Secure Data Transmission Using AES Cryptography In Color Image Steganography

A Dissertation submitted in the partial fulfilment for the award of

Degree of Master of Technology

In

Software Engineering

By

Firoz Alam

2K15/SWE/09

Under the Guidance of:

Prof. Kapil Sharma

DEPARTMENT OF COMPUTER ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

BAWANA ROAD, DELHI

Certificate

It is certified that the work contained in this thesis entitled ”**Secure Data Transmission using AES cryptography in Color Image Steganography**” by **Firoz Alam** is an authentic work which has been carried out under my supervision. The content embodied in this thesis has not been submitted elsewhere for the award of any degree to the best of my knowledge and belief.

Dr. Kapil Sharma

Head of Department,

Department of IT

Declaration

I hereby want to declare that the thesis entitled "**Secure Data Transmission using AES cryptography in Color Image Steganography**" which is being submitted to the **Delhi Technological University**, in the partial fulfilment of the requirements for the award of degree in **Master of Technology in Software Engineering** is an authentic work carried out by me. The material contained in the thesis has not been submitted to any institution or university for the award of any degree.

Firoz Alam

Master of Technology,

Software Engineering,

Department of CSE

Acknowledgement

I take this opportunity to express my deep sense of gratitude and respect towards my guide **Dr.Kapil Sharma,** Department of Information Technology.

I am very much indebted for his generosity, expertise and guidance I have received from him while working on this project. Without his support and timely guidance the completion of the project would not be possible. In this respect I find myself blessed to have my guide. He have guided not only with the subject matter, but also taught the proper style and techniques of documentation and presentation.

I would like to express my gratitude to the university for providing us with the laboratories, infrastructure, testing facilities and environment which allowed us to work without any obstructions.

I would also like to thanks to the Almighty God with his blessings I had an opportunity and strength to do this wonderful project and studies, as well as to my parents who always support me and guide me in the right direct direction with their incredible experiences of life.

Firoz Alam

2K15/SWE/09

M.Tech, SWE

Table of Contents

CHAPTER 1 1
INTRODUCTION 1
1.1 Motivation 3
1.2 Problem Statement 4
1.3 Scope of work 5
1.4 Organization of thesis 6
CHAPTER 2 7
DATA SECURTIY APPROACHES 7
2.1 Cryptography 7
2.1.1 Types of Cryptography 9
2.1.2 Encryption algorithms 10
2.2 Steganography 12
2.2.1 Different types of Steganography 14
CHAPTER 3 25
RELATED WORK 25
CHAPTER 4 28
Data Security Approaches Used 28
4.1 Diffie-Helman 28
4.2 AES 29
4.3 Steganography encoding and decoding algorithms 34
4.4 A combined Security approach 36
CHAPTER 5 38
EXPERIMENTAL RESULTS AND ANALYSIS 38
5.1 Programming Tools and software used 38
5.2 Stegano-Crypto framework 39
5.3 Performance analysis parameters 42
5.4 Comparative analysis of LSB, DCT, DWT 44
CHAPTER 6 51
CONCLUSION AND FUTURE WORK 51
REFRENCES 52

List of figures

Figure 1 Process of hiding data 13
Figure 2 File formats used in steganography 14
Figure 3 Different types of image steganography 16
Figure 4 DCT of an image 20
Figure 5 First step were 1D DWT applied along the rows 22
Figure 6 The vertical operation 23
Figure 7 AES structure 30
Figure 8 AES Encryption and Decryption Process 31
Figure 9 Substitute Bytes 32
Figure 10 Shift row operation 32
Figure 11 Mix Column operation 33
Figure 12 Add Round Key operation 33
Figure 13 Embedding process and performance evaluation 36
Figure 14 Extracting process 37
Figure 15 Stegano-Crypto framework 39
Figure 16 Process of stegano-crypto framework 40
Figure 17 Histogram plot for RGB plane of peppers.jpg 41
Figure 18 Histograms of output image stego.jpg 41
Figure 19 Sample test images 44
Figure 20 PSNR comparison for LSB, DCT and DWT for different image formats 46
Figure 21 PSNR comparison for different secret data size in LSB for peppers.jpg 47
Figure 22 PSNR for different text size in DCT for peppers.jpg 47
Figure 23 PSNR for different text size in DWT for peppers.jpg 48
Figure 24 SSIM comparison for different technique for different image format 49

List of Tables

Table 1: Sample Image Properties…….44

Table 2: MSE and PSNR values for LSB technique……..45

Table 3: MSE and PSNR values for DCT technique……..45

Table 4: MSE and PSNR values for DWT technique……..46

Table 5: SSIM values for LSB, DCT and DWT techniques……48

Table 6: Comparison with several other parameters …….49

List of Abbreviations

AES - Advanced encryption standard

LSB - Least Significant bit

DCT - Direct cosine transform

DWT - Direct Wavelet transform

MSE - Mean Square Error

PSNR - Peak Signal to Noise ratio

SSIM - Structural Similarity index

DES - Data Encryption Standard

BMP - Bitmap image

GIF - Graphics Interchange Format

JPG - Joint Photographic Experts Group

PNG - Portable Network Graphic

TIF - Tagged Image File Format

WWW – World Wide Web

TDES - Triple Data Encryption standard

HTML - Hyper Text Markup Language

MD - Message Digest

ABSTRACT

With the advent of internet to the most part of the world and its usage in vast variety of fields, the sensitive information being communicated through network has also increased exponentially. As the data travels through network of nodes i.e routers, gateways etc, the third party (intruder) may get unauthorized access to the confidential data. The increase in private data communication has invited security measures so that the sensitive information may not land in intruder's hands.

Cryptography and steganography have both come a long way in making the data transmission secure. They both have been used in isolation as well as in combination. Their blending has provided two level security and therefore stronger privacy.

The intruder mostly "pick and choose" data from network. The encrypted information using cryptography technique is unreadable by the intruder but the sender or receiver may still be susceptible to various passive attacks and even make it unreadable for the recipient. The data chosen by intruder for further attacks is mostly encrypted as that will contain the confidential data being communicated. Steganography is hiding the private data in image, audio, video etc which serves the purpose of hiding the existence of confidential information.

This work incorporates color image steganography on the private data which is encrypted using AES (Advanced Encryption Standard) to provide two level data security over the network. The Steganography techniques for images, LSB(least significant bit), DCT(Direct Cosine Transform) and DWT(Direct wavelet transform) are compared and analyzed for various parameters like PSNR(Peak Signal to Noise Ratio), SSIM(Structural similarity index) etc. Steganography overcomes the limitation of cryptography by hiding the fact that some confidential transmission is taking place. Hence these two techniques alone cannot work as efficiently as they do together.

CHAPTER 1

INTRODUCTION

Corporations in both the private and public domains have become more and more reliant on automated data processing. Massive amounts of digital data is being collected and stored in huge computer data bases and transmission of data taking place between terminal devices and computers that are connected together in complex communications networks. Without appropriate security measures, the data is susceptible to interception when in transmission, or data may be physically removed or copied while in storage. This would result in undesirable exposures of data and potential invasions of privacy. Data are also vulnerable to unauthorized deletion, alteration or addition during transmission or storage. This can result in illegal access to computing resources and services, distortion of personal data or business records, or the conduct of fraudulent transactions, which may also lead to increase in credit authorizations, funds transfers' modification, and the issuance of unauthorized payments [16].

Security authorities after recognizing that the secrecy and integrity of certain data must be secured, have passed regulations to help prevent these problems. But regulations alone cannot attacks or eradicate threats to data processing systems. Additional steps are required to preserve the secrecy and integrity of digital data. Among the security measures that have come up as a solution are cryptography and steganography, which uses methods for rendering data unintelligible to unauthorized parties.

Cryptography and steganography are the techniques that are used for safeguarding information transferred via communication-networks that uses communications satellites, land lines, and microwave technology. Steganographic procedures hide the existence of secret data thereby adding a new facet to digital security. Cryptographic procedures beside secrecy also provide extra security to data using digital signatures, message authentication, and personal identification for authorizing electronic funds transfer and credit card transactions.

Eavesdropping is a passive interception that is mostly used by intruders to hurt the secrecy of information being transmitted over a network. The intruder eavesdrops or records information which is being transmitted. An attack which involves only eavesdropping is called a passive attack. If the opponent in addition to recording information also amends transmitted information or inserts information into the communication path, then the attack is called an active attack.

There is possibility that an attacker can capture data by eavesdropping. There are a number of ways used for eavesdropping:-

Wiretapping:- Interception of single transmission over communication lines by the help of hardwired connections.

Electromagnetic Eavesdropping:- Interception of wireless transmissions, for example, microwave and radio transmissions, or information holding electromagnetic energy originating from electronic devices.

Acoustic Eavesdropping:- Intercepting sound waves generated via human voice or by printing, punching, or transmitting equipment.

In a passive attack, a recording over a tape is made of digital data which is intercepted from a communication path. The data may be reconstructed by examining the recording tape or directing it back into suitable receiving equipment like terminal, modem. In an active attack, a modem and terminal that are compatible with the transmission line are essential and in some cases a minicomputer is used which can quickly modify intercepted information.

Cables between corporate offices and telephone company junction boxes for the connection are particularly vulnerable to eavesdropping or wiretapping. A wiretap is sometimes easy as no special technical skills are needed and the essential equipment are relatively economical. However once the lines are beyond the premise of the building and until they touch telephone company switching facilities, access to selected lines gets more testing. Effective attacks are however still possible and hence the communication of data should be properly protected from unknown third party.

Intervention to microwave and radio transmissions poses a particularly great threat to private communication because there is no requirement of physical connection (tap) to the transmission link. Though as microwave links and also those used in satellite communications, may contain several number of channels, sophisticated and expensive equipment are required to intercept and separate the signals of channel. But even though it is costly, the prize for a successful attack may be enormously rewarding and hence a constant threat to the confidential communication.

In the absence of strong security measures, an eavesdropping third party may get to know substantial information about the operational procedures of the system, including passwords, to overthrow any weak security mechanisms which may lead to catastrophic loss to a person or an organization. Cryptography and steganography forms the backbone of secret communication and their amalgam have provided wonderful results.

1.1 Motivation

Some of the conventional methods for data security while transmission over a network involves either cryptography or steganography and their combination is rarely used. This work combines both the approaches to add a second line of defense and maintaining the confidentiality of information while transferring over the network.

As the encrypted message travels through network, it can be analyzed by the intruder for various attacks. If a hacker gets to know the IP address of a server, the attacker can do Distributed Denial of Service attack on it which may lead to service becoming unavailable for the intended user. The attacker can also probe for the running services and also if any ports is open on the host and would try to exploit it in some manner like FTP, Mail, Mysql etc.

Hence the need for steganography arises so that the existence of confidential communication is hidden within a stego object. If intruder analyses the data, there will be no hint to as any confidential communication is taking place therefore protecting the sender and receiver from a number of possible attacks. But even if the intruder do perform steganalysis on the stego object the AES encryption would stand on its own and intruder would not able to get the access to the confidential information.

Both the techniques perform the same function to protect important data but rather in a distinctive way. Cryptography do hides the content of secret message but not the presence of secret message and Steganography hides the presence of any secret communication. This work presents various methods where steganography and cryptography are blended to perform encryption and also to hide the data in the image. Hence providing two levels of security to the data being transferred. Thereby resulting in a more secure system as compared to when used alone.

As there are number of image steganography techniques available, an additional work that needs to be carried out is the comparative analysis of the techniques with respect to various image formats and performance parameters.

1.2 Problem Statement

Cryptography and Steganography are two of the popular techniques that are widely used to manipulate information in order to cipher or hide their existence respectively. Our aim is to build a system that combines both the approaches and build a secure system for data communication. The stego object used for hiding the data is image in our case.

Firstly the data is to be encrypted using AES (Advanced Encryption Standard) and after that cipher text produced should be stored in the image using some steganography technique like LSB, DCT, DWT. The technique would be used for different types of images and the techniques are compared to select a better one according to the suitability.

Image steganography techniques, LSB, DCT, DWT are compared with various performance analysis parameters like PSNR(Peak Signal to Noise Ratio), MSE(Mean Square Error) and SSIM(structural similarity index) so that proper technique could be applied according to the requirement. The PSNR is the factor that helps in comparing the quality of Stego image as compared to the original image. MSE is the other parameter that computes the magnitude of average error between the original image and stego image. SSIM measures the structure similarity of two images and gives consistent result with respect to human optical system. Hence these three parameter and other factors like capacity, robustness, invisibility etc helps in selecting better steganography technique.

1.3 Scope of work

A secure file transfer solution that need not to be too costly or complicated has been the area of research from so long with the so much important data crossing over network. Secure data transmission explores the security threats involving transmission of data, like eavesdropping and decrypting. It debates why and how to establish protected data communication and also the ways to avoid or hinder attacks on these secure channels. The work is primarily carried out for anyone who is trying to design a completely secure system for data transmission with the help of encryption and data hiding. Any person or organization involved with transmitting sensitive data-whether in a business that communicates private information, requires a robust and secure system for communication. This vary from various corporates like banks, e-commerce websites, government offices, and what not whether private or public.

