Steganography Techniques: Concepts and Overview

Bhavneet Kaur[1], Pooja Nagpal[2], Harish Kundra[3]
[1]Research Scholar,[2]Assistant Professor, [3]HOD
[1, 2, 3] Department of Computer Science & Engineering, Rayat Institute of Engineering and Information Technology, Railmajra SBS Nagar, (Punjab), India
[1]bk.cgctc@gmail.com, [2]poojanagpal48@gmail.com, [3]hodcseit@rayatbahra.com

Abstract: **Steganography is the art of inconspicuously hiding data within data. Steganography's goal in general is to hide data well enough that unintended recipients do not suspect the steganographic medium of containing hidden data. The software and links mentioned in this article are just a sample of the steganography tools currently available. As privacy concerns continue to develop along with the digital communication domain, steganography will undoubtedly play a growing role in society. For this reason, it is important that we are aware of digital steganography technology and its implications. Equally important are the ethical concerns of using steganography and steganalysis. Steganography enhances rather than replaces encryption. Messages are not secure simply by virtue of being hidden.**

***Keywords*: Stego-object, Cover-object, Steganalysis, Cover Image, embedding key, extraction key, Steganography,**

I. INTRODUCTION

Steganography 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. It is a form of security through obscurity. The word steganography means "concealed writing" meaning "covered or protected writing". The advantage of steganography 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, steganography can be said to protect both messages and communicating parties. The art and science of hiding information by embedding messages within other, seemingly harmless messages. Steganography works by replacing bits of useless or unused data in regular computer files (such as graphics, sound, text etc) with bits of different, invisible information. This hidden information can be plain text, cipher text, or even images. Steganography sometimes is used when encryption is not permitted. Or, more commonly, steganography is used to supplement encryption. An encrypted file may still hide information using steganography, so even if the encrypted file is deciphered, the hidden message is not seen. Steganography (literally meaning covered writing*)* dates back to ancient Greece, where common practices consisted of writing messages in wooden tablets and covering them with wax, and tattooing a shaved messenger's head, letting his hair grow back, then shaving it again when he arrived at his contact point.

II. GENERAL CONCEPTS

In this section we go over the concepts and definitions used in the field of steganography. We first start by going over the framework in which steganography is usually presented and then go over some definitions. The modern formulation of steganography is often given in terms of the prisoner's problem where Alice and Bob are two inmates who wish to communicate in order to hatch an escape plan. However, all communication between them is examined by the warden, Wendy, who will put them in solitary contentment at the slightest suspicion of covert communication.

- Specially, in the general model for steganography, illustrated in Fig. below, we have Alice wishing to send a secret message m to Bob. In order to do so, she "embeds" m into a cover-object c, and obtains a stego-object s. The stego-object s is then sent through the public channel. Thus we have the following definitions:
- ***Cover-object****:* refers to the object used as the carrier to embed messages into. Many different objects have been employed to embed messages into for example images, audio, and video, and html pages to name a few.
- ***Stego-object****:* refers to the object which is carrying a hidden message. So given a cover object, and a messages the goal of the steganographer is to produce a stego object which would carry the message.
- In a pure steganography framework, the technique for embedding the message is unknown to Wendy and shared as a secret between Alice and Bob. However, it is generally considered that the algorithm in use is not secret but only the key used by the algorithm is kept as a secret between the two parties.
- The secret key, for example, can be a password used to seed a pseudo-random number generator to select pixel locations in an image cover-object for embedding the secret message (possibly encrypted).

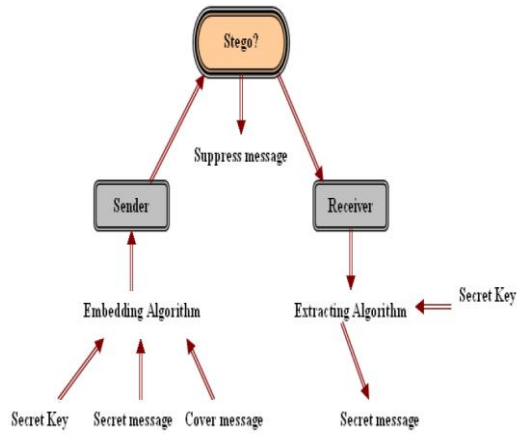

Figure1.

