Comprehensive Review of Various Types of Steganographic Implementations

Gunit Malik[1], Siddhant Modi [2]
Department of Computer Engineering
[1,2] Student, Mukesh Patel School of Technology Management & Engineering,

Abstract- *Since the rise of the Internet, security has been one of the most important factors of information technology and communication. Cryptography was created as a technique to secure the secrecy of communication by encrypting and decrypting data to keep the message secret. Unfortunately, this is not enough to keep the contents of the message a secret. It is sometimes important to hide the existence of the message itself so as to keep the message secret. The technique implemented to achieve this is called steganography.*

Steganography, coming from the Greek words stegos, meaning roof or covered and graphia which means writing, is the art and science of hiding the fact that communication is taking place. Using steganography, a secret message can be embedded inside an unsuspicious piece of information. The purpose of this is covert communication- to hide the information or even the existence of the message from a third party. Using steganography, information can be hidden in different embedding mediums, known as carriers. These carriers can be images, audio files, video files, and text files. This paper emphasizes on the methods used for applying steganography on different types of media, focusing on image files, audio files, text files and on the web. The main focus is on explaining various types of steganographic techniques and the types of media it's used in, which is described in other research work. This provides a concise summary for readers.

Keywords- Steganography, Cryptography, Data hiding, Image, Audio, Web, Text.

I. INTRODUCTION

There are multiple techniques to implement secure communication using Steganography. The other ways to secure data such as cryptography and watermarking are also very effective but each has their drawbacks. Cryptography focuses on encrypting the data but if an unofficial user acquires the key then it can easily be hacked. Watermarking is similar to Steganography as they both hide the data over a media but in watermarking the data is related to the media whereas in Steganography there is no relation to the data.

Although the term steganography was only coined at the end of the 15th century, the use of steganography dates back several millennia. In ancient times, messages were hidden on the back of wax writing tables, written on the stomachs of rabbits, or tattooed on the scalp of slaves. Invisible ink has been in use for centuries-for fun by children and students and for serious espionage by spies and terrorists [1].

The steganography process generally involves placing a hidden message inside an appropriate carrier (transport medium) using a steganography key. This is then sent to the receiver, who using the key can retrieve the original message.

An effective steganographic scheme should possess the following desired characteristics:

- Secrecy: No one should be able to extract message from the host medium without the knowledge of the proper secret key used in the extraction procedure.
- Imperceptibility: The stego file should be indistinguishable from the original cover file. It should be impossible for one to notice there is a hidden message.
- High capacity: The stego file should have a high capacity for the secret message without affecting the quality of the message.
- Accurate extraction: When the secret message extracted at intended receiver side the secret message should be accurate and without any distortion.

In the following paper steganography is divided into the following categories namely

- Image steganography
- Web steganography
- Text steganography
- Audio Steganography

Peak Signal to Noise Ratio (PSNR) and Mean Square Error (MSE) are used to measure the quality of the stego-file, which is its imperceptibility. They are inversely proportional to each other.

Entropy is used to measure the level of security of encrypted information.

II. STEGANOGRAPHY TERMINOLOGIES

There are a few terminologies used in the field of steganography and which are essential to be understood to understand steganography. They have been mentioned in Fig.1. as well and are explained below.

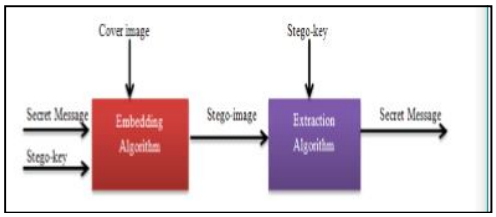

Fig.1. The Basic model of steganography [2]

Cover Object: In Steganography cover objects are media in which the secret data is hidden. The cover object has to be appropriately chosen according to the type of message. In Image Steganography it is an image, in Audio Steganography is a digitized audio file and in Text Steganography it is text file.

Secret Message: In Steganography, the secret message is the message in transmission hidden in the cover object. The secret message format depends on the type of steganographic process.

