STEGANOHIDE: AN INFORMATION HIDING SYSTEM USING STEGANOGRAPHY TECHNIQUE

[1]Shobhit Sharma, [2]Shivam Sharma and [3]A. Charan Kumari

[1]Student, Department of Electrical Engineering, Dayalbagh Educational Institute, Dayalbagh, Agra, India;

[2]Student, Department of Electrical Engineering, Dayalbagh Educational Institute, Dayalbagh, Agra, India;

[3]Assistant Professor, Department of Electrical Engineering, Dayalbagh Educational Institute, Dayalbagh, Agra, India

Email: { [1]shobhitsharma686@gmail.com, [2]shivamsharma156506@gmail.com, [3]charankumari@dei.ac.in}

Abstract

Steganography, a subset of data hiding techniques, seeks to conceal sensitive information within any form of digital media. The goal of a steganography plan at the sender's end is to incorporate a secret message into an innocent-looking image or any other file format known as a cover object. Meanwhile at the receiver's side its goal is to extract the hidden information or message. Steganalysis, on the other hand, is the inverse process of steganography and attempts to detect steganography. The role and development of steganalysis has increased when steganography was used for vulnerable activities like the embedding of harmful content. Moreover, the development of effective steganographic methods has made it very difficult to decipher them. This paper discusses in depth about the various existing steganalysis techniques and suggests improvements that can be brought about to prevent the abuse of steganography. It includes implementation of an encryption and decryption algorithm which evades usage of a secret key in the process. An interface that facilitates the same has also been developed and discussed in detail. It also discusses the further scope and challenges in both steganography and steganalysis.

Key Words - steganalysis, steganography, stego object, data security, LSB encoding

1. Introduction

Data and information have become extremely important in the today's digital world. The problem of storing, sending, and sharing this private information over an unsecured communication route is still open [1]. Researchers have shown a great deal of interest in data protection methods like cryptography, watermarking, and steganography in this regard. Steganography in particular, of various types, has sparked a lot of interest. The skill of finding and extracting concealed data from digital media files, also known as steganalysis, has been created to combat this digital dishonesty. Steganalysis is a procedure that concentrates on figuring out whether digital files contain hidden data and where the concealed data is kept inside the file. Now if we talk about detection of steganography i.e., steganalysis, over the years many methods and techniques have been developed to be at par with steganography [2]. The aim of steganalysis can vary depending upon the usage. Steganalysis can be used for detecting the steganographic communication, to break the channel of the communication or to also decode the message that was being sent secretly [3]. Based on these, algorithms are constructed. Various scenarios are possible in steganalysis depending upon the goals and approaches. The three main types of scenarios are termed active, passive and malicious [4]. Passive steganalysis involves observing and analyzing conversation without interfering in order to find hidden messages. In order to prevent steganography, the transmission is purposefully interfered with during active steganalysis. In malicious steganalysis, the communication is hacked by pretending to be one of the steganography scheme's participants.

The remainder of this paper is structured as follows: Section 2 provides a brief description of the types of steganography. Section 3 discusses about the various techniques that have been developed in steganography as of date. Section 4 contains various techniques of steganalysis. Section 5 contains the implementation details of the proposed algorithm. Section 6 presents the implementation of the algorithm. Section 7 finally concludes by describing the final inferences and future work.

2. Types of Steganography

The following section describes the various types of steganography based on the format of the files. Steganography is performed on many different formats of digital data storage. Image steganography is the most popular of them all. The various formats on which it is

performed are-

- Text Steganography- Information is being concealed by text steganography in the text files. To create readable texts, it may be necessary to alter the structure of already existing text, change the words within a text, create random character sequences, or use context-free grammars. Several methods, including the Format Based Method, Random and Statistical Generation, and Linguistic Method, are used to conceal the data in the text.

- Image Steganography- Image steganography is the practice of concealing data by using the cover object as the picture. Images are frequently used as a cover source in digital steganography because they contain a large number of bits in their digital form. There are many techniques for concealing information within an image- Least Significant Bit Insertion, Masking and Filtering, Redundant Pattern Encoding, Encrypt and Scatter and Coding and Cosine Transformations [5]. Image steganography and its analysis will be discussed at length in this paper.