In this work a secure system is developed to transfer data using the amalgam of cryptography and steganography. It uses AES to provide the robust encryption which is practically unbreakable. On top of encryption an image is used to hide the encrypted data. The steganography technique to hide the data is a variation of traditional LSB technique.

The scope of the work can be summarized as:

- Implementing the AES (Advanced Encryption Standard) algorithm which uses 256 bit key
- Designing the user interface for steganography technique for hiding the data in image
- Merging both AES and steganography on image so that encrypted data is hidden in the image
- Designing the reverse technique which would decode the encrypted data from the image which can then be decrypted to get the original message
- Comparing LSB, DCT, DWT image steganography technique for various parameters so that suitable technique could be selected according to the requirement.

1.4 Organization of thesis

Rest of our work can be summarized as below:-

Chapter 2 discusses various methods used for data security with respect to cryptography and steganography

Chapter 3 reveals about the previous research work done in the field, with combination of steganography and cryptography

Chapter 4 explains the usage of AES in color image steganography and building a secure data transmission system and comparing different steganographic technique for better selection

Chapter 5 illustrates the working of our experiments and comparison with previous approaches.

Chapter 6 concludes the work with inferences drawn from our analysis and explained about its future work

CHAPTER 2

DATA SECURTIY APPROACHES

The two fields, steganography and cryptography have proven their worth in securing the private data over communication lines. This chapter focuses on various methods used in these two techniques for better understanding of the historical development in both approaches. The key concepts related to these two techniques have been detailed further in this chapter.

2.1 Cryptography

In the late 1960s and early 1970s data security began to be recognized as a major design concern for data processing (DP) systems. During this period, systems were designed to operate reliably only in environments subjected to "random noise"-power line disturbances, spurious electromagnetic radiation, equipment malfunction, programming errors etc. Very few precautions were taken to protect the secrecy of computer data, or to defend it against "intelligent noise"-the deliberate actions of people intent on subversion. As a result, many systems were vulnerable to attack. Transmitted data could be intercepted and data could be modified, deleted, or added to a system. But today data processing system designers are more aware of these threats, and cryptography is recognized as an important factor in the design of secure systems.

A basic problem in cryptography is developing methods to transform messages (plaintext) into cryptograms (cipher text) that can survive intense cryptanalysis which are the techniques applied by intruder to penetrate encrypted communications and recover the original information.

The procedures or methods used to achieve such transformations involve either a code system or a cipher system. Code systems require a code book or dictionary that translates words, phrases, and sentences of plaintext vocabulary into their equivalent cipher text code groups. However, the number of plaintext groups that can be converted depends on the size of the code book. Therefore,

not every message can be encoded, and the versatility of these code systems is limited.

On the other hand, cipher systems are versatile. In cipher system two basic element are required, a cryptographic algorithm (a set of rules or a procedure) and cryptographic keys used for encryption or decryption. A key is a relatively short, secret sequence of characters or numbers chosen by the user.

This chapter discusses two particularly useful ciphers: block ciphers and stream ciphers. Both conventional algorithms (e.g., DES) and public-key algorithms (e.g., the RSA algorithm and the trapdoor knapsack algorithm) are covered under the subject of block ciphers.

Both block and stream ciphers can be used in communications and data processing systems: With a block cipher, data are encrypted and decrypted in blocks, whose length are predetermined by the algorithm's designer. With a stream cipher, the algorithm's user determines the length of data to be encrypted and decrypted. This flexibility requires that stream ciphers, in addition to the algorithm and key, employ another parameter defined as an initializing vector. Different modes of encryption can be obtained with block and stream ciphers by employing feedback methods (chaining), which establish dependencies to past information. Chaining not only strengthens a cipher, but can also be used to authenticate data even when privacy is not required.

The cryptographic algorithm can be thought of as an extremely large number of transformations, the particular transformation in effect depending on the cryptographic key being used. Each transformation changes sequences of intelligible data (plaintext) into sequences of apparently random data (cipher-text). The conversion from plaintext to ciphertext is known as encipherment encryption. Each transformation must have a unique inverse operation which would also be identified by a cryptographic key. The reverse transformation from cipher-text to plain-text is known as decipherment or decryption.

There are two categories of cryptographic algorithms, conventional and public key. With a conventional cryptographic algorithm, the enciphering and deciphering keys are either identical, or, if different, are such that each key can be simply calculated from the other. Thus knowing the

enciphering key is equivalent to knowledge of the deciphering key-when you have one, you also have the other.

A public-key algorithm, on the other side, permits many users or nodes within a communications system to encipher data using the same public key, but only the specific user or node possessing the secret deciphering key can "unlock" or recover the data. In contrast, a conventional cryptographic algorithm provides effective data security between two users or nodes within a communications system only if these users or nodes have knowledge of the same secret key.The cryptographic key used in a conventional cryptographic algorithm and the private key used in a public-key algorithm are examples of cryptographic variables. They are analogous to the secret combination for a safe.

There are many widely available encryption algorithms that are used in securing data. They are divided into Symmetric (private) and Asymmetric (public) keys cryptography. In Symmetric or secret key encryption, the encryption or decryption is carried out by using only one key. In Asymmetric cryptography two different keys are used for encryption and decryption. Encryption is carried out by using public key and decryption is done by private key (e.g. RSA algorithm). The basis for public key encryption are mathematical functions which are computationally intensive. The various examples of strong and weak keys of cryptography algorithms are DES, AES. AES uses various 128,192,256 bits keys depending on various factors while DES uses one 64-bits key. The larger the key length the better is the cipher provided by it.
AES with larger key size has proven to be practically unbreakable and hence is the standard for most of the confidential information sharing. AES if implemented properly is worth everything that a secret communication require over WWW(world wide web). The various types of cryptography algorithms are described in detail.

2.1.1 Types of Cryptography

The cryptography procedures are distinguished based on their key selection. This section throws light on the merits and demerits of different cryptographic approaches [11].

Private or symmetric Cryptography (also termed as secret-key encryption) It involves carrying out encryption and decryption with the same key. The one who has the key can decipher it but the problem here is how to share the key securely and efficiently. This approach proves to be effective and fast when comparison is done with the asymmetrical key cryptography. In simple words, key would be generated by the encryption algorithm which is then send to the receiver where decryption can be done with the same key.

Asymmetric (Public) Cryptography In a public-key encryption scheme, there are two keys involved one is public known to all and other is private key known to specific user. First of all a network user asks for a public and private key pair. A public administrator can provide the intended receiver public key to the user who wants to communicate a secret message. A basic application of public-key cryptosystems is the session keys or key used in symmetric cryptography distribution. The method tends to provides better security when comparison is carried out with private key cryptography. However the problem here is that extra hardware is needed since more processing time and more energy needed. Due to escalation in the computational mathematical unit the outlays are high in asymmetric cryptographic systems.

2.1.2 Encryption algorithms

Data Encryption Standard (DES)

DES is one of the most widely used cryptographic systems which is publicly available and accepted today by internet community. It was designed by IBM in 1970s. Later on it was taken by US govt. as a standard algorithm for digital communication in 1976. It requires a fix length 56-bit crypto key to carry out encryption of the 64-bit block of data. The DES algorithm takes 56-bit keys and 64-bit plaintext messages as inputs and outputs a 64-bit cryptogram and the algorithm performs 16 iterations for each block with transformation operations [33].

Blowfish

Blowfish is a symmetric-key block cipher which is developed in 1993 by Bruce Schneier and have been used in a number of encryption products. Till now no operative cryptanalysis has been developed for Blowfish. However, the AES(Advanced Encryption Standard) now has gained more

attention and popularity. Bruce designed the Blowfish as a general-purpose algorithm, the purpose was to serve as a substitute to the DES. The motivation helped him to get it free of the problems and constraints free as associated with other algorithms. In the course of release of this algorithm, many other designs were copyrights, hindered by patents or were commercial or government secrets. Bruce opened the algorithm for all and hence for public use. He said that "Blowfish is unpatented, and will remain so in all countries. The algorithm is hereby placed in the public, and can be freely used by anyone." Main features of the design include a highly complex key schedule and key-dependent S-boxes.

Triple DES (TDES)

It was designed in 1998 and is a derivative of DES. It is triple DES because DES cipher algorithm is applied 3 times to every block of data. There is increment in key size in Triple DES to ensure added security through encryption capabilities. The data block is of 64 bit. There are three keys used which are called as bundle keys, each is of 56 bits. As triple DES uses three keys each of 56 bits which amounts to 168 bit key length and hence better security with increased key length. The key length in TDES is of 168 bits but due to man in middle attack the effectiveness of key security falls to 112 bits.

Advanced Encryption Standard (AES)

AES is a symmetric data encryption technique which uses 128-bit block data designed by Belgian cryptographers Joan Daemen and Vincent Rijmen. The algorithm was adopted as its encryption standard by U.S government in October 2000, which replaced the DES algorithm which was used before. AES works at several network layers simultaneously. The NIST(National Institute of Standards and Technology) of the U.S. Department of Commerce choose the algorithm, called Rijndael algorithm , out of a collection of five algorithms under review. While the terms AES and Rijndael are used interchangeably, there are some differences between the two. AES has a fixed block size of 128-bits and a key size of 128, 192, or 256-bits, while Rijndael can be specified with any key and block size is a 32 bit multiple, with range which vary from 128 bit to 256 bits.

RSA

RSA is an asymmetric cryptographic system where two seperate keys are used to carry out encryption and decryption. It was designed in 1978 by a team of Ron Rivest, Adi Shamir and Leonard Adleman. The operations in RSA includes three phases- generation of key, encryption and decryption phases. RSA process starts by generating key, it starts by choosing two large prime numbers and then consequently by using various mathematical properties the keys are generated. The RSA user publishes the product of those prime numbers along with some public key. Anyone can encrypt the message using the public key but decryption requires the prime numbers which are difficult to factor from the product. Due to its flaws it has not been preferred for commercial purpose. The strength of the RSA depend upon the prime numbers selected, if the numbers selected are small the attacker can easily factor the product and therefore decrypt. But if large prime numbers are chosen then performance reduced in comparison to DES due to more time consumption. Also the lengths required of both the numbers should be same which is difficult to fulfill.

Diffie-Hellman

Diffie Hellman algorithm is a secure method for exchanging crypto-keys. This technique was introduced in 1976 developed by Whitfield Diffie and Martin Hellman. The method allows for sharing keys over an insecure channel. Earlier the keys would be transferred physically. The simples and the original proposal of the Diffie-Hellman protocol uses the multiplicative group of integers modulo p, where p is prime, and g is a primitive root modulo p. This method proves secure against eavesdropper if proper care is taken by Alice and Bob in selection of finite cyclic group G of order n and a generating element g in G.

2.2 Steganography

Steganography is synonymous term for covert or hidden communication. It operates by hiding messages in some cover objects that will be then sent to the intended destination. The most essential prerequisite of any steganographic framework is that it ought to be unimaginable for a intruder to recognize difference between ordinary object and objects that contain mystery

information. Steganography in its modern form is relatively young. Until the early 1990s, this unusual mode of secret communication was used only by spies. At that time, it was hardly a research discipline because the methods were a mere collection of clever tricks with little or no theoretical basis that would allow steganography to evolve in the manner we see today. With the resulting unconstrained move of communication from analog to digital, this antiquated field encountered an explosive innovation. Hiding messages in electronic documents for the purpose of covert communication seemed easy enough to those with some background in computer programming. Soon after, steganographic applications showed up on the Internet, giving the masses the capacity to conceal documents or messages in pictures, text or audio. At the same time, steganography caught the attention of researchers and quickly developed into a rigorous discipline. With it, steganography came to the forefront of discussions at professional meetings, such as the Electronic Imaging meetings annually organized by the SPIE in San Jose, the IEEE International Conference on Image Processing (ICIP), and the ACM Multimedia and Security Workshop. In 1996, the first Information Hiding Workshop took place in Cambridge and after that a series of workshops has since become the premium event to present the latest developments in theory and applications of data hiding.

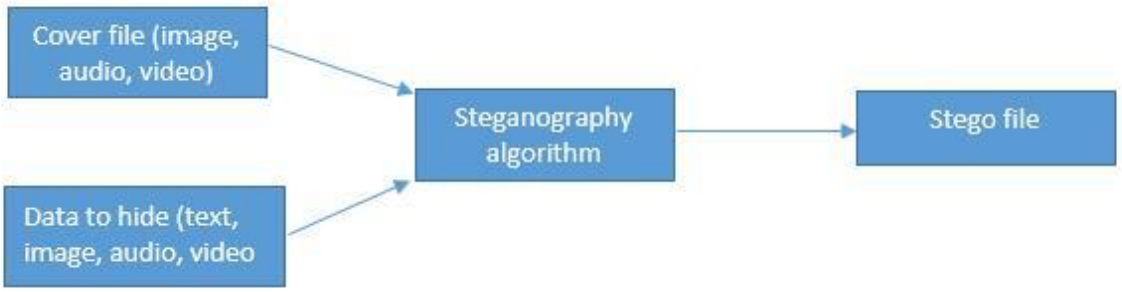


Figure 1 Process of hiding data

Steganography shares many common features with the related but fundamentally quite different field of digital watermarking. In late 1990s, digital watermarking dominated the research in data hiding due to its numerous lucrative applications, such as digital rights management, secure media distribution, and authentication. As watermarking matured, the interest in steganography and steganalysis gradually intensified, especially after concerns had been raised that steganography

might be used by criminals. The steganography has gain popularity since many government have put a lid on putting limits to power of cryptographic systems and also forbidden them thereby individuals have shifted focus on other technique for private data exchange. Various organizations have also adopted this technique for communicating trade secrets or product information.