Stego Object: The stego object is the final product after hiding the secret message in cover message. The stego object is used for transmission and then at receiver side processing is done on stego object to extract message from it.

III. LITERATURE SURVEY

Many researchers have thoroughly studied Steganography, the benefits, drawbacks and most importantly the implementation of it .The following paper gives a sufficient review and insight based on the study of these papers related to our work.

3.1- Image Steganography

Images are the most popular cover objects used for steganography. In the domain of digital images, there exist various image file formats. For different file formats, there exist different steganographic algorithms. The most common algorithm used for image steganography is Least Significant Bit (LSB) algorithm [3]. Images consist of a rectangular map of the image's pixels (represented as bits) where each pixel is located and its colour. These pixels are displayed horizontally row by row. The number of bits in a colour scheme, called the bit depth, refers to the number of bits used for each pixel. Depending on the type of image, the bit depth can range from 8 bit (Gray scale images), up to 24 bits (RGB images). The image size increases with bit depth.

The LSB algorithm is the simplest to implement. It converts the secret message to binary and starts replacing the last bits of the cover image pixels, known as the least significant bits with the bits of the secret message. In 8 bit Gray Scale images, the 8th bit is replaced. When using a 24-bit colour image, a bit of each of the red, green and blue colour components can be used, so a total of 3 bits can be stored in each pixel. Hence the bit count is more when using 24 bit colour images. For example [3], the following is the representation of a single pixel of a 24-bit colour image:

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

When the character A, which binary value equals 10000001, is inserted, the following grid results:

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

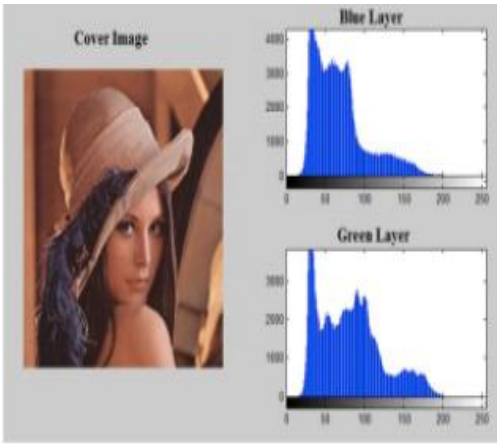

Fig.2. Cover Image "Lena" and histograms [2]

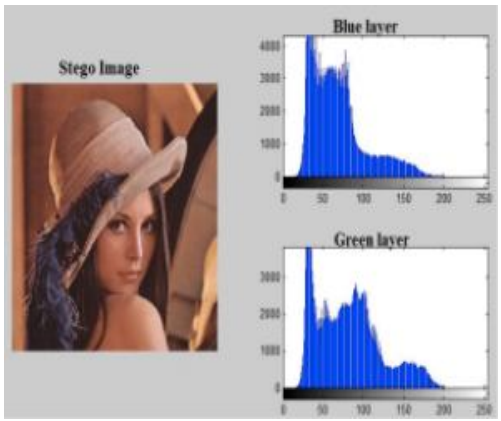

Fig.3. Stego Image "Lena" and histograms [2]

Another method proposed a comparatively efficient LSB technique, which increases imperceptibility and hiding capacity of the stego image. [7]

Images have a lot of redundancy because adjacent pixels are similar, so compression is relatively easy. If secret data is a digital image, then compression leads to better hiding capacity and imperceptibility.

The proposed method uses all the cover image pixels for data hiding which leads to less difference between stego and cover image. This also led to better quality of picture
The crux of the proposed method is to break the secret data into two segments and then using LSB method to embed the segments into two cover pixels of one.

PSNR and MSE are calculated according to following formula where Si and Ci are ith pixel of secret and cover pixel:

$$PSNR=10\times\log\left(255^2/MSE\right)\ dB$$
$$MSE=\frac{1}{w\times h}\sum_{i=1}^{w\times h}(S_i-C_i)^2 \quad (1)$$

In the proposed method the secret data is broken into remainder and quotient after being divided by m (Gray scale Value), which is then hidden separately over the cover pixels.