- Audio Steganography- The hidden message is inserted into an audio signal during audio steganography, changing the binary order of the associated audio file. If we compare, then, image steganography is much easier to use than digital audio steganography for the purpose of concealing secret communications. Least Significant Bit Encoding, Parity Encoding, Phase Coding, and Spread Spectrum are a few different ways to encrypt audio files.

- Video Steganography – Various types of data can be concealed in digital video formats using video steganography. The benefit of this type is that a significant quantity of data can be concealed inside and that it is a moving stream of sounds and images. This can be compared to an amalgamation of audio and image steganography. There are two primary types of video steganography: embedding data directly into the compressed data stream and embedding data in raw, uncompressed video.

- Network Steganography- It is a method of incorporating data into network management protocols used for data transmission, including TCP, UDP, ICMP, and others. Steganography can be applied in the covert channels in the OSI model. For instance, one can obfuscate data in some optional or header sections of a TCP/IP packet.

Although all formats of steganography find implementation in various fields, the paper is focused only on image steganography, which is the hidden integration of information into digital images. Internet's widespread use of digital pictures makes them the most common application for steganography [6]. Large amounts of material can be covered by the image file's enormous size. The Human Visual System finds it difficult to distinguish between a normal picture and an image with hidden information. Digital images frequently have a lot of unnecessary bits, which is why they are the most widely used cover items for steganography. In order to use steganography, any project employs an image as a cover file. Different image formats, such as JPEG, BMP, TIFF, PNG, or GIF files, can be used.

The very basic working of image steganography concerns concealing a secret message in an image which is exposed, called the cover image. The cover image along with the message is called as stego object. This stego object is transmitted across various channels to the receivers to whom it is supposed to be delivered. A special secret key is introduced in many algorithms to decode the message. With or without using the key the receiver is able to decrypt the message. This is how, broadly talking, an image steganography exchange takes place.

There are three architectural designs that concern steganography- cover selection, cover synthesis, and cover modification. By selecting a cover image with a hidden meaning, the sender can convey a secret message using steganography by cover selection. The sender creates his own cover image to convey the secret message in steganography by cover synthesis. Steganography by cover modification involves the sender making changes to a cover image to conceal the hidden message.

3. Techniques of Steganography

The following section highlights the various existing steganography techniques. Several techniques have been deployed to achieve steganography. Some the broad categories are described below-

- Spatial Domain Techniques – In spatial domain techniques the data concealing is carried out directly on the pixel value of the cover image in order to make the message invisible on the cover image. The following categories are used to group spatial domain methods:
  - LSB Encoding- One of the methods in the spatial domain is called LSB. Although it is straightforward, LSB is vulnerable to lossy compression and picture manipulation. In order to hide the data, some bits are changed immediately in the image's pixel values [7]. For the human eye, LSB value changes are undetectable.
  - Pixel Value Differencing- The first two successive pixels are chosen to embed the data in PVD. Calculating the disparity between two regular pixels yields the payload.

- BPC- The noise component of image complexity is measured using the Binary Pattern complexity method. Binary Pattern, which is mapped from the secret data, is used to replace the noisy part. When the reverse noise component is calculated, the image won't change.