- Wendy has no knowledge about the secret key that Alice and Bob share, although she is aware of the algorithm that they could be employing for embedding messages between Alice and Bob can be passive or active. A passive warden simply examines the message and tries to determine if it potentially contains a hidden message. If it appears that it does, she suppresses the message and/or takes appropriate action, else The warden Wendy who is free to examine all messages exchanged be-she lets the message through without any action.

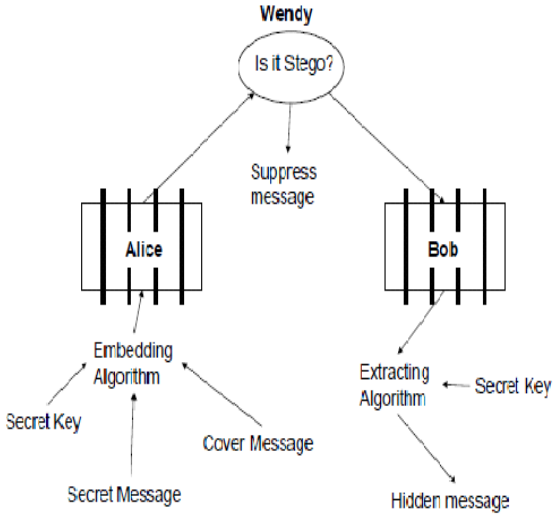

Figure2.

An active warden, on the other hand, can alter messages deliberately, even though she does not see any trace of a hidden message, in order to foil any secret communication that can nevertheless be occurring between Alice and Bob. The amount of change the warden is allowed to make depends on the model being used and the cover-objects being employed. For example, with images, it would make sense that the warden is allowed to make changes as long as she does not alter significantly the subjective visual quality of a suspected stego-image. In this tutorial we assume that no changes are made to the stego-object by the warden Wendy. Wendy should not be able to distinguish in any sense between cover-objects (objects not containing any secret message) and stego-objects (objects containing a secret message). In this context, steganalysis refers to the body of techniques that aid Wendy in distinguishing between cover-objects and stego-objects. It should be noted that Wendy has to make this distinction without any knowledge of the secret key which Alice and Bob may be sharing and sometimes even without any knowledge of the specific algorithm that they might be using for problem. However embedding the secret message. Hence steganalysis is inherently a difficult, it should also be noted that Wendy does not have to glean anything about the contents of the secret message m. Just determining the existence of a hidden message is enough. This fact makes her job a bit easier. The development of techniques for steganography and the wide-spread availability of tools for the same have led to an increased interest in steganalysis techniques. The last two years, for example, have seen many new and powerful steganalysis techniques reported in the literature. Many of such techniques are specific to different embedding methods and indeed have shown to be quite effective in this regard. We will review these techniques in the coming sections.

III. APPLICATIONS

A. Usage in modern printers:

Steganography is used by some 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.

B. Use by terrorists:

When one considers that messages could be encrypted steganographically in e-mail messages, particularly e-mail spam, the notion of junk e-mail takes on a whole new light. Coupled with the "chaffing and winnowing" technique, a sender could get messages out and cover their tracks all at once.

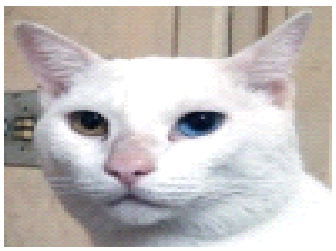

Figure3

An example showing how terrorists may use forum avatars to send hidden messages. This avatar contains the message "Boss said that we should blow up the bridge at midnight." encrypted with mozaiq using "växjö" as password.Rumors about terrorists using steganography started first in the daily newspaper USA Today on February 5, 2001 in two articles titled "Terrorist instructions hidden online" and "Terror groups hide behind Web encryption". In July the same year, an article was titled even more precisely: "Militants wire Web with links to jihad". A citation from the article: "Lately, al-Qaeda operatives have been sending hundreds of encrypted messages that have been hidden in files on digital photographs on the auction siteeBay.com". Other media worldwide cited these rumors many times, especially after the terrorist attack of 9/11. Immediate concerns also include the use of cyberspace for covert communications, particularly by terrorists but also by foreign intelligence services; espionage against sensitive but poorly defended data in government and industry systems; subversion by insiders, including vendors and contractors; criminal activity, primarily involving fraud and theft of financial or identity information, by hackers and organized crime groups..."