2.2.1 Different types of Steganography

All digital file format like image, text etc can be utilized for steganography, however the formats those are more appropriate are those with a high level of repetition or redundancy. Repetition can be characterized as the bits of an object that give precision far superior than would normally be essential for the object's display and use. The repetitive bits of an object are those bits that can be modified without the adjustment being identified easily. Pictures and audio records particularly agree to this prerequisite, while study has likewise revealed other file format that can be utilized for data hiding.

Figure 1 demonstrates the four principle classes of file formats that can be utilized for steganography.

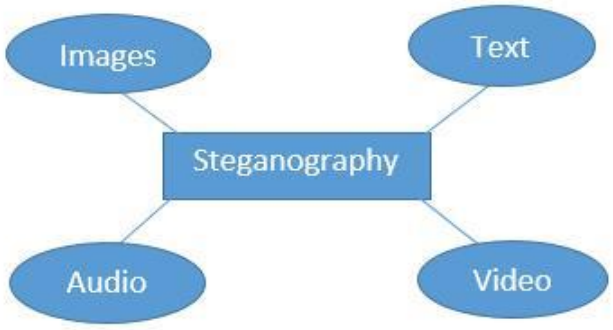


Figure 2 File formats used in steganography

Concealing data in text content is verifiably the most important and oldest strategy for steganography. An apparent strategy was to conceal a confidential message in each nth letter of each word of a text document. It is just since the start of the Internet and because of all the diverse and advanced file format that it has diminished in significance. Text content steganography using digital files is not used regularly since text content have a little amount of repetitive or redundant

information which could be used for hiding purpose.

Given the explosion of computerized images, particularly on the web, and given the substantial amount of repetitive bits introduced in the digitalized illustration of an image, they are the most prevalent cover objects for steganography. This work will concentrate on concealing data in images.

To conceal data in audio files comparable methods as for image files are used. A distinctive method of audio steganography which exploits the properties of the human auditory system to conceal data unnoticeably is known as masking. A faint however audible sound ends up unnoticeable when present with another louder and clear sound. This property makes a channel in which we can conceal the data. But as they are equivalent to pictures in stenographic potential, the bigger size of significant audio documents makes them less prevalent to use as compared to images.

2.2.1.1 Image steganography

As described in previous section, images proves to be the most commonly used carrier objects for steganography. In the field of digital images there exist several image file formats, most of them for some specific applications. Different steganographic algorithms exist for these different type of image file formats. The methods gives different level of concealment depending on the format used [32].

Image steganography methods can be broadly categorized into two groups, those that are in the Image or Spatial Domain and those in the Transform or Frequency Domain. Spatial domain techniques directly make use of the pixel intensities to embed messages, whereas in Transform domain, before embedding the message images are first transformed and then embedding is carried out.

Spatial domain techniques involves bitwise methods which apply bit insertion and noise manipulation and are sometimes referred to as “simple steganography systems”. The lossless

formats of the images proves to be most proper for image domain steganography. The techniques mostly depend on the image format.

Frequency domain steganography techniques comprises of the manipulation of standard steganography algorithms and the image transforms. These methods makes the concealment more robust by hiding messages in more substantial areas of the cover image. Most of these methods are independent does not depend on image format and the embedded message may persist with transformation between lossy and lossless compression. The major types of image steganography has been shown below:

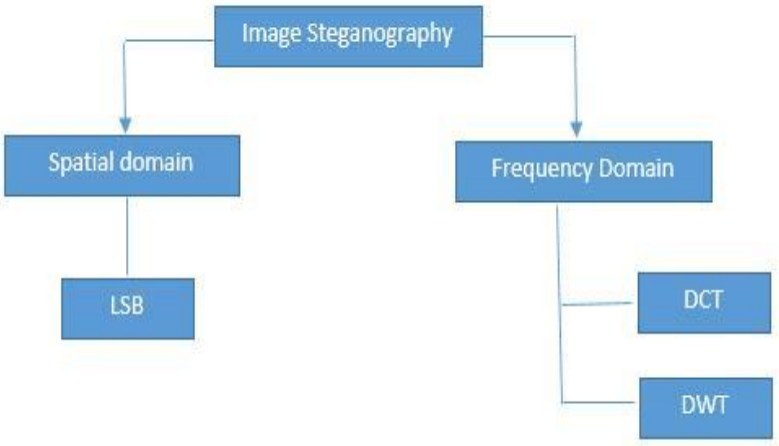


Figure 3 Different types of image steganography

To a computer machine, an image is an accumulation of numeric values that constitute diverse light powers in various regions which represents some color of the image. These collection of numbers creates a matrix or a grid and the individual units are called as pixels. The images are comprised of a 2D grid which have subdivisions as atomic unit called pixel which store a number representing the color intensity. An important term with respect to images is bit depth which denotes the number of bits required to represent a pixel. Mostly the minimum bit depth in present color schemes is 8, which refers to eight bit needed to mention the color for every pixel in the image. Greyscale and Monochrome images have bit depth of 8 and hence they exhibit 256 shades of a particular color or grey. Digitalized color images uses the RGB color model and are mostly

stored in 24-bit files. Various color variants for the pixels of a 24-bit image are derivatives from three primary colors red, green and blue, and each of the primary color is represented by 8 bits. Hence in a single given pixel, there can be 256 different quantities of red, green and blue, which adds up to more than 16-million of combinations, therefore it results in more than sixteen million colors. It can be derived that the more the amount of colors that can be displayed, larger will be the resulting file size.

In the upcoming sections steganographic algorithms will be explained according to the domain in which they are used and also described based on the image format.

Spatial Domain

Least Significant Bit approach

Least significant bit (LSB) insertion is the most simple and common technique which is used to embed information in the image. The LSB or the 8^{th} bit of some or all of the bytes of pixel inside an image is used to hold a bit of the confidential information [3]. When using a color 24-bit image each plane of red, blue and green is represented by a byte and hence a bit of each of three can be used to embed the message. In simpler words, if we use only LSB to store information then 3 bits in each pixel can be used to store the message. An 200 × 300 pixel image would therefore stores a aggregated sum of 180,000 bits or 22,500 bytes of encoded message.
For instance a pixel of RGB image with three components R, G, B and hence a 24-bit image can be seen as shown below:

(00101100 00011101 11011101)
(10100111 11000101 00001111)
(11010000 10101100 01100010)

Take a decimal number 100 which could be converted to binary format 1100100 which is then

embedded into the LSB of the each pixel of the image as shown in above grid, the resulting matrix will be altered as shown below:

(00101101 00011101 11011100)
(10100110 11000101 00001100)
(11010010 10101100 01100011)

While the decimal numerical value is inserted into the first 7 bytes of the grid, the change occurred only in 4 of the underlined bits as visible in the embedded information. On an average, we can say that only half of the bits in an image will be required to be modified to hide the confidential data using the maximum cover size. Meanwhile as each of the primary color has 256 potential intensities, the intensity of the colors would have minimal effect by changing the LSB of a pixel. These slight differnence would not be noticed by the human eye, thereby making the message obscurity successful. If the image is well chosen, we can also conceal the message in the least as well as second to the LSB and see no difference still.

In the example above, as the successive bytes of the image have been used to incorporate the information. This technique can be detected easily by steganalysis. A slightly more secure system would be sharing a secret key by both sender and receiver that will specify merely certain pixels that are to be changed. If an opponent become suspicious that LSB steganography is being used, he would have no way of getting to know which pixels stores the secret message without access to the secret key.

In its most simple and common form, BMP images were used to implement LSB steganography because they uses lossless compression. But to hide a confidential data in a BMP file, we would need a very big cover image. Now a days, BMP images of large size are not regularly used on the Web and might lead to suspicion by the intruder. For the similar reason, LSB steganography has additionally been designed for use with other image file formats.

LSB for Palette Based Images

Palette based pictures like GIF image have also been popular image format which was designed primarily for utilization on the Internet. The bit depth for a GIF image can't be more than eight, therefore the most extreme number of colors is 256 which GIF holds. These type of images are indexed pictures where the colors utilized as a part of the image are put away in a palette or color lookup table. Every pixel is denoted by a solitary byte and the pixel information is a index to the lookup table. The arrangement of color is carried out in the palette for faster look up or decreased query time.

LSB steganography can also be applied to the GIF image but extra care is needed. The main problem here in GIF palette based images is that if someone makes change in the LSB of the pixel it would result in totally different color as pixel then would point to the different index of color palette. If the neighboring entries to palette are similar then the change would little or non-noticeable but if the neighboring palette entries are largely dissimilar then a change would be evident. One possible answer to the problem would be to perform sorting of palette so that the difference in the color in the successive colors is reduced. Another possible solution would be to addition of new colors which have visual similarity to the colors already existing in the palette. The requirement would be that the image should have less no of colors thereby less uniqueness in variety of colors. Utilizing this technique one ought to therefore take care when picking the correct cover image. However any messing with the palette of a listed picture leaves a clear mark, hence makes the detection easier. Another solution would be to make use of greyscale images where there are 256 different colors and the difference between the colors is very slight, thereby it makes the detection difficult.

Frequency Domain

Frequency or Transformation domain methodologies depend on the manipulation of the transformation of the image instead of the image itself. Transformation domain approaches are based on image processing according to the frequency components. The main principle behind the transform domain techniques for image enhancement comprises of the calculating a 2-D discrete

unitary transformation of the image, for example the 2-D DFT (Direct Fourier Transform), manipulating the transform coefficients by an operator M, and then performing the inverse transform. The orthogonal transformation of the image has two parts phase and magnitude. The magnitude comprises of the frequency content of the image. The phase is utilized to reestablish the image back to the spatial domain. The typical transform domain allows for the operation on the frequency component of the image, and thereby edges and other subtle information which are high frequency content can easily be improved.

Discrete Cosine Transform method (DCT)

DCT coefficients have been used for applying compression in JPEG images. DCT divides the image into different parts of significance [2]. It helps in transforming an image or signal from the spatial domain to the transform or frequency domain. It divides the image into various frequency components i.e, high, middle and low. Low frequencies in image means there is slight variations in pixel over space, while high frequency content refers to fast changes in pixel values over space. The low frequency sub-band contains the significant visual part of the image. So the confidential message is usually embedded by performing modification in the coefficients of the middle frequency sub-band, so that the image visibility is not much affected.

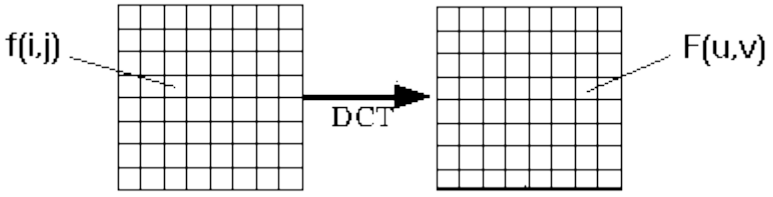


Figure 4 DCT of an image

The DCT is a approach for transforming a signal into elementary frequency components. It shows an image as a summation of sinusoids of varying frequencies and magnitudes. For an input image X, we can calculate the DCT coefficients of the transformed output image Y, by using Eq. 1 for 1D Data. In the following equation X is an input image that possess N x M pixels, Y(u,v) is DCT coefficient in uth row & vth column of the DCT matrix for 2D image and X(m,n) is the intensity of the pixel in mth row & nth column of image matrix.

The 1D (N data items) DCT equation is defined by the following equation where u = 1, 2, 3…..N-1.

$$Y(u) = \text{a(u)} \sum_{i=0}^{N-1} X_i \left(\cos \frac{(2i+1)u\pi}{2N} \right) \qquad (1)$$

The 2D (N × M image) DCT equation is defined by the following equation:

$$Y(u,v) = \text{a(v)} \sum_{i=0}^{n-1} \left[a(u) \sum_{i=0}^{N-1} X(m,n) \left(\cos \frac{(2i+1)u\pi}{2M} \right) \right] \times \cos \frac{(2i+1)v\pi}{2N} \qquad (2)$$

Where u = 1, 2, 3…..N-1 and v=1,2,….M-1. The size of the input image is N × M. X(i, j) represents intensity of the pixel in row i and column j and Y (u, v) is the DCT coefficient in row u and column v of the DCT matrix. The image is broken into 8x8 blocks of pixels for DCT computations. DCT transformation is carried out by applying the computation from left to right, top to bottom to each block. Quantization table is used to compress each block to scale the DCT coefficients and message is then embedded in DCT coefficients where LSB of coefficient is generally used.