- Transform Domain Techniques- It is a more complex method of concealing data in a picture. To conceal information in the pictures, various algorithms and transformations are used. When a signal is embedded in the frequency domain, it does so much more effectively than when it is embedded in the time domain. Techniques in the transform domain conceal information in pictures that are less subject to compression, image processing, and cropping than do techniques in the spatial domain [8]. Some transform domain methods perform both lossless and lossy format conversions without regard to the image format. Various categories of transform domain methods, including the discrete Fourier transformation (DFT), the discrete cosine transformation (DCT), and the discrete wavelet transformation (DWT)
  - Discrete Fourier Transform (DFT)- Signals with discrete time are transformed into harmonics with discrete number. A finite combination of complex sinusoids ordered by their frequencies is transformed by DFT from a finite list of evenly spread function samples into a list of coefficients. The sampled function is said to be converted from its initial domain, which is frequently time or position along a line, to the frequency domain. The Fourier transform of Discrete Time makes use of discrete time but changes it to continuous frequency. On contemporary computers, the algorithm used to calculate the DFT is extremely quick. This method, also known as Fast Fourier Transform or FFT, uses the Inverse Discrete Fourier Transform to achieve the same outcome as the Discrete Fourier Transform (DFT).
  - Discrete Cosine Transform (DCT)- The Discrete Fourier Transform and this technique are comparable. The signal or picture is transformed using DCT from the spatial domain to the frequency domain. The effect of spreading the location of the pixel values over a portion of the picture is created by the mathematical transformations that are applied to the pixels [9]. The inverse discrete cosine transforms, or IDCT, is used in the decompression procedure to reconstruct an image when necessary.
  - Discrete Wavelet Transform (DFT)- The picture is changed from the spatial domain to the frequency domain using it. DWT determines the high frequency and low frequency information of each image pixel during the steganography procedure. It is a mathematical instrument for hierarchically decomposing images [10]. Non-stationary signal processing is its primary application. DWT operates in the two-dimensional plane and in one degree. The DWT is a multi-resolution image description and is a more precise model than the DFT or the DCT.

- Spread Spectrum - This method makes use of the spread spectrum idea. The secret data is dispersed over a large frequency bandwidth in this technique. Every frequency band's signal-to-noise ratio must be so low that it is challenging to identify the presence of data. There would still be enough information available in other bands to recover the data, even if portions of the data were removed from a few bands. As a result, it is challenging to fully remove the data without also completely destroying the cover. It is an extremely effective strategy used in combat communication.
- Statistical Technique- Using the method, the message is embedded by altering a number of the cover's properties. One communication bit is embedded in each block after the cover is divided into segments. Only when the size of the message bit is one does the cover block need to be changed; otherwise, there is no need.
- Distortion Technique- By distorting the signal, the distortion technique is used to store the secret data. A series of modifications are made to the cover picture by an encoder, and a secret key is used in the decoder phase to translate the encrypted data back into the original data.

4. Types of Steganalysis

This section focuses on the existing steganalysis techniques. There are specific kinds of classification algorithms that can be found in the literature; the classification is a supervised process that requires prior training to divide the data into normal and stego data. The steganographic algorithm (SA) might or might not be necessary for the steganalysis method. Hence steganalysis is categorized as Specific and Generic [11]. Few algorithms are steganographic algorithms dependent, and few are not. Basically, they fall into two categories:

- Specific Steganalysis- The steganographic algorithm (SA) is well-known, and the steganalysis method is designed using SA. The SA is necessary for the steganalysis method. This kind of steganalysis examines how an image's statistical characteristics change after insertion. The use of specific steganalysis has the benefit of producing extremely accurate findings. The specific or target embedding method is the only real option for steganalysis. As a result, not all kinds of

algorithms can use it completely. Additionally, not all image types are supported.

- Blind Steganalysis- The steganalysis algorithm is not universally accepted in the approach. Therefore, anyone can create a detector to look for the hidden information without using steganalysis algorithms. In contrast to specialized steganalysis, universal is more widespread and less effective [12]. However, because it is independent of the SA, universal steganalysis is more frequently used than specialized one. The subject of this study is general steganalysis. The next two stages are feature extraction from the data and grouping the results into two categories.
  - Feature Extraction- It is a method for developing a collection of unique statistical characteristics for a picture. These qualities are referred to as features. A dimension decrease is all that feature extraction is. Wavelet decompositions, moment of image statistic histograms, empirical transition matrices, moment of image statistic from spatial and frequency domain, and co-occurrence matrices are some of the feature extraction techniques. The extracted features must be sensitive to the embedding objects and the Image quality metrics.
  - Classification- It is a method of classifying the pictures into groups based on the values of their feature attributes. One of the main divisions of steganalysis is supervised learning. Learning that is partially supervised is permitted. In this learning, the classifier is trained using a collection of training inputs that includes input features. Based on the provided features, a class label is determined following training.