- "International interest in R&D for steganography technologies and their commercialization and application has exploded in recent years. These technologies pose a potential threat to national security. Because steganography secretly embeds additional, and nearly undetectable, information content in digital products, the potential for covert dissemination of malicious software, mobile code, or information is great."

- "The threat posed by steganography has been documented in numerous intelligence reports."

Moreover, an online "terrorist training manual", the "Technical Mujahid, a Training Manual for Jihadis" contained a section entitled "Covert Communications and Hiding Secrets Inside Images."

C. Alleged use by intelligence services:

In 2010, the Federal Bureau of Investigation revealed that the Russian foreign intelligence service uses customized steganography software for embedding encrypted text messages inside image files for certain communications with "illegal agents" (agents under non-diplomatic cover) stationed abroad.

IV. BASIC STEGANOGRAPHY SYSTEM

We can hide the data inside images using steganography technique. We can hide all the data that has been inputted within the image to protect the privacy of the data. The system provides an image platform for user to input image and a text box to insert texts. In this, user can send the stego image to other computer user so that the receiver is able to retrieve and read the data which is hidden in the stego image by using the same proposed system. Thus, the data can be protected without revealing the contents to other people. Steganography Imaging System (SIS) is a system that is capable of hiding the data inside the image. The system is using 2 layers of security in order to maintain data privacy. Data security is the practice of keeping data protected from corruption and unauthorized access. The focus behind data security is to ensure privacy while protecting personal or corporate data. Privacy, on the other hand, is the ability of an individual or group to seclude them or information about themselves and thereby reveal them selectively. Data privacy or information privacy is the relationship between collection and dissemination of data, technology, the public expectation of privacy, and the legal issues. Data privacy issues can arise from a wide range of sources such as healthcare records, criminal justice investigations and proceedings, financial institutions and transactions, biological traits, residence and geographic records and ethnicity. Data security or data privacy has become increasingly important as more and more systems are connected to the Internet. There are information privacy laws that cover the protection of data or information on private individuals from intentional or unintentional disclosure or misuse. Thus, hiding the data in a kind of form such as within an image is vital in order to make sure that security or privacy of the important data is protected.

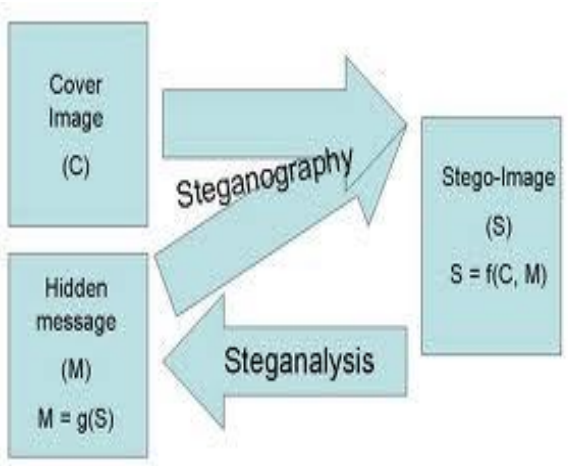

Figure4.