Discrete Wavelet Transform Technique (DWT)

It successively decomposes an image acting as a mathematical tool. It proves to be helpful for processing of non-stationary signals [10]. The transform depends on small waves called wavelets of changing frequency and limited time. Wavelet transform gives both frequency and spatial depiction of a picture. Dissimilar to regular Fourier transform, temporal data is reserved in this transform procedure. Wavelets are made by translation and expansions of a function that is fixed called mother wavelet. This part investigates appropriateness of DWT for image steganography as compared to other transforms. When we perform DWT on 2-D images, then the image is processed by 2-D filters in both dimensions. These filters decompose the input image into four parts. These parts are various sub-bands which are multi resolution and non overlapping, called as LL, LH, HL,HH. Generally most of the image energy is stored at lower frequency sub LL so Steganography in these sub-bands may put down the quality of image. However embedding in low frequency sub-bands would lead to increase in robustness. On the other side the edges and textures of an image

are represented by the high frequency sub-bands. Usually people do not notice little changes in edges and textures, so high frequency sub-bands proves to be more suitable for embedding without being notice by the human eye.

The Haar DWT is mostly used in steganography approach. A 2D Haar DWT contains two operations, first is the horizontal operation and the other one is the vertical operation. Complete steps involved in a 2D Haar DWT have been discussed below:

I. Firstly 1 Dimensional DWT is applied which results in low-pass and high pass filtered images. The result is obtained by performing the sum and difference operations on adjacent pixels. Then storing the difference on the right and sum on the left as shown in figure 1.5. Repeating the operation on all of the rows . The result is two images where sum denotes L(low frequency) and difference denotes H(high frequency part).

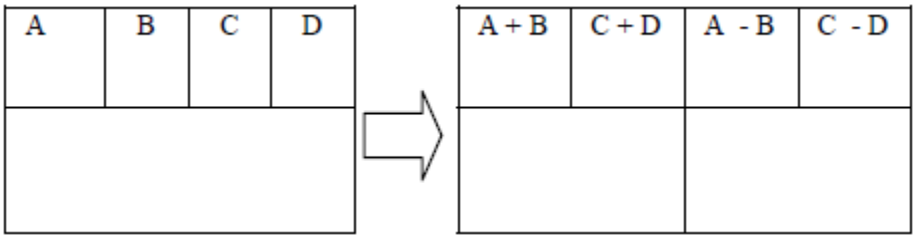


Figure 5 First step were 1D DWT applied along the rows

II. In second step vertical scanning of the pixels is done on both the images from first step. The sum and difference is carried out in verical direction leading to creation of four sub-bands as shown in Fig 1.6. Lastly 4 sub-bands denoted as LL, HL, LH, and HH obtained. The LH band tries to minimize horizontal features of an image whereas HL diminishes vertical feature.The LL sub-band is the low frequency portion and hence has close similarity to the original image. The technique discussed is known as the first order 2D Haar DWT.

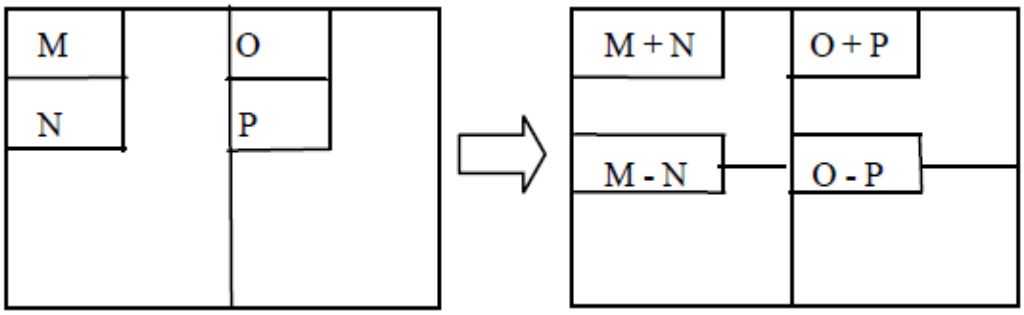


Figure 6 The vertical operation

2.2.1.2 Audio Steganography

Audio steganography refers to concealment of digital information, for example text messages, binary files and documents, into audio records, for example, WAV, MP3, and RM records [15]. The carrier file is produced which is the output audio file which is send to the recipient. Audio steganography exploits the Human Auditory System that can't perceive the little variation in the audio frequencies at the high frequency side of the audible spectrum and in this way, sound steganography can make use of this sort of frequencies to conceal confidential information without affecting the size much and also without causing any damage to the quality of audio file. The abundance and popularity of audio file make them eligible for transmitting confidential data. Thereby, numerous scientists begun to explore how audio signal and its properties can be utilized as a part of data hiding approaches. Various techniques were perceived, the most common ones are, Echo hiding, Least Significant Bit, Hiding in Silence Interval, Amplitude Coding, Spread Spectrum, Phase Coding and Discrete Wave Transform.

2.2.1.3 Text Steganography

Concealing data in plain content could be possible in a wide range of ways. A few procedures comprise of changing the layout of the carrier content, for example, including white-spaces or modifying the case of specific characters in order to embed the confidential data. Others, comprise of relating the characters to cover up with the characters of the carrier content, making a reference dictionary that maps words from the private content with words from the carrier content. This

segment reveals throws insight into the different methods utilized as a part of text steganography including hiding in HTML, hiding by selection, hiding using whitespace, semantic-based hiding, line and word shifting and abbreviation-based hiding techniques.

CHAPTER 3

RELATED WORK

Cryptography and Steganography are widely researched areas and various techniques have been developed combining several approaches of both the fields. It has resulted in several security systems being designed with extra layer of protection for the secret data. There has been a great amount of research in the cryptography since the advent of internet but steganography is relatively new. Several approaches of both the techniques have proven their worth over time. Although both techniques are used to provide data security, a study is made to combine both cryptography and steganography methods into single system for better confidentiality and security.

In this chapter we have described the literature survey of previously existing data security techniques:-

X. Qing, et al., [1] gave a new approach where message is concealed in RGB components of an image thereby making use of all the planes and hence taking advantage of limits of Human visual system.

In [20] author suggested LSB approach with some improvement is published. The suggested work states confidential data is only embedded in blue component of the RGB model. This approach helps in decreasing the deterioration of the output stego image as only blue components are being used to insert the confidential data in the image carrier.

H. Yang, et al. proposed a variation of LSB steganography known as adaptive LSB image steganography, which uses a pixel adjustment approach for improvement of encoded image quality. This technique also help in providing high payload capacity.

Authors in [20] have encoded the private data with the help of vector quantization table. This approach has also improved the hidden data capacity.

Nouf A. Al-Otaibi, et al.[12], developed a new approach of blending steganography and cryptography resulting in two layer data security for concealing the secret data on personal computers. They divided the system in 2 layers i.e, steganography layer and cryptography layer. LSB algorithm is used is used for steganography laye. For cryptography DES is used. Authors has also done study to improve hidden data capacity. Drawback of this method is that DES is not fully secure hence this method may fail to secure private data.

K. B. Raja, et al. [4] made use of various blending of steganographic techniques like LSB, DCT and also the compressing is carried out to provide better security.

In [22] authors proposed an approach for protection of image in open wireless channel. The secret image is embedded in the cover image using LSB technique from spatial domain. Then the stego image is divided into 8*8 blocks. The encryption of the divided stego image is done by double random phase encoding which transforms the image into white stationary noise.

In [10] suggested a better approach to steganography using DWT and the analysis using PSNR shows that the proposed technique gives better results.

In [25] one of the paper authors integrated RSA cryptography and audio steganography. The secret data is converted to encrypted text with the help of RSA algorithm and the encrypted text is hidden in audio using LSB audio technique. By combining steganography and cryptography it produces the higher level of security.

In [29] authors proposed a new approach of image steganography on gray image combined with cryptography. The secret data is converted to cipher text using Vernam cipher and the data is encoded in the cover image using LSB with shifting. Here the sender and the receiver shares one time pad key for Vernam cipher. The authors claim that message concealing capacity of this approach has increased drastically.

In one of the paper authors presented 2 new approaches to secure data. In the first approach each byte of the secret image is encrypted using S-DES algorithm to produce an array of encrypted

pixels. Each element of array is then divided into 2 parts where first part contains first 4 MSB's and second part contains remaining LSB's. Then each pixel value is converted with alphabets from A to P where A is assigned to 0000 and P to 1111. The output will be an encrypted image containing text. The encrypted image is then embedded in cover image by XOR method. In the second approach they simply encrypted the secret image using S-DES algorithm and embed it in the cover image as stated above.

In [19] authors has given a hybrid method for image security that provides good encryption quality. The secret image is encrypted using blowfish algorithm to produce the cipher image. Then the encrypted image is embedded using LSB technique in the cover image. Blowfish algorithm is lossless and highly secured encryption technique.

In [18] authors proposed a method that increase the security of data transfer by combining cryptography and steganography. Mp3 file is taken as the cover media and the secret message is encrypted using AES algorithm using a key that has been processed by MD5 hash function. The secret message was inserted in the homogeneous frame in mp3 files with addition of a key code. The MD5 algorithm is being used widely cryptographic hash function used to verify data integrity.

CHAPTER 4

Data Security Approaches Used

A digitalized image consists of a number of atomic units called pixels. Here we have made use of color image for the purpose of steganography. The pixel of a color image is formed by combination of red, green and blue component hence three plains as compared to single plane in grey image.. The binary representation of a color level of each component is done in 8 bits. Hence the total number of bits required to represent a pixel is 24 bits. Thereby an image could be represented by 2D array of pixel values with every value correspond to some blending of red, green and blue. We can use successive bytes to store the information but before that our message is to be converted to encrypted text which is then converted to bits and then stored in the image.

This security approach is based on two main sub-parts:

i. Transforming the private data to cipher or encrypted text by using AES Cryptography

ii. Concealing the encrypted text received from above step into image by a steganographic technique like LSB, DCT, DWT

AES Cryptographic algorithm takes a key as input which can be of variable length depending on the implementation and then perform encryption by converting the plain text into encrypted text. This encrypted text can then be embedded into a cover image using any image Steganographic technique like DCT, LSB, DWT on any image format. These technique were then analyzed for different image formats and comparative study is performed.

4.1 Diffie-Helman

Diffie-Hellman is not an encryption mechanism as we regularly consider them, in that we don't commonly utilize DH to encrypt data. Rather, it is a strategy for secure exchange of the keys that encrypt data. Diffie helman performs this protected exchange by making a "shared secret" (once

in a while called a "Key Encryption Key") between two devices. The shared secret then encrypts the symmetric key for secure transmittal. The symmetric key is some of the time called a "Trafifc Encryption Key" or "Data Encryption key". The procedure starts when every side of the correspondence generates a private key. Every side then produces an public key, which is a derivative of the private key. The two systems then exchange their public keys. Every side of the correspondence now has its own private key and the other systems's public key. This technique helps to defend against “men in middle attack”.

Diffie-Hellman Key Exchange

Step	Alice	Bob
1	Parameters: p, g	
2	$A = \text{random}()$ $a = g^A \pmod p$	$\text{random}() = B$ $g^B \pmod p = b$
3	$a \longrightarrow$ $\longleftarrow b$	
4	$K = g^{BA} \pmod p = b^A \pmod p$	$a^B \pmod p = g^{AB} \pmod p = K$
5	$\longleftarrow E_K(data) \longrightarrow$	

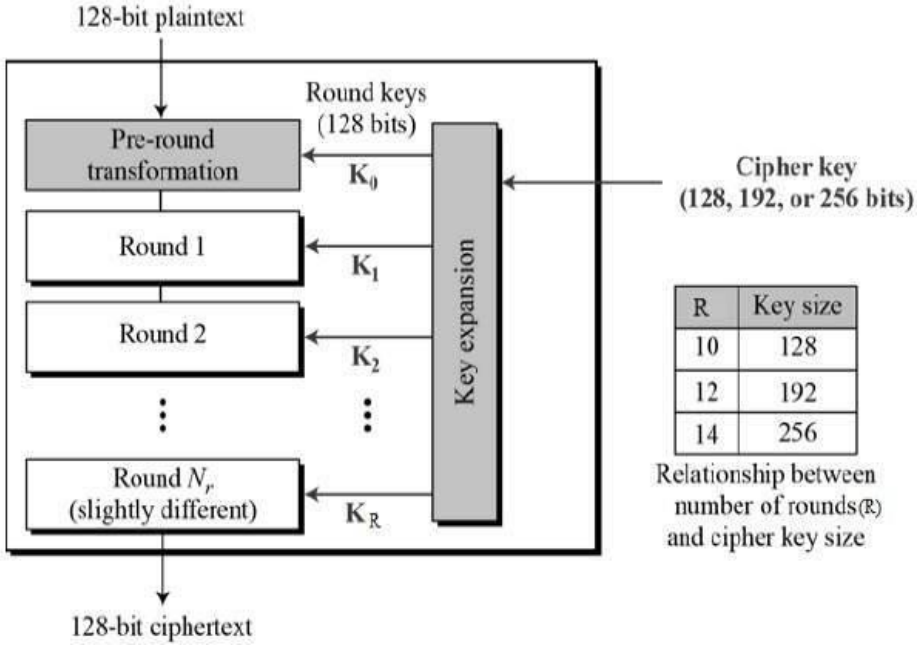
4.2 AES

The most widely used and adopted symmetric encryption algorithm has gained immense popularity is the Advanced Encryption Standard (AES) [27]. AES performance is better than the TDES as it is faster. Also DES was not as secure because of short key length. As the computer system’s computation powers have increased exponentially in the last decade or so, DES has gone secure to vulnerable from various key attacks. After that TDES was developed to improve on key length of DES but it is found to be too slow.