The secret data is divided which gives quotient QI either 0 or 1 and remainder UI. Each secret data is now a quotient QI, remainder UI pair. A remainder sequence R is formed.

The adjacent quotient bits are grouped together, which form quotient sequence Q. The secret data value has been encoded into two sequences: Remainder sequence R and grouped quotient sequence Q.

Which are then embedded into the cover pixels using LSB method and transmitted to the targeted receiver, the remainder and quotient sequences found separately and combined to get secret data. The drawback of this method is that it is not effective if there is less repetition in the secret data.

3.2- Audio Steganography

Audio encoding involves converting an analog signal to a bit stream.[5] Analog sound-voice and music are represented by sine waves of different frequencies. Sound, being an analog signal is continuous. Storing it digitally requires the continuous sound wave to be sampled. After sampling, a sequence of bits (zeros and ones)is obtained. These samples obtained are converted into voltage levels. Then, pulse code modulation (PCM) is used is used to convert the obtained samples into a numeric value. The device used to perform this function is called coder-decoder or codec. After the bits are obtained, the method uses LSB substitution as discussed above to achieve the process of steganography.

After the embedding is finished, the code is converted back to the analog signal and the stego file is ready to be transmitted.

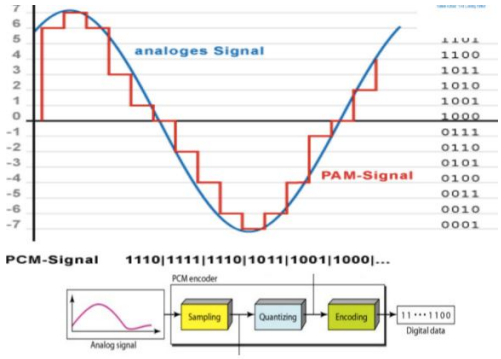

Fig.4. Pulse Code Modulation

Another method for the implementation of audio Steganography Genetic algorithm provides a significant advantage over more commonly used methods [6]. **Genetic algorithm** (**GA**) is inspired by the process of natural selection, which means the survival of the fittest. Genetic algorithms are commonly used to generate high-quality solutions for optimization, which in Steganography means the difference between cover file and stego file (Imperceptibility).

In genetic algorithm set of chromosomes are taken, but here instead of taking chromosomes we are taking the strings of binary data and then the binary data of best fitness value. It is stated that Human Auditory System is very sensitive compared to the Human Visual system therefore GA is for optimization. Due to the lack of robustness in LSB method some researchers theoretically proposed the concept of GA

A researcher introduced the concept of hybrid genetic algorithm. In this application of GA it is used to the optimal position for the Least Significant Bit of the audio file to embed the secret message in. The audio file is divided into parts to embed the secret message, which is in a text or audio format. When the audio file is divided into parts, each part represents

a chromosome. The length of each chromosome represents the length of the message.

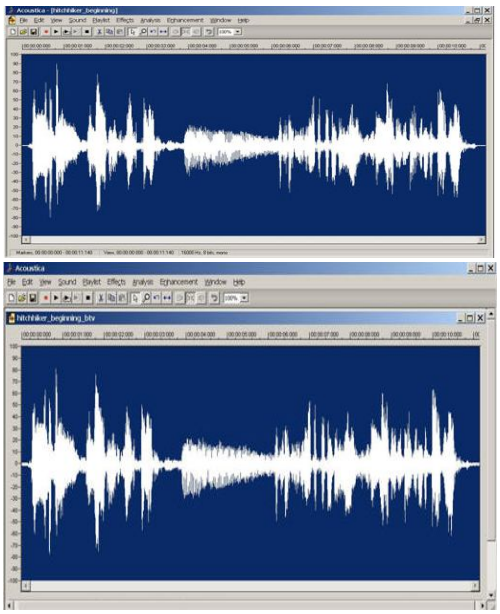

Fig.5-6. Comparison between 2 wav files before and after embedding data [1]

3.3- Web Steganography

This research paper gives another skyline to safe correspondence through information hiding on Internet.
The techniques described in the paper hide the secret data in the source code of Web pages. This is called Web Steganography. The following techniques discussed are all based on tags [8].