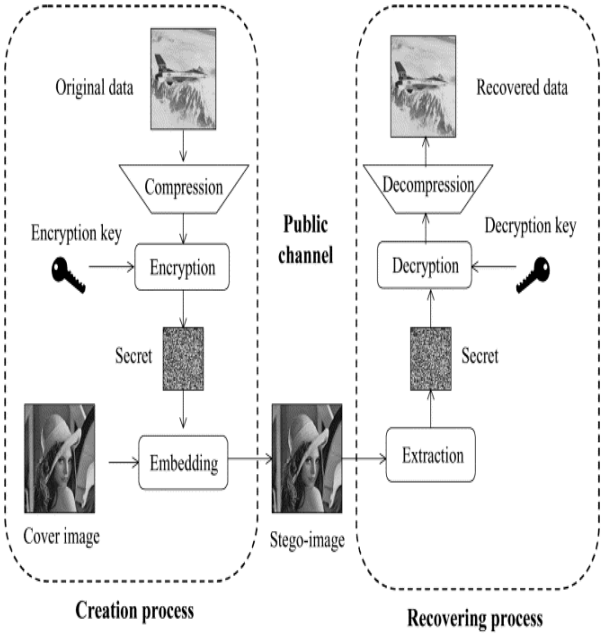

Figure5.

V. ADVANTAGES OF STEGANOGRAPHY

A. It does not attract the attention: Encrypting a message gives away that there is something of value and this will attract unwanted attention.

B. Packet sniffing barrier: Encrypted PGP email messages start with a line identifying them as an encrypted PGP message, making it easy for a packet sniffer on an ISP to flag encrypted PGP emails by just scanning for the word PGP or GnuPG, this can not be used against steganography.

C. Makes Internet surveillance difficult: If someone's Internet activities are being monitored visiting Flickr and uploading personal family photos with hidden messages will not trigger any alarm but sending encrypted messages and visiting a political discussion forum will.

D. Difficult to prove it exists: In some countries like the United Kingdom you can be required by the police to

provide the password to your encrypted files, refusing to do so carries a prison sentence, if the data has been hidden inside a photograph the police would first have to show beyond reasonable doubt that there is definitely something hidden inside the file.

VI. CHARACTERIZATION OF STEGANOGRAPHY SYATEMS

Steganographic techniques embed a message inside a cover. Various features characterize the strength and weaknesses of the methods. The relative importance of each feature depends on the application.

A. ***Capacity:*** The notion of capacity in data hiding indicates the total number of bits hidden and successfully recovered by the Stego system.

B. ***Robustness :*** Robustness refers to the ability of the embedded data to remain intact if the stego-system undergoes transformation, such as linear and non-linear filtering; addition of random noise; and scaling, rotation, and loose compression.

C. ***Undetectable :*** The embedded algorithm is undetectable if the image with the embedded message is consistent with a model of the source from which images are drawn. For example, if a Steganography method uses the noise component of digital images to embed a secret message, it should do so while not making statistical changes to the noise in the carrier. Undetectability is directly affected by the size of the secret message and the format of the content of the cover image.

D. ***Invisibility (Perceptual Transparency) :*** This concept is based on the properties of the human visual system or the human audio system. The embedded information is imperceptible if an average human subject is unable to distinguish between carriers that do contain hidden information and those that do not. It is important that the embedding occurs without a significant degradation or loss of perceptual quality of the cover.

E. **Security :** It is said that the embedded algorithm is secure if the embedded information is not subject to removal after being discovered by the attacker and it depends on the total information about the embedded algorithm and the secret key.

VII. TYPES OF STEGANOGRAPHY

Hiding data is the process of embedding information into digital content without causing perceptual degradation. In data hiding, three famous techniques can be used. They are watermarking, steganography and cryptography. Steganography is defined as covering writing in Greek. It includes any process that deals with data or information within other data. According to Lou et al., steganography is hiding the existence of a message by hiding information into various carriers. The major intent is to prevent the detection of hidden information. Research in steganography technique has been done back in ancient Greek where during that time the ancient Greek practice of tattooing a secret message on the shaved head of a messenger, and letting his hair grow back before sending him through enemy territory where the latency of this communications system was measured in months. The most famous method of traditional steganography technique around 440 B.C. is marking the document with invisible secret ink, like the juice of a lemon to hide information\ However, the majority of the development and use of computerized steganography only occurred in year 2000. The main advantage of steganography algorithm is because of its simple security mechanism. Because the steganographic message is integrated invisibly and covered inside other harmless sources, it is very difficult to detect the message without knowing the existence and the appropriate encoding scheme. There are several steganography techniques used for hiding data such as batch steganography, permutation stehanography, least significant bits (LSB), bit-plane complexity segmentation (BPCS) and chaos based spread spectrum image steganography (CSSIS).