AES uses iterative approach in place of Feistel cipher. The basis of AES was several substitution, permutation operations. Various interconnected operations form the AES. One of the operation is

substitution which include replacement of some input by some specific output and other operation like permutation involves shuffling of bits.

The computations of AES is done on bytes instead of bits, for instance the algorithm would treat the 128 bit plain text as 16 bytes. A matrix of 4x4 is formed by those 16 bytes. In AES the number of rounds are not fixed and they would be based on the basis of key length chosen. If the key length is 128 bit then number of round would be 10, if key length is 192 bit then 12 rounds and if key length is 256 bit then 14 rounds. A 128 bit round key is computed which is computed from original key and is different for each round.



R	Key size
10	128
12	192
14	256

Relationship between number of rounds(R) and cipher key size

Figure 7 AES structure

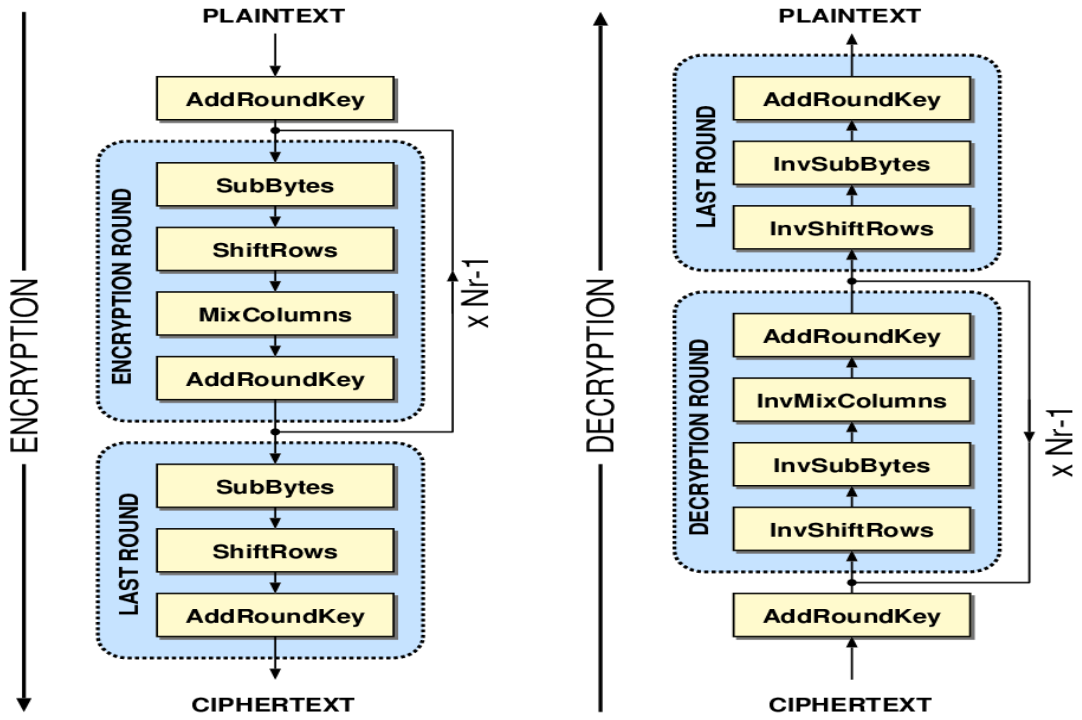


Figure 8 AES Encryption and Decryption Process

Each of the round comprises of four sub-processes. Every sub-process is described diagrammatically.

Byte Substitution (SubBytes)

The substitution process is done by looking up a fixed table called S-box and replacing the 16 input bytes. The output is stored in a 2D array of four columns and four rows.

The substitution process is described below with the help of a diagram:

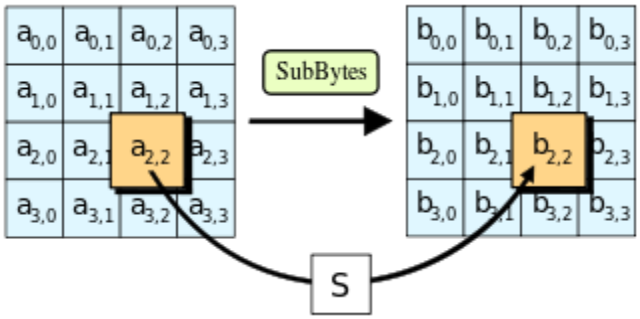


Figure 9 Substitute Bytes

Shiftrows

After substitution the left shifting is performed on each of the four rows. The shifting is circular nd the entries falling off are inserted to the right side of the row. Shifting is done as follows –

No shifting of first row is done then shifting of second row is done one (byte) position to the left. This process goes on till fourth row with increment of one in next row shifting.

There would be same 16 bytes in the 2D array but only shifting takes place in reference to each other.

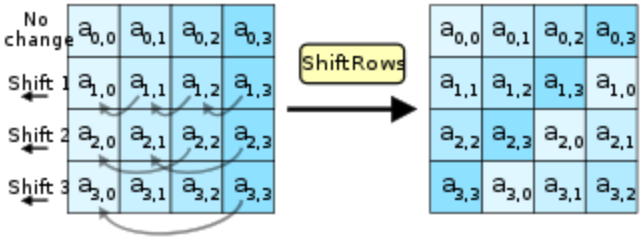


Figure 10 Shift row operation

MixColumns

In this round column transformation takes place with the help of some mathematical function. The 4 byte of a column is taken as input by this function which would then result in 4 transformed new bytes, and replacing the original column bytes.The output is completely new matrix which consists of 16 new bytes. In last step this step is not performed.

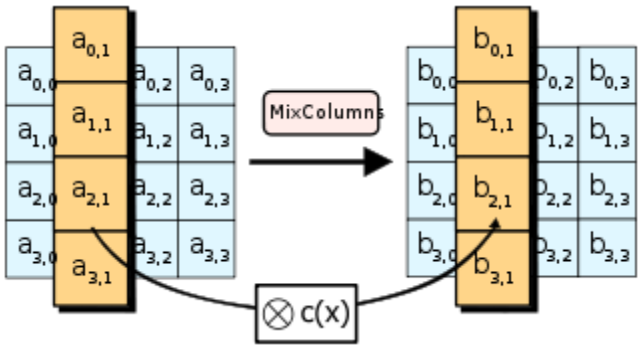


Figure 11 Mix Column operation

Addroundkey

Now XOR operation is done between 128 bit of 2D array and the 128 bits of the round key. The cipher text would be the output of this step after last round. Otherwise the process is repeated treating 128 bits as 16 bytes.

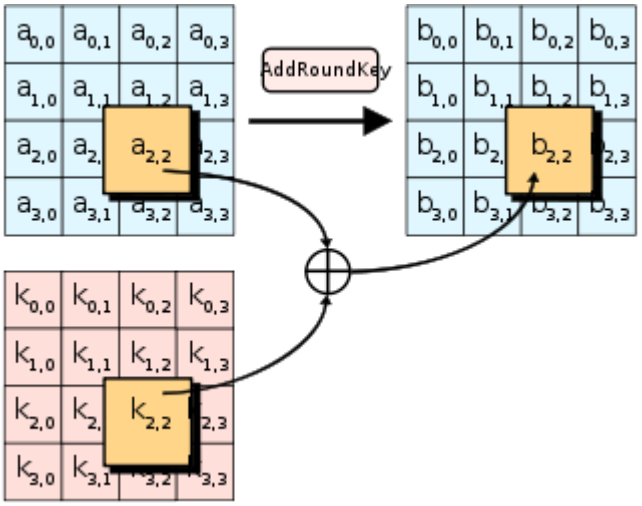


Figure 12 Add Round Key operation

Nowadays both hardware and software supports AES and is being used widely. Uptill AES has proven to practically full proof agains cryptanlytic attacks. AES also has in-built flexibility of key lengths, that allow a degree of "future proofing" against advancement in the ability to perform exhaustive key searches. However as in DES, the security of AES is only assured if it has good key management and is implemented correctly.

4.3 Steganography encoding and decoding algorithms

A. **The LSB based encoding steganography algorithm**

The steps can be described as follows:

i. Taking the seret message as an input and choosing a cover image for message concealment.
ii. Conversion into grey image from color image.
iii. Converting the secret text message into binary format.
iv. Finding the LSB of every pixel of the cover image.
v. Replacing the LSB of cover image with each bit of secret text message from first bit to last one.
vi. Writing the stego image
vii. Computation of SSIM, MSE and PSNR of the encoded image to analyse the performance of this method.

B. The LSB decoding algorithm

The steps can be described as follows:

i. Reading the enoded image.
ii. Calculating the LSB of every pixels of stego image.
iii. Retrieving bits and converting each 8 bit into letter or character.

C. The DCT Based Steganography encoding algorithm

The steps can be described as follows:

i. Reading the carrier image.
ii. Reading text message and convert it into binary format.
iii. The carrier image is then broken into 8x8 blocks of pixel.
iv. 128 is substracted from every block of pixel working horizontally as well as vertically.
v. Direct cosine transform is then applied to each of the block.
vi. Compression of every block is then performed using the quantaization table.
vii. Computing the LSB of each direct cosine coefficient and then replacing with each bit of the private text.

viii. Writing the stego image.

ix. Computation of the SSIM metric, MSE and PSNR of the encoded image to analyse the performance of this method.

D. The DCT decoding algorithm

i. Reading the encoded image

ii. The stego image taken as input to this method would first divide the input image as blocks of 8x8 pixels.

iii. Inverse DCT is then performed on every block of 8x8 pixels.

iv. Then compression is carried out on every block by using quantization table.

v. Then the LSB is extracted from every Direct Cosine Coefficient.

E. The DWT Based Steganography encoding algorithm

i. Reading the carrier image and secret text that is to be concealed in the carrier image.

ii. Converting the secret text into the binary format. Apply 2D Haar transformation on the carier image.

iii. Obtain the four sub-bands with different features like LL, LH, HL, HH. Then performig the embedding on the DWT coefficient of the LL band.

iv. Obtaining stego image.

v. Computing the MSE and PSNR of the encoded image to analyse the performance of this method.

F. The DWT decoding Algorithm

i. Take stego image as input.

ii. The coefficients i.e, both horizontal and vertical are obtained from the stego image.

iii. Extracting the data bit by bit and re-composing the secret data

iv. Converting the data into message vector. Comparing it with original message

4.4 A combined Security approach

Embedding process using cryptography and steganography

Both the approaches are joined by performing encryption of the message using cryptography and concealing the encrypted text based on steganography. The Advanced Encryption Standard(AES) is used for performing encryption. The conversion of plain text to cipher text therefore takes place ttext using AES. The cipher text is then hidden in a cover image which could have any format like jpg, bmp, png etc. The stegno method could be any of the steganography algorithms(LSB, DCT, DWT) based on the suitaibility considering several performance factors like capacity, security and robustness. The output is the stego image embedded with the cipher text.

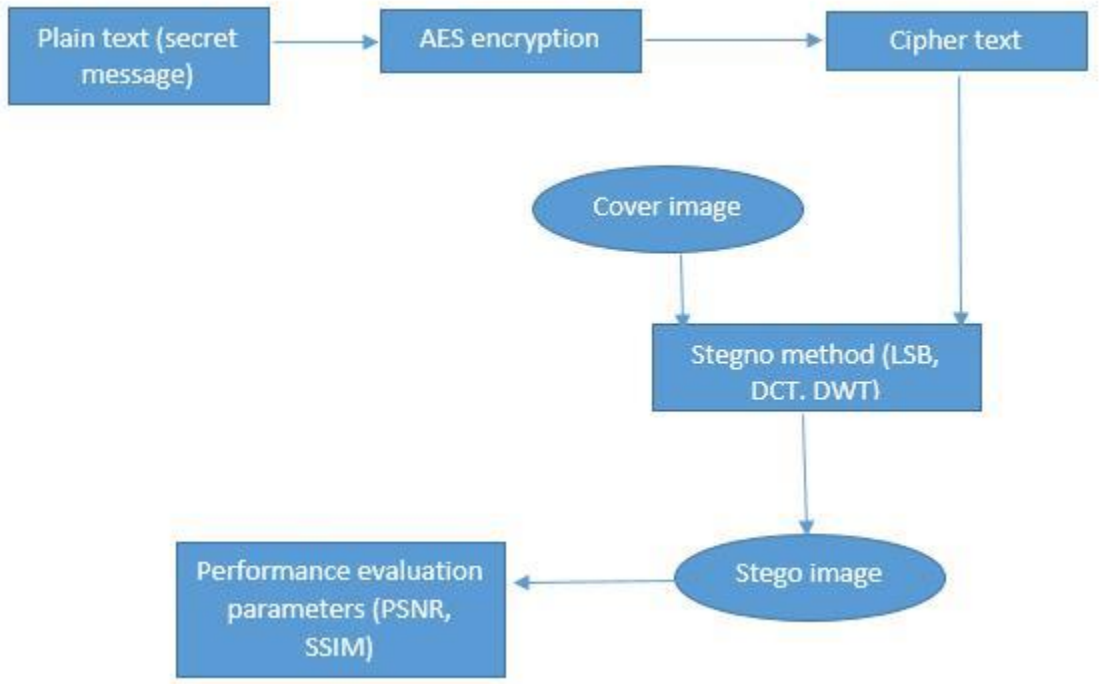


Figure 13 Embedding process and performance evaluation

Extracting process

The extraction of message take place in reverse order of embedding process. The stego image with hidden cipher text is fed to the steganography algorithm chosen in the embedding process. The

corresponding decoding method returns the cipher text which is then pass through AES decryption method to get the secret or plain text.

