal value of n bits hidden in that pixel.

Step5: If (d1~d2)<=(2^n)/2then no adjustment is made in that pixel.

Else

Step6: If(d1<d2)d = d

–

2^n. If(d1>d2)d = d + 2^n.

This„d" is converted to binary and written back to pixel

(Amirtharajan et al., 2010).

This method of substitution is simple and easy to retrieve the data and the imagequality better so that it provides good security.

6.3 Compression

The need for an efficient technique for compression of Images ever increasing because the raw images need large amounts of disk space seems to be a big disadvantage

during transmission & storage. Even though there are so many compression technique already present a better technique which is faster, memory efficient and simple surely suits the requirements of the user. In this paper we proposed the Lossless method of image compression and decompression using a simple coding technique called Huffman coding. This technique is simple in implementation and utilizes less memory. A software algorithm has been developed and implemented to compress and decompress the given image using Huffman coding techniques.

Huffman coding technique

Huffman code procedure is based on the two observations [3].

a. More frequently occurred symbols will have shorter code words than symbol that occur less frequently.

b. The two symbols that occur least frequently will have the same length.

The Huffman code is designed by merging the lowest probable symbols and this process is repeated until only two probabilities of two compound symbols are left and thus a code tree is generated and Huffman codes are obtained from labelling of the code tree.

Original source reduction							Source reduction
S	P	code	1	2	3	4	
a2	0.4	1	0.4	1	0.4	1	0.4
1	0.6	0					
a6	0.3	00	0.3	00	0.3	00	0.3
00	0.4	1					
a1	0.1	011	0.1	011	0.2	010	0.3
01							
a4	0.1	0100	0.1	0100	0.1	011	
a3	0.06	01010	0.1	0101			
a5	0.04	01011					

At the far left of the table I the symbols are listed and corresponding symbol probabilities are arranged in decreasing order and now the least t probabilities are merged as here 0.06 and 0.04 are merged, this gives a compound symbol with probability 0.1, and the compound symbol probability is placed in source reduction column1 such that again the probabilities should be in decreasing order. so this process is continued until only two probabilities are left at the far right shown in the above table as 0.6 and 0.4. The second step in Huffman's procedure is to code each reduced source, starting with the smallest source and working back to its original source [3]. The minimal length binary code for a two-symbol source, of course, is the symbols 0 and 1. As shown in table III these symbols are assigned to the two symbols on the right (the assignment is arbitrary; reversing the order of the 0 and would work just and well). As the reduced source symbol with probabilities 0.6 was generated by combining two symbols in the reduced source to its left, the 0 used to code it is now assigned to both of these symbols, and a 0 and 1 are arbitrary appended to each to distinguish them from each other. This operation is then repeated for each reduced source until the original course is reached. The final code appears at the far-left in table 1.8. The average length of the code is given by the average of the product of probability of the symbol and number of bits used to encode it. This is calculated below:

$$L_{avg} = (0.4)(1) + (0.3)(2) + (0.1)(3) + (0.1)(4) + (0.06)(5) + (0.04)(5) = 2.2$$

Huffman encoding and decoding algorithm

Step1- Read the image on to the workspace of the mat lab.

Step2- Convert the given colour image into grey level image.

Step3- Call a function which will find the symbols (i.e. pixel value which is non-repeated).

Step4- Call a function which will calculate the probability of each symbol.

Step5- Probability of symbols are arranged in decreasing order and lower probabilities are merged and this step is continued until only two probabilities are left and codes are assigned according to rule that :the highest probable symbol will have a shorter length code.

Step6- Further Huffman encoding is performed i.e. mapping of the code words to the corresponding symbols will result in a compressed data.

Step7- The original image is reconstructed i.e. decompression is done by using Huffman decoding.

Step8- Generate a tree equivalent to the encoding tree.

Step9- Read input character wise and left to the table II until last element is reached in the table II.

Step10-Output the character encode in the leaf and return to the root, and continue the step9 until all the codes of corresponding symbols are known.

5. FUTURE SCOPE

In the current semester we have implemented the client side module which includes

- Encryption with 3D substitution technique
- Steganography with LSB Technique
- Compression using Huffman Technique

In the next semester we will work on the receiver side module of the project.At receiver side similar kind of procedure will be followed but in reverse order of the steps at client side. It includes

- Decompression
- Extraction of data back from image
- Decryption

The future work on this project is to improve the compression ratio of the image to the text. This project can be extended to a level such that it can be used for the different types of image formats like .bmp, .jpeg, etc., in the future. The security using Least Significant Bit Algorithm is good but we can improve the level to a certain extent by varying the carriers as well as using different keys for encryption and decryption.

6. BIBLIOGRAPHY

[1] Complete reference(Java,J2EE)
[2] Ivan Bayross(Database)
[3] Java Swing by Marc Loy,Robert Eckstein
[4] Roger S.Pressman(Software Engineering)
[5] Fourozen (Computer Networking)
[6] William Stallings (Cryptography and network Security)
[7] Investigator's Guide To Steganography by Gregory Kripper
[8]Data Compression Book By Marc Nelson
[9]Graphic Swing 2 by David M.Geary
[10] Digital Image Processing by William Burger.

7. REFERENCES

Research Papers

[1]IEEE Research paper on Steganalaysis:Data hiding method by Evan Jee Zoung,2011

[2] International Journal on Data compression using Huffman Technique by R.K.Rawat,2010

Websites

[1] www.google.com

[2] http:\\www.howstuffworks.com

[3] www.w3schools.com