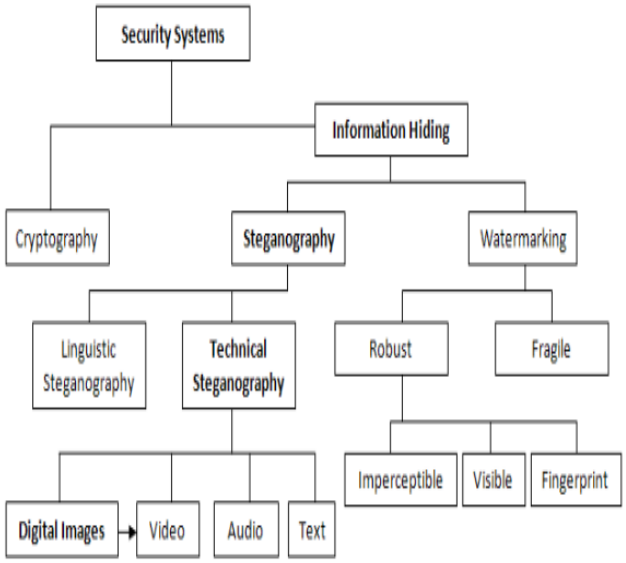

Figure6

A. Encoding the Secret Messages in Text

Encoding secret messages in text can be a very challenging task. This is because text files have a very small amount of redundant data to replace with a secret message. Another drawback is the ease of which text based Steganography can be altered by an unwanted parties by just changing the text itself or reformatting the text to some other form (from .TXT to .PDF, etc.). There are numerous methods by which to accomplish text based Steganography.

(a) Line-shift encoding involves actually shifting each line of text vertically up or down by as little as 3 centimeters. Depending on whether the line was up or down from the stationary line would equate to a value that would or could be encoded into a secret message.

(b) Word-shift encoding works in much the same way that line-shift encoding works, only we use the horizontal spaces between words to equate a value for the hidden message. This method of encoding is less visible than line-shift encoding but requires that the text format support variable spacing.

(c) Feature specific encoding involves encoding secret messages into formatted text by changing certain text attributes such as vertical/horizontal length of letters such as b, d, T, etc. This is by far the hardest text encoding method to intercept as each type of formatted text has a large amount of features that can be used for encoding the secret message. All three of these text based encoding methods require either the original file or the knowledge of the original files formatting to be able to decode the secret message.