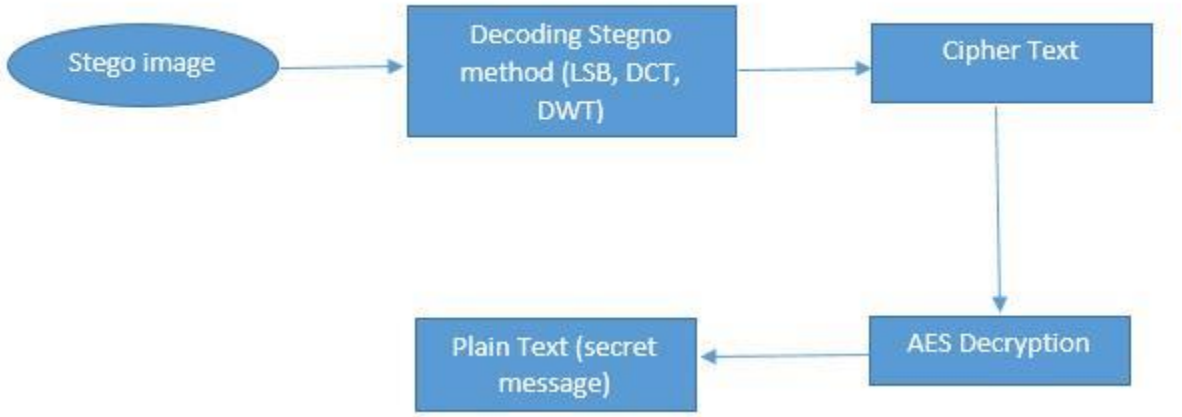


Figure 14 Extracting process

CHAPTER 5

EXPERIMENTAL RESULTS AND ANALYSIS

In this chapter, simulation of AES cryptography in color image steganography is presented. Various steganography methods like LSB, DCT, DWT are also compared with respect to various parameters so that a better technique could be applied according to the requirement and domain.

Here we will discuss the experimental setup of the research work done. First section will discuss the sample images used for steganaography followed by programming tools used for programming. In the next section, a case study is discussed to elaborate the working. In the last section summary of the chapter is given.

5.1 Programming Tools and software used

We have used four images to carry out our analysis. All the images have different file format. The file format of the images are jpg, png, bmp, tif. The specification of the images are described in later section. The implementation of the stegno-crypto framework is carried out in Visual Studio 2013.

The later analysis of steganographic method is done using MATLAB. To evaluate the performance of the different approaches we have used different metrics like PSNR, MSE, SSIM etc. Specification of system used to simulate our work are:

Operating System: Windows 8.1

Language used: C#, MATLAB

Memory: 3 GB

Processor: i3

5.2 Stegano-Crypto framework

The stegno- crypto application is developed that gives a robust security system. The LSB-AES blending is used to provide two level security. The performance of the method can be measured with some tests such as MSE, PSNR. The LSB AES based image data hiding technique is implemented.

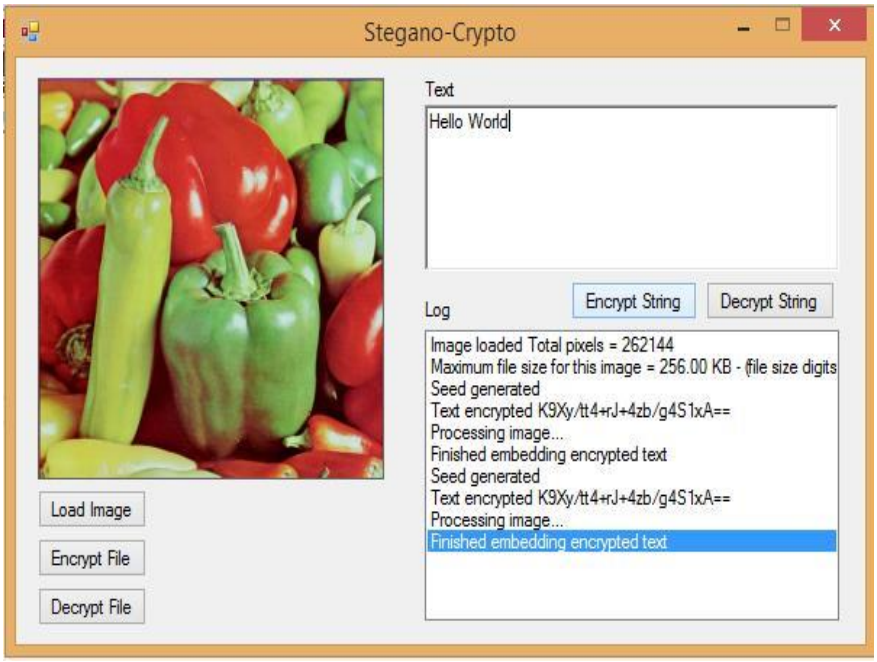


Figure 15 Stegano-Crypto framework

To analyze the performance of the security scheme, a number of different types of colored image formats (BMP, JPEG and TIF) are tested. The stego images looks undistorted, therefore it can be inferred that the implemented system produces high quality of the stego images.

The above implemented framework takes a color image and a confidential message or secret file as input. The input is then encrypted using AES and it is then embedded in the image. The output is saved in the stego image. The decrypt file or decrypt string can then be used to obtain the secret message or file embedded in the stego image. The framework works to embed a file or message in

any type of image format like *.jpg, *bmp, *.png. The process of embedding is shown in the figure 5.2 below.

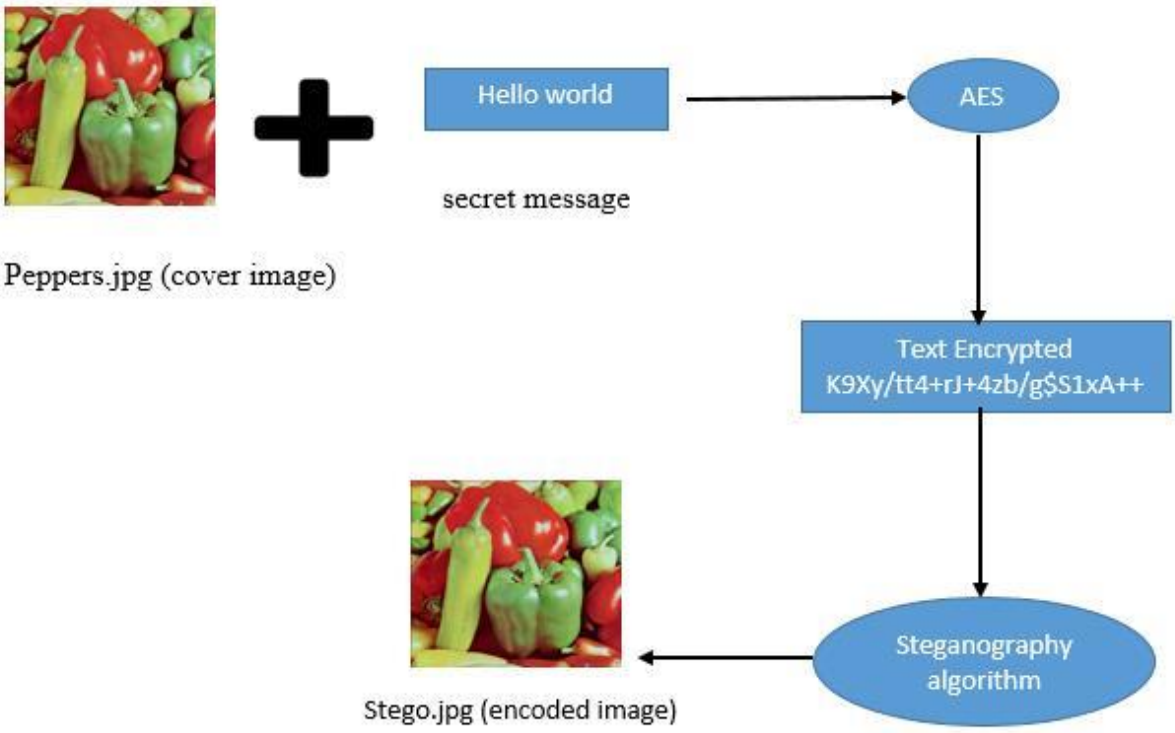


Figure 16 Process of stegano-crypto framework

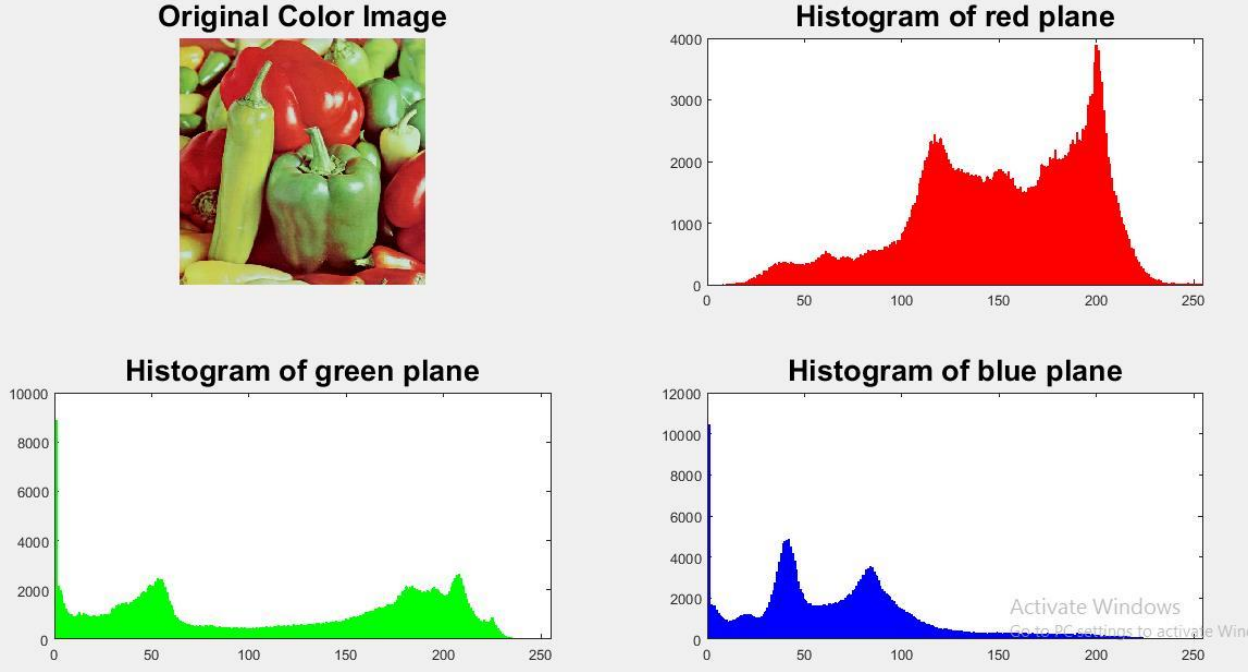


Figure 17 Histogram plot for RGB plane of peppers.jpg

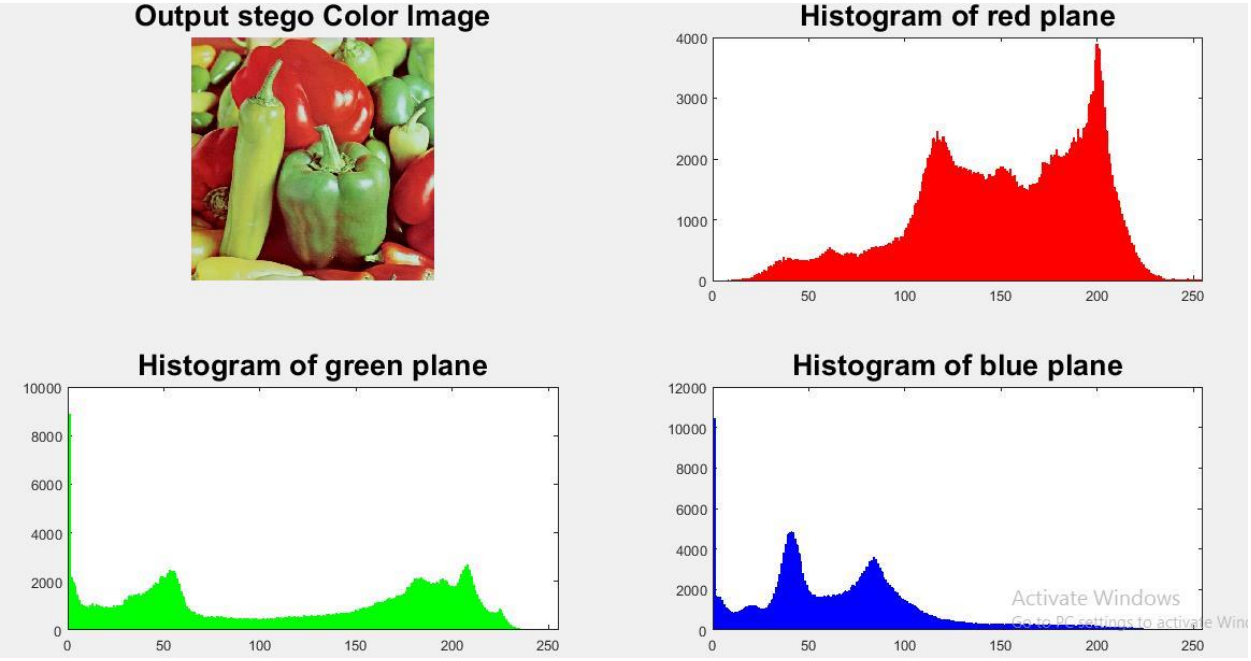


Figure 18 Histograms of output image stego.jpg

The histogram analysis shows that there would be little difference in histograms of cover and stego image which is not visible as such. This means the stego image does not get suspicious on first viewing by the attacker.