5. Implementation

This section details on the implementation of the proposed technique. For development of a steganalysis scheme or algorithm, various scenarios can be considered. Some such scenarios are-

- for analysis, only the stego object is accessible.
- analysis of both the cover and the stego object is possible.
- knowing the message and comparing it to the stego object
- analysis of the stego entity and stego algorithm is possible.
- the original object and the stego-object are both accessible, and the steganography tool algorithm is known.

Here in, we take the last scenario and follow a simple algorithm to encode and decode the message contained within. The proposed algorithm is a modified version of Least Significant Bit (LSB) method.

The encryption algorithm is as follows-

1. First the cover image and the message to be sent are obtained.
2. The message string is converted to binary string.
3. The image is taken as input in RGB format and converted to a bit array.
4. The conversion of RGB image to bits and its inverse are facilitated using hex codes.
5. After converting to binary, we substitute the LSB bit of the image with the binary values of the secret message.
6. A binary breakpoint is added to identify the message length.
7. With this the stego object with hidden secret message is created.

Fig. 1 depicts the encoding algorithm-

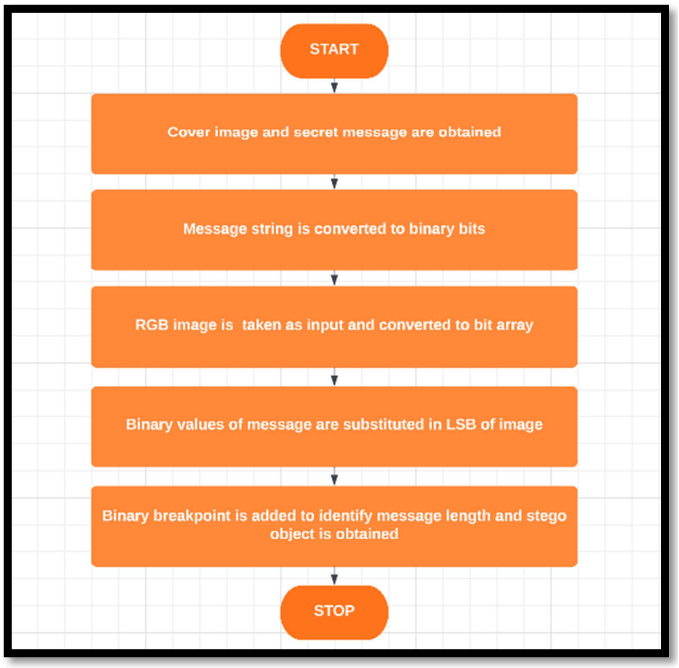

Fig. 1 Encoding Algorithm

The decryption algorithm follows as-

1. Firstly, the stego image is obtained.
2. The image is opened in the desired RGB format.
3. Now, extraction is performed on the LSB bits, with the breakpoint indicating the length
4. The extracted bits are reconstructed from binary to string format and message is hence decoded.