B. Encoding the Secret Messages in Images

Coding secret messages in digital images is by far the most widely used of all methods in the digital world of today. This is because it can take advantage of the limited power of the human visual system (HVS). Almost any plain text, cipher text, image and any other media that can be encoded into a bit stream can be hidden in a digital image. With the continued growth of strong graphics power in computers and the research being put into image based Steganography, this field will continue to grow at a very rapid pace. Before dividng into coding techniques for digital images, a brief explanation of digital image architecture and digital image compression techniques should be explained. As Duncan Sellars explains "To a computer, an image is an array of numbers that represent light intensities at various points, or pixels. These pixels make up the images raster data." When dealing with digital images for use with Steganography, 8-bit and 24-bit per pixel image files are typical. Both have advantages and disadvantages, as we will explain below. 8-bit images are a great format to use because of their relatively small size. The drawback is that only 256 possible colors can be used which can be a potential problem during encoding. Usually a gray scale color palette is used when dealing with 8-bit images such as (.GIF) because its gradual change in color will be harder to detect after the image has been encoded with the secret message. 24-bit images offer much more flexibility when used for Steganography. The large numbers of colors (over 16 million) that can be used go well beyond the human visual system (HVS), which makes it very hard to detect once a secret message, has been encoded. The other benefit is that a much larger amount of hidden data can be encoded into a 24-bit digital image as opposed to an 8-bit digital image. The one major drawback to 24-bit digital images is their large size (usually in MB) makes them more suspect than the much smaller 8-bit digital images (usually in KB) when sent over an open system such as the Internet. Digital image compression is a good solution to large digital images such as the 24-bit images mentioned earlier. There are two types of compression used in digital images, lossy and lossless. Lossy compression such as (.JPEG) greatly reduces the size of a digital image by removing excess image data and calculating a close approximation of the original image. Lossy compression is usually used with 24-bit digital images to reduce its size, but it does carry one major drawback. Lossy compression techniques increase the possibility that uncompressed secret message will lose parts of its contents because of the fact that lossy compression removes what it sees as excess image data. Lossless compression techniques, as the name suggests, keeps the original digital image in tact without the chance of loss. It is for this reason that it is the compression technique of choice for steganographic uses. Examples of lossless compression techniques are (.GIF and .BMP). The only drawback to lossless image compression is that it doesn't do a very good job at compressing the size of the image data. We will now discuss a couple of the more popular digital image encoding techniques used today. They are least significant bit (LSB) encoding and masking and filtering techniques. Least significant bit (LSB) encoding is by far the most popular of the coding techniques used for digital images. By using the LSB of each byte (8 bits) in an image for a secret message, you can store 3 bits of data in each pixel for 24-bit images and 1 bit in each pixel for 8-bit images. As you can see, much more information can be stored in a 24-bit image file. Depending on the color palette used for the cover image (i.e., all gray), it is possible to take 2 LSB's from one byte without the human visual system (HVS) being able to tell the difference. The only problem with this technique is that it is very vulnerable to attacks such as image changes and formatting (i.e., changing from .GIF to .JPEG).Masking and filtering techniques for digital image encoding such as Digital Watermarking (i.e.- integrating a companies logo on there web content) are more popular with lossy compression techniques such as (.JPEG). This technique actually extends an images data by masking the secret data over the original data as opposed to hiding information inside of the data. Some experts argue that this is definitely a form of Information Hiding, but not technically Steganography. The beauty of Masking and Filtering techniques are that they are immune to image manipulation which makes there possible uses very robust. As a side note, there are many other techniques that are not covered in this paper that should be researched by anyone interested in using digital images for steganographic purposes. Techniques that use complex algorithms, image transformation techniques and image encryption techniques are still relatively new, but show promise to be more secure and robust ways to use digital images in Steganography.

C. Encoding Secret Messages in Audio

Encoding secret messages in audio is the most challenging technique to use when dealing with Steganography. This is because the human auditory system (HAS) has such a dynamic range that it can listen over. To put this in perspective, the (HAS) perceives over a range of power greater than one million to one and a range of frequencies greater than one thousand to one making it extremely hard to add or remove data from the original data structure. The only weakness in the (HAS) comes at trying to differentiate sounds (loud sounds drown outquiet sounds) and this is what must be exploited to encode secret messages in audio withoutbeing detected.There are two concepts to consider before choosing an encoding technique for audio. They are the digital format of the audio and the transmission medium of the audio.There are three main digital audio formats typically in use. They are Sample Quantization, Temporal Sampling Rate and Perceptual Sampling.Sample Quantization which is a 16-bit linear sampling architecture used by popular audio formats such as (.WAV and. AIFF). Temporal Sampling Rate uses selectable frequencies (in the KHz) to sample the audio. Generally, the higher the sampling rate is, the higher the usable data space gets. The last audio format is Perceptual Sampling. This