5.3 Performance analysis parameters

The performance evaluation should be carried out so that proper technique could be selected as and when required. All the three algorithms i.e, LSB, DCT and DWT need to be compared against various parameters for instance payload capacity, robustness, PSNR, SSIM etc. PSNR is mainly used for evaluation of these techniques, it is the ratio of peak signal to the noise. This ratio better illustrates the quality difference between two images. If PSNR ratio is high then image produced when compared with original image is of high quality. Due to some drawbacks in PSNR, SSIM i.e, structural similarity metric is now being used for better comparison.

The various parameters for performance analysis are stated as down under:

1. **Capacity**

   It represents the amount of information that can be hidden in the carrier. It is given by the formula

$$Capacity = \frac{total\ no\ of\ bits\ embedded\ in\ the\ cover\ image}{total\ no\ bits\ in\ the\ cover\ imge} \quad (3)$$

2. **MSE(Mean Square error)**

   It refers to the square of error between cover image and the stego image. MSE helps to measure the distortion in the image.

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \quad (4)$$

I(i,j) and *K(i,j)* are the pixel values of cover image and stego image respectively. M and N represents the number of rows and columns in the input image.

$$MSE = \frac{1}{3mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \quad (5)$$

As RGB color image has three planes thereby the equation further divided by a factor of 3.

3. PSNR(Peak signal to noise ratio)

It is defined as the ratio of the peak or maximum signal to noise with respect to original and stego image [28]. PSNR is measured in decibels (dB). Its formula is depicted by equation 6:

$$PSNR = 10\log_{10}\left(\frac{MAX^2}{MSE}\right) = 20\log_{10}\left(\frac{MAX}{\sqrt{MSE}}\right) \quad (6)$$

MAX is the largest possible value in an image.

MSE is the mean square error between pixel of two images

If bit depth is 8 bit, then the value of *MAX* would be 255. A higher PSNR generally indicates that the re-construction of image has resulted in high quality. For color image with three components of R,G,B values per pixel, the PSNR definition would be same except the MSE is computed as the sum over all squared value differences divided by image size and by three. PSNR measuring unit is dB(decibels).

4. SSIM(Structural Similarity Index)

The SSIM is a method developed for estimating the perceptual quality of digital image and other media. This method finding similarity between two images when noise is added on the basis of human visual perception [7]. The equation for the same is given by:

The computation of SSIM index is done on various windows of an image. If we are given two windows a and b of same size MxM then similarity measure would be computed by:

$$SSIM(a,b) = \frac{(2\mu_a\mu_b + c_1)(2\sigma_{ab} + c_2)}{(\mu_a^2 + \mu_b^2 + c_1)(\sigma_a^2 + \sigma_b^2 + c_2)} \quad (7)$$

μ_a and μ_b are the mean of the two windows a and b respectively.

σ_a and σ_b are the variance of the two windows a and b respectively.

$\sigma_a\sigma_b$ is the covariance of a and b.

$C_1=(k_1L)^2$ and $C_2=(k_2L)^2$ where L= $2^{\text{no of bits per pixel}}$-1 and $k_1=0.01$, $k_2=0.03$.

5.4 Comparative analysis of LSB, DCT, DWT

Below images are used to carry out the analysis of various method of steganography. Below table shows the properties of images used in testing. All the test images used are of different format for better analysis.

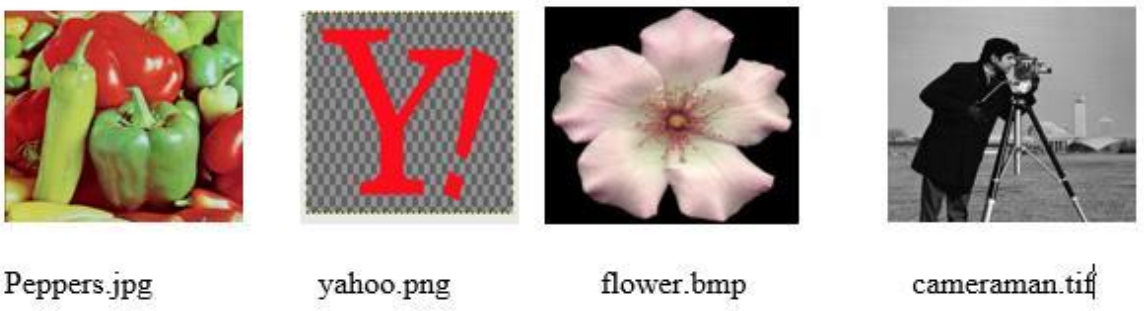


Figure 19 Sample test images

Image type	**Image size**	**Total pixels**
Peppers.jpg	256KB	262144
yahoo.png	183KB	187704
Flower.bmp	21.7KB	22201
Cameraman.tif	64KB	65536

Table 1: Sample image properties

LSB based substitution

The LSB based Substitution steganography is applied on the four images with different format and various performance evaluation parameters are evaluated as shown in table below:

Image type	**MSE**	**PSNR**
Peppers.jpg	0.1716	60.5565
yahoo.png	0.0352	67.4334
Flower.bmp	0.3432	57.5468
Cameraman.tif	0.0205	65.0084

Table 2: MSE and PSNR values for LSB technique

DCT based substitution

Image type	**MSE**	**PSNR**
Peppers.jpg	3.7913	40.8593
yahoo.png	25.9718	29.2145
Flower.bmp	22.2600	29.8843
Cameraman.tif	8.7765	38.6976

Table 3: MSE and PSNR values for DCT technique

DWT based substitution

Image type	MSE	PSNR
Peppers.jpg	0.7635	49.3025
yahoo.png	0.0364	65.3225
Flower.bmp	0.8354	43.0435
Cameraman.tif	0.0139	66.6861

Table 4: MSE and PSNR values for DWT technique

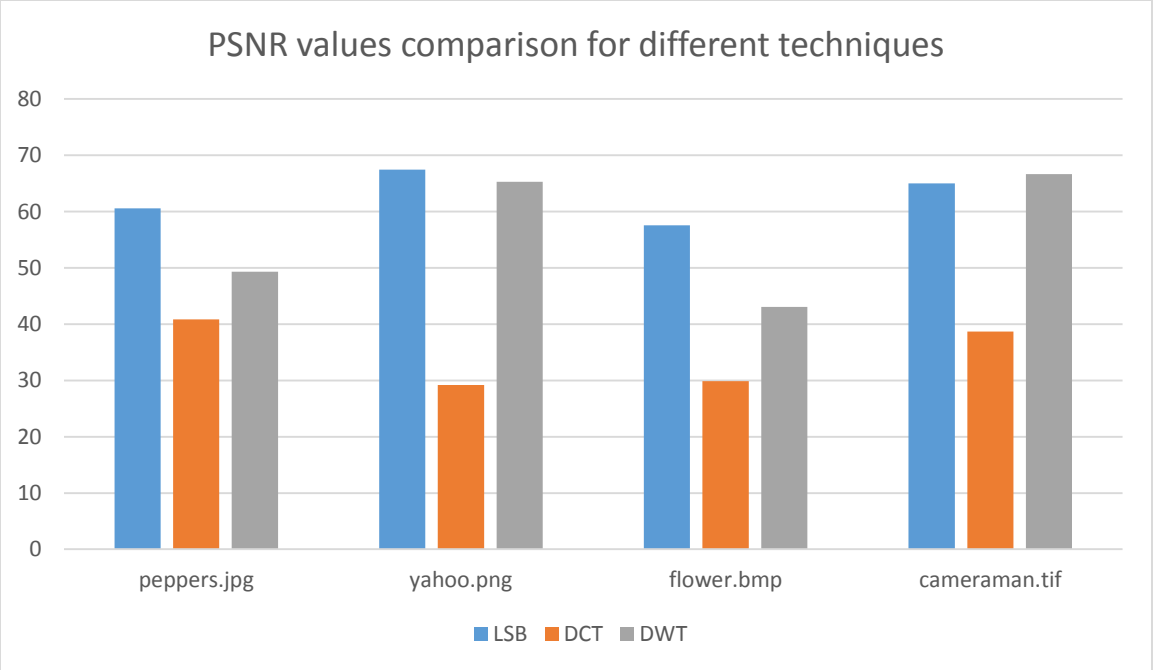


Figure 20 PSNR comparison for LSB, DCT and DWT for different image formats

The PSNR analysis shows that LSB and DWT gives best results and therefore best embedding. The embedding by LSB is easy to detect by steganalysis as compared to LSB. In the work carried

out the PNG and TIF formats provide better results in comparison with other file format. The MSE is proportional to PSNR as their respective values shows with DWT and LSB having least error.