1. White spaces in tags

In this method white spaces are either added or removed at the beginning or the end of the HTML tag definition (After the bracket open or before bracket close). One bit of data is embedded per tag.

2. Appearing order of the elements

In this method, the secret message is embedded by exchanging the order of elements in tags. Each pair of interchanged element represents one bit of data.

3. Change case of Letters in tags

As HTML tags are case insensitive, this method used this hide a message within a document by changing the case of in a tag's name without affecting the meaning. It has a large capacity but is easily detectable compared to other methods.

4. Appearing order of the attributes

In this method the secret data can is hidden by interchanging the order of attributes in the element. One bit of data is covered per by interchanging order of attributes.

2. Change quotation marks of attribute values in tags

Attribute values are enclosed with single inverted comma, double inverted comma or without commas. It does not affect the output of the HTML page.
After examining and analyzing the methods described the researcher derived the following method, which improves on each of these methods.
The derived algorithm is divided into two parts

A. Embedding process

This embeds the secret text into the webpage. It takes the original message and encrypts the message using cryptography using key k. The encrypted is converted into ASCII and then in binary code. A webpage is selected as the cover file and any of the 4 methods discussed above are used to embed the binary code into the webpage. This webpage is made live and then the targeted receiver can extract the encrypted message

B. Extracting Process

The webpage is first opened and the binary code is extracted then converted to ASCII code. Original message will be obtained by decrypting the text using key k.
The following table shows the comparison between the proposed methods and the existing methods [8]

Techniques	Imperceptibility	Change in file size	Security	LEC
Change case of letters in tags	Weak	No	Weak	100 %
By using white space	Good	Yes (minor)	Yes	100 %
Appearing order of the attributes	Good	No	Strong	5%
Change of quotation marks in attribute values of tags	Medium	Yes (minor)	medium	80%
Proposed Method	Very Good	Yes (minor)	Very Strong	285 %

Fig.7. Comparison between proposed and existing methods. Large embedding capacity (LEC)

3.4-Text Steganography

In the reviewed paper the researcher used a combination of cryptography and steganography [4].

The researcher used Variable block size data encryption algorithm for encrypting the secret text data and then following two Steganography methods:
1) Modified LSB Technique: In this method the bits of the secret text data are embedded to 2 or 3 Least Significant Bits of the cover image. This benefits the hiding capacity over normal LSB technique but at the price of increased variability.
2) Raster Scan Technique: This technique is inspired from television, the secret text data bits are embedded into the cover image from left to right followed by right to left or it can even take another pattern top to bottom or bottom to top for data hiding.

In the proposed method the R, G, B planes of RGB image are extracted from the cover image and the cipher text bits are embedded into each planes by always XORing the least significant bit using both modified and raster scan technique of Steganography. This process is continued till the whole cipher text will be hidden in cover image.

The researcher then combined R, G, B planes to obtain the cover image with the secret text data. The researcher recorded PSNR after transmission. The results are depicted in the table below [4]