format changes the statistics of the audio drastically by encoding only the parts the listener perceives, thus maintaining the sound but changing the signal. This format is used by the most popular digital audio on the Internet today in ISO MPEG (MP3).Transmission medium (path the audio takes from sender to receiver) must also be considered when encoding secret messages in audio. W. Bender [8] introduces four possible transmission mediums:
1) Digital end to end - from machine to machine without modification.
2) Increased/decreased resampling - the sample rate is modified but remains digital.
3) Analog and resampled - signal is changed to analog and resampled at a different rate.
4) Over the air - signal is transmitted into radio frequencies and resampled from a microphone.
We will now look at three of the more popular encoding methods for hiding data inside of audio. They are low-bit encoding, phase-coding and spread spectrum.Low-bit encoding embeds secret data into the least significant bit (LSB) of the audio file. The channel capacity is 1KB per second per kilohertz (44 kbps for a 44 KHz sampled sequence). This method is easy to incorporate but is very susceptible to data loss due to channelnoise and resampling. Phase coding substitutes the phase of an initial audio segment with a reference phase that represents the hidden data. This can be thought of, as sort of an encryption for the audio signal by using what is known as Discrete Fourier Transform (DFT), which is nothing more than a transformation algorithm for the audio signal.Spread spectrum encodes the audio over almost the entire frequency spectrum. Itthen transmits the audio over different frequencies which will vary depending on what spread spectrum method is used. Direct Sequence Spread Spectrum (DSSS) is one such method that spreads the signal by multiplying the source signal by some pseudo random sequence known as a (CHIP). The sampling rate is then used as the chip rate for the audio signal communication. Spread spectrum encoding techniques are the most secure means by which to send hidden messages in audio, but it can introduce random noise to the audio thus creating the chance of data loss.There are many applications for Steganography, some good and some bad, which brings us to the closing section of our in-depth look at Steganography in which we will look at Steganalysis. Steganalysis is the art and science of stopping or detecting the use of all steganographic techniques mentioned earlier. In Steganalysis, the goal is to be able to compare the cover-object (cover message), the stego-object (the cover message with the hidden data embedded in it) and any possible portions of the stego-key (encryption method) in an effort to intercept, analyze and/or destroy the secret communication. As Fabien A.P. Petitcolas points out in his book, there are six general protocols used to attack the use of Steganography.
1) Stego only attack - only the stego object is available for analysis.
2) Known cover attack - the original cover object and the stego object are available for analysis.
3) Known message attack - the hidden message is available to compare with the stego-object.
4) Chosen stego attack - the stego tool (algorithm) and stego-object are available for analysis.
5) Chosen message attack - takes a chosen message and generates a stegoobject for future analysis.
6) Known stego attack - the stego tool (algorithm), the cover message and the stego-objects are available for analysis.
Being that Steganalysis is a broad topic and one that merits a paper on just it, I will close this discussion of Steganalysis by showing the reader one example of how someone could detect the use of steganographic tools that change the least significant bit (LSB) of an image in order to embed secret data inside of it.Generally, bitmap images (.BMP) have known and predictable characteristics. One such characteristic is the probability of near duplicate colors. Bitmap images get their colorfrom a central color table, which by its nature have little, or no near duplicate colors. When hidden data is embedded into the (LSB) of a bitmap image, it increases the number of near duplicate colors dramatically. Generally speaking, any bitmap image with more than fifty near duplicate colors should raise the suspicion of embedded data being present.

D. Encoding Secret Messages in Video

Video Steganography is a technique to hide any kind of files in any extension into a carrrying Video file.This project is the application developed to embed any kind of data(File) in another file, which is called carrier file. The carrier file must be a video file. It is concerned with embedding information in an innocuous cover media in a secure and robust manner. This system makes the Files more secure by using the concepts Steganography and Cryptography.

VIII. CONCLUSIONS

In this paper, we have discussed the various ways in which steganography can be used and its application areas and advantages in the present scenario. In the next paper, we shall propose a new steganography system which gives us more security and higher value of PSNR.

REFERENCES

[1]SANS Security Essentials, (volume 1.4, chapter 4) Encryption and Exploits, 2001.2) Petitcolas, Fabien A.P., “Information Hiding: Techniques for Steganography and Digital Watermarking.”, 2000.

[2]StegoArchive,“Steganography Info, Software and News to enhance your Privacy”, 2001.

[3]Petitcolas, Fabien A.P., “The Information Hiding Homepage: Digital Watermarking and Steganography”,

[4] Johnson, Neil F., “Steganography”, 2000, 6) The WEPIN Store, “Steganography (Hidden Writing)”, 1995.

[5] Sellars, D., “An Introduction to Steganography”

[6] Bender, W., “Techniques for Data Hiding”, IBM Systems Journal, Vol. 35, Nos 3+4, Pgs 313-336, 1996.

[7] Krinn, J., “Introduction to Steganography”, 2000.