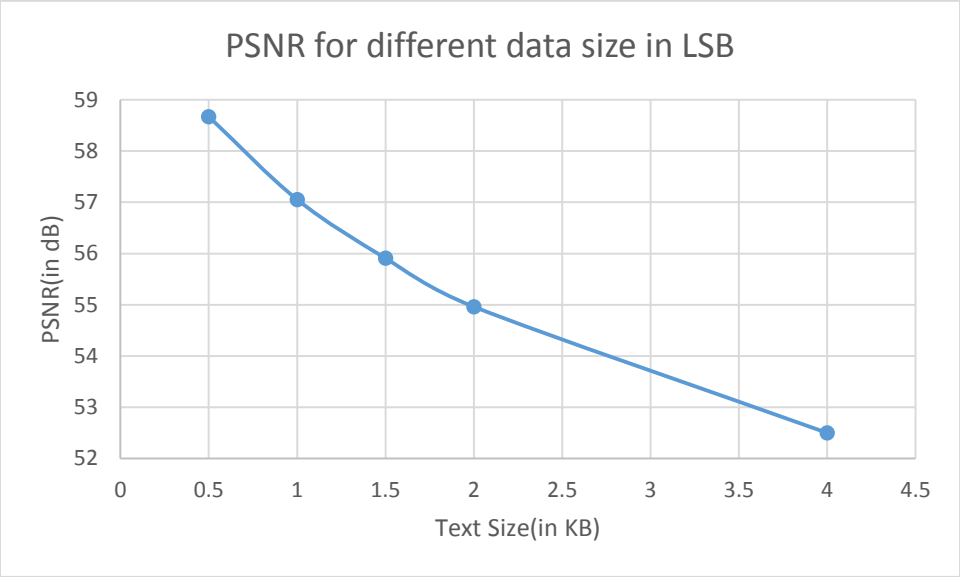


Figure 21 PSNR comparison for different secret data size in LSB for peppers.jpg

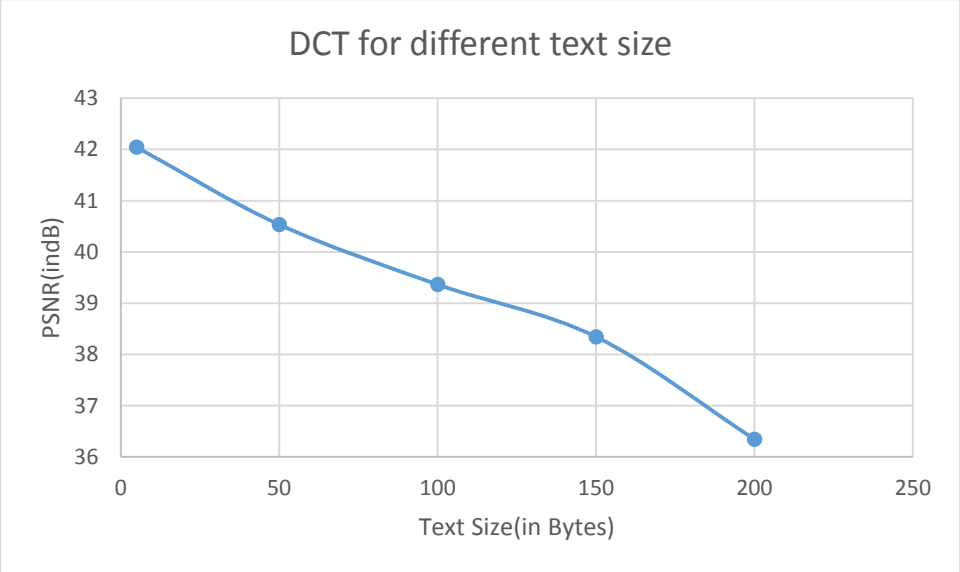


Figure 22 PSNR for different text size in DCT for peppers.jpg

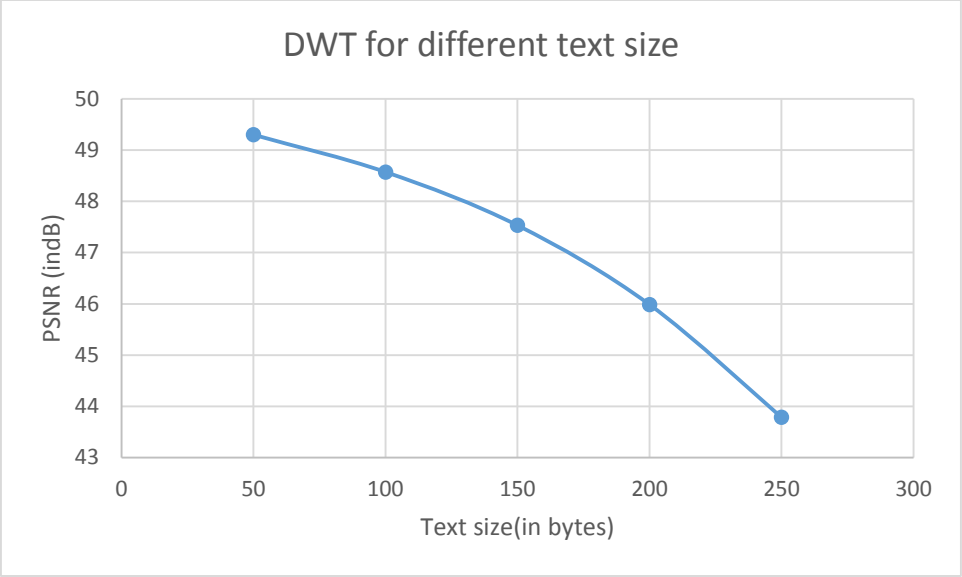


Figure 23 PSNR for different text size in DWT for peppers.jpg

PSNR method proves to be inconsistent with the human visual perception. SSIM(structural similarity index) was designed to make improvement on traditional methods like PSNR, MSE etc. PSNR do pixel to pixel matching and error finding whereas SSIM deals with the window matching hence covering a pixel as well as its neighbors. Thereby SSIM covers more than one pixel i.e, window matching and hence is more consistent to the human visual perception.

SSIM evaluation for different techniques

Image type	**SSIM for LSB**	**SSIM for DCT**	**SSIM for DWT**
Peppers.jpg	1.0000	0.9981	0.9997
yahoo.png	0.9999	0.8975	1.0000
Flower.bmp	0.9991	0.8762	0.9996
Cameraman.tif	0.9998	0.9540	0.9999

Table 5: SSIM values for LSB, DCT and DWT techniques

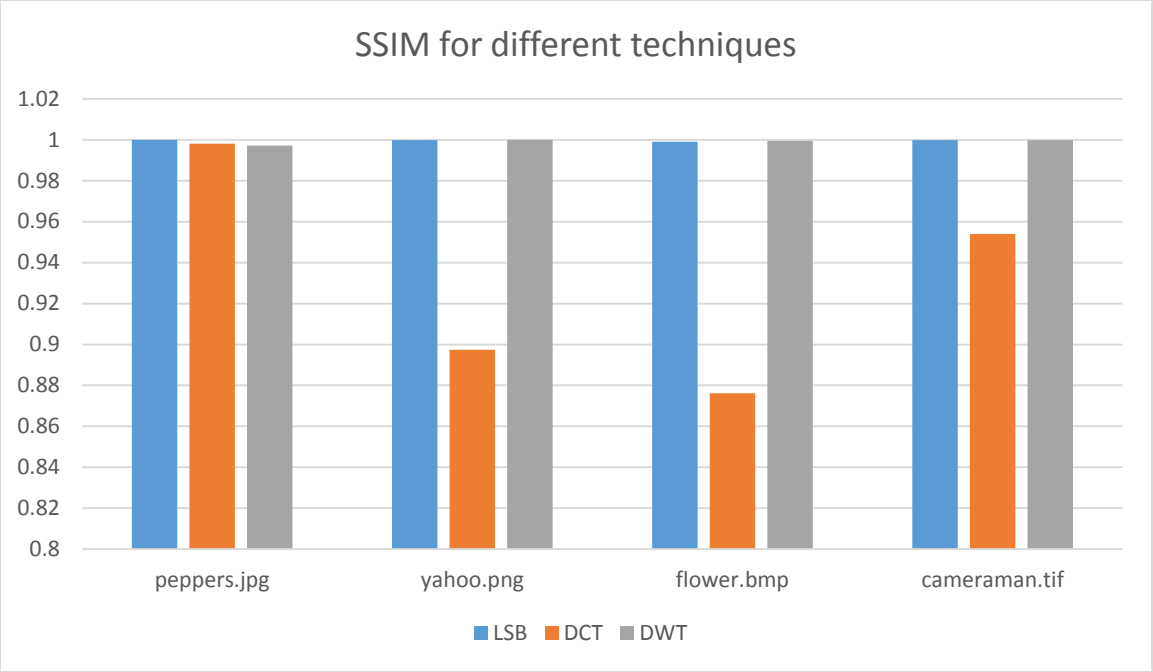


Figure 24 SSIM comparison for different technique for different image format

SSIM analysis shows the same results as compared to PSNR i.e, the LSB and DWT gives more structural similarity whereas DCT performs poor in this metric as well. DWT gives slightly better results to LSB and also detection in LSB is easier hence DWT trump over LSB. A point also be noted is that JPG performs best for SSIM in all the three techniques and hence most difficult to detect any hidden text with optical perception.

Parameters	**LSB**	**DCT**	**DWT**
Payload capacity	High	Medium	Low
Robustness	Low	Medium	High
Imperceptibility	High	Low	High
MSE	Low	High	Low
PSNR	High	Low	High
SSIM	Medium	Low	High
Encoding and decoding time	Low	High	Medium

Table 6: Comparison with several other parameters

Although the payload capacity for LSB method is high in comparison to DCT and DWT but the robustness in other two methods i.e, detection through steganalysis especially small messages is difficult as message bits are stored in transformed coefficients.

CHAPTER 6

CONCLUSION AND FUTURE WORK

The attackers or intruders are more equipped than ever before and hence data security while in transmission over a network becomes of utmost importance. The two layer security of using AES encryption algorithm along with steganography in color image discussed in this work can prove its worth in transmission of secret data over an insecure channel. The histogram comparison of cover and stego image does not show any visual change. The steganographic methods compared have proven that DWT gives better result when compared with LSB and DCT with respect to several parameters like robustness, PSNR, SSIM, invisibility etc. DWT takes less encoding and decoding time as compared to other two technique. As LSB is prone to statistical attack therefore DCT or DWT will be much better to use. The techniques have been applied to different image file format which results in JPG format trumping over other format when SSIM index is compared for all the techniques.

From this work it could be concluded that there is compromise between the three characteristics of steganographic algorithm i.e, imperceptibility, embedding capacity and robustness. The future work should involve around designing new technique which preserve all the three main factors. Hash functions could be used with some other cryptographic algorithm for integrity and security along with steganography for a better system. Hamming code can be used for further improving the PSNR.

REFRENCES

[1] X. Qing, X. Jianquan and X. Yunhua, "A High Capacity Information Hiding Algorithm in Color Image", Proceedings of 2nd IEEE International Conference on E-Business and Information System Security, Wuhan, China, 2010.

[2] Patel H, Dave P. Steganography Technique Based on DCT Coefficients. International Journal of Engineering Research and Applications, 2(1):713–7, 2012.

[3] Nikhil Patel, Shweta Meena, "LSB Based Image Steganography Using Dynamic Key Cryptography", International Conference on Emerging Trends in Communication Technologies (ETCT), 2016.

[4] K. B. Raja, C. R. Chowdary, K. R. Venugopal and L. M. Patnaik, "A Secure Image Steganography using LSB, DCT and Compression Techniques on Raw Images", Proceeding of 3rd IEEE International conference on Intelligent Sensing and Information Processing (ICISIP), 2005.

[5] P. Campisi, O. Kundur, O. Hatzinakos and A. Neri, Compressive data hiding: an unconventional approach for improved color image coding, EURASIP 1. on Applied Signal Processing, 152- 163, 2002.

[6] https://en.wikipedia.org/wiki/Structural_similarity

[7] H. Mehra, V. Kapse, T. K. Sahu, G. Tiwar, "An Extensive Literature Survey on Steganography using Genetic Algorithm", International Journal of Scientific Progress and Research (IJSPR) vol.13,123-126, 2015.

[8] X. Luo, F. Liu, C. Yang, S. Lian, and Y. Zeng, "Steganalysis of adaptive image steganography in multiple gray code bit-planes," Multimed. Tools Appl., vol. 57, no. 3, pp. 651-667, 2012.

[10] Mandal J.K. and Sengupta M., "Authentication/Secret Message Transformation Through Wavelet Transform based Sub-band Image Coding (WTSIC).", Proceedings of International Symposium on Electronic System Design, IEEE Conference Publications, pp 225 – 229, 2010.

[11] Idrizi, Florim, Dalipi, Fisnik & Rustemi, Ejup. "Analyzing the speed of combined cryptographic algorithms with secret and public key". International Journal of Engineering Research and Development, e-ISSN: 2278-067X, p-ISSN: 2278-800X, 2013

[12] N. A. Al-Otaibi and A. A. Gutub, "2-Leyer Security System for Hiding Sensitive Text Data on Personal Computers", Lecture Notes on Information Theory, vol. 2, no. 2, (2014).

[13] Chan CK, Cheng LM Hiding data in images by simple LSB substitution. Pattern Recognition,37:469–474, 2004.

[14] Johnson NF & Jajodia S Exploring steganography: seeing the unseen. Comput Pract,26–34, 1998.

[15] Swanson M, Kobayashi M, Tewfik A Multimedia data embedding and watermarking technologies, Proc IEEE 86(6):1064–1087, 1998.

[16] S. Kartalopoulos, "Security of Information and Communication Networks", Wiley-IEEE Press, 2009.

[17] RiniIndrayani, Hanung AdiNugroho, Risanuri Hidayat, Irfan Pratama, "Increasing the Security of MP3 Steganography Using AES Encryption and MD5

Hash Function", International Conference on Science and Technology-Computer (ICST), IEEE, 2016.

[18] Y. Lee, L. Chen, "High capacity image steganographic model", IEEE Proceedings on Vision, Image and Signal Processing, 147, 288 -294,2000.

[19] Moresh Mukhedkar, Prajkta Powar and Peter Gaikwad, "Secure non real time image encryption algorithm development using cryptography & Steganography", IEEE INDICON, 2015.

[20] Chang CC, Lin MH, Hu YC A fast and secure image hiding scheme based on LSB substitution. Int J Pattern Recog, 16(4):399–416, 2002.

[21] Wang RZ, Lin CF, Lin JC (2001) Image hiding by optimal LSB substitution and genetic algorithm. Pattern Recogn, 34(3):671–683, 2001.

[22]Singla, Deepak, and RupaliSyal. "Data Security Using LSB & DCT Steganography in Images." International Journal Of Computational Engineering Research 2, 359-364, 2013.

[23] Walia, Ekta, Payal Jain, and NavdeepNavdeep. "An analysis of LSB & DCT based steganography." Global Journal of Computer Science and Technology 10.1, 2010.

[24] Kingslin S, Kavitha N. Evaluative Approach towards Text Steganographic Techniques. Indian Journal of Science and Technology, 8(29) Doi:10.17485/ijst/2015/v8i1/84415, Nov 2015.

[25] Ankit Gambhir and Sibaram Khara, "Integrating RSA Cryptography & Audio Steganography", IEEE ICCCA, 2016.

[26]https://www.tutorialspoint.com/cryptography/advanced_encryption_standard.html

[27] https://en.wikipedia.org/wiki/Advanced_Encryption_Standard

[28] https://en.wikipedia.org/wiki/Peak_signal-to-noise_ratio

[29] Kamaldeep Joshi, RajkumarYadav, "A New LSB-S Image Steganography Method Blend with Cryptography for Secret Communication", IEEE ICIIP, 2015.

[30] Chih-Ching Thien and Ja-Chen Lin, A simple and high-hiding capacity method for hiding digit-by-digit data in images based on modulus function, Pattern Recognition 36, 2875-2881, 2003.

[31] Anderson, R.J. & Petitcolas, F.A.P, On the limits of steganography. IEEE Journal of Selected Areas in Communications, 16(4), pp.474-81, 1998.

[32] Hwang, R.J., Shih, K.T., Kao, C.H. & Chang, T.M. Lossy compression tolerant steganography, 2001.

[33]DES Encryption Tropical Software, http://www.tropsoft.com/strongenc/des.htm