Fig. 2 depicts the decoding algorithm-

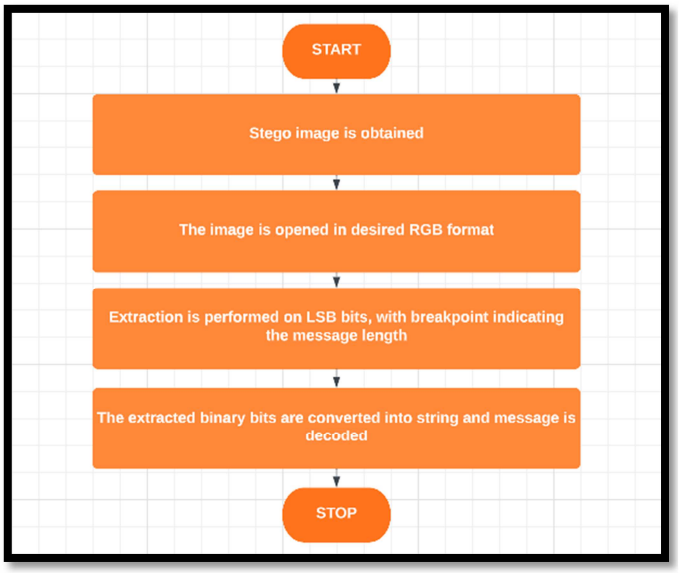

Fig. 2 Decoding Algorithm

The advantage of using this algorithm for the encryption and decryption is that there is no need to supply any secret key at the time of encryption. Also using hex codes as a means for conversion makes it more efficient.

5.1 Technology Stack

The algorithm can be implemented easily in any programming language. A simple functional-programming based program is designed. For implementation using Python, various Python modules are used to write a program based on this algorithm which include pillow, optparse, binascii, codecs, bitstring. Pillow is used for to write in and out the image, bitstring, codecs, binascii are used for manipulation of data and optparse is used for creating the helper function to execute other functions. For the development of the web application, Django was used as the backend engine and Angular.js was used on the frontend.

6. Results

The following section contains the steps and screenshots of the execution of the web application based on the above-mentioned algorithm. It consists details of all the components present in the web application.

Step 1: The first page of the application is as depicted in fig. 3.

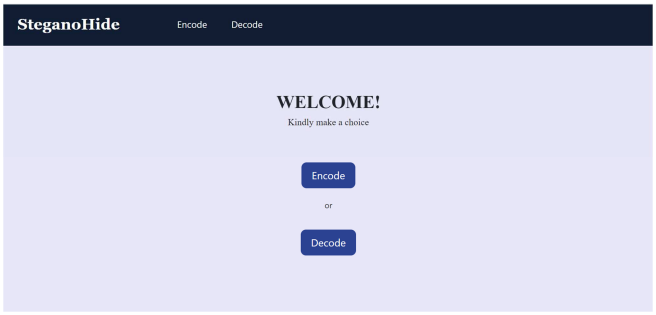

Fig. 3 Landing Page

Step 2: For encoding, select encode option from the menu bar as depicted in fig. 4.

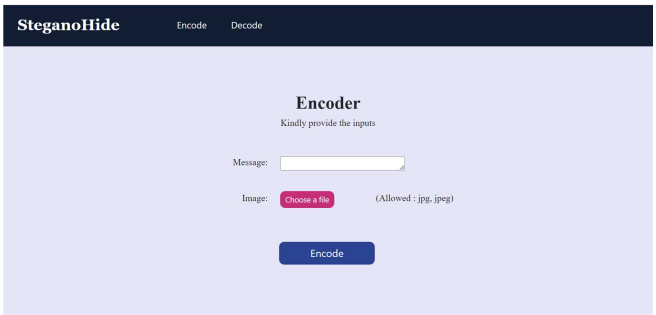

Fig. 4 Encode Page

Step 3: Now upload the cover image from your system and enter the message you want to encode. Fig. 5 depicts this process.

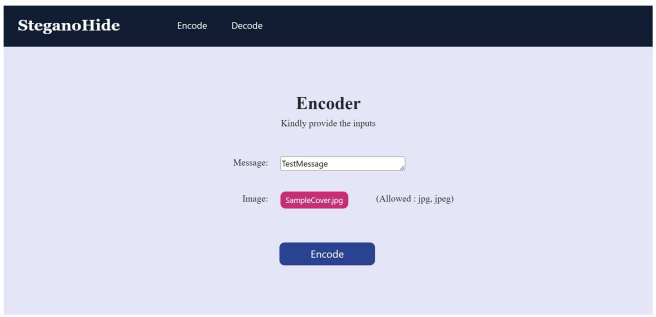

Fig. 5 Encryption process

Step 4: Upon uploading of image and message, you can click the encode button. A pop-up will indicate that the message has been encoded and the stego image will be downloaded on your system. The same process is depicted in fig. 6.