Y-axis: PSNR in DBs
X-axis: Image Name

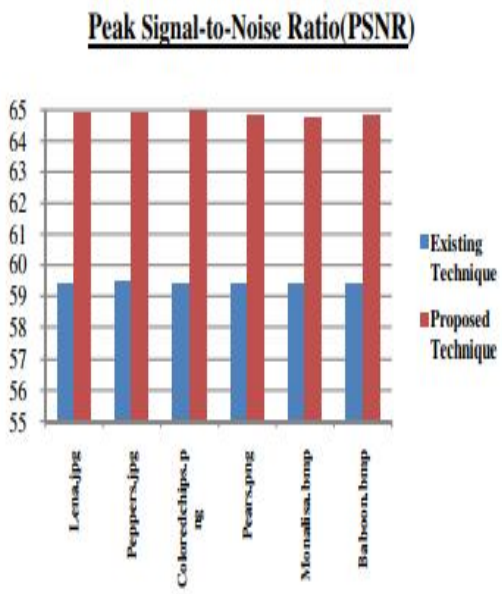

Fig.8. Comparison between the PSNR of the existing and proposed technique

The researcher used extra metric Entropy that was not used in existing methods. The proposed algorithm as compared to existing algorithm gives result less MSE and more PSNR value hence better quality of imperceptibility.

Inference

Overall, steganography is an extremely viable way to provide security to any type of file. Especially now as modern-day techniques which are becoming extremely common and exposed to unauthorized people this makes steganography have considerable advantages. For files and sensitive content which are relatively small in size, steganography is the most effective which is showcased in the table below[9]

Criteria	Encryption	Watermarking	Steganography
History	Modern process	Modern process	Extremely ancient with modern implementations in digital era
Fails when	Decrypted during transmission	Removed or Replaced	File detected as stego-file
Objective	Data security	Copyright preservation	Secret transfer
Visibility	Visible	Possible	Never
Detection	Blind	By cross-checking	Blind
Input files	One	Minimum 2 (unless self-embedding)	Minimum 2 (unless self-embedding)

Fig 9. Encryption vs Watermarking vs Steganography

In the described methods of image steganography, both the methods have a drawback of high computational power due to compression. These techniques give a better PSNR, which enhances stego image qualities.

In audio steganography the sampling of the analog signal does not give accurate conversions to digital and there exists a quantization error. Genetic Algorithm is still highly experimental and there is no wide scale implementation due to lack of research and high complexity even though it gives a significant advantage over existing methods. The process described for web steganography has advantages, which outweigh its drawbacks such as increasing the web page size exponentially.

IV. CONCLUSION

This paper provides a framework on the different types and methods of steganography and various ways to implement it. Steganography has been researched a lot but during the review it was evident that it has significant

advantages over watermarking and cryptography therefore there needs to be a higher understanding and intense research on less common topics like Genetic Algorithm and steganography as a whole. The goal of secure internet and communication can be achieved by steganography in combination with cryptography and watermarking.

REFERENCES

[1] Gary C. Kesler, "An overview of Steganography for the Computer Forensics Examiner", February 2015
[2] Marwa M. Emam, Abdelmgeid A. Aly, Fatma A.Omara,"An Improved Image Steganography Method Based on LSB Technique with Random Pixel Selection", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 3, 2016
[3] J.R. Krenn, "Steganography and Steganalysis", January 2014
[4] Shivani Chauhan, Jyotsna, Janmejai Kumar and Amit Doegar,"Multiple layer Text security using Variable block size Cryptography and Image Steganography", 3rd IEEE International Conference on "Computational Intelligence and Communication Technology",9-10 Feb. 2017,page 1-8
[5] Sangeeta Roy, Avinash Kumar Singh, Jyotirmayee Parida, Ashok Singh Sairam, "Audio Steganography Using LSB Encoding Technique with Increased Capacity and Bit Error Rate Optimization", October 2012
[6] Prashant Johri,Arun Kumar,Amba, "Review paper on text and audio steganography using GA", International Conference on Computing, Communication & Automation,15-16 May 2015, page 1-6
[7] Nadeem Akhtar, Vasim Ahamad, Hira Javed, "A Compressed LSB Steganography Method", 3rd IEEE International Conference on "Computational Intelligence and Communication Technology",9-10 Feb. 2017 page 1-8
[8] Lipi Kothari,Rikin Thakkar,Satvik Khara,"Data hiding on web using combination of Steganography and Cryptography", International Conference on Computing, Communication & Automation,1-2 July 2017, Page 1-5
[9] Namrata Singh,"Survey Paper on Steganography",International Refereed Journal of Engineering and Science, Volume 6, Issue 1 (January 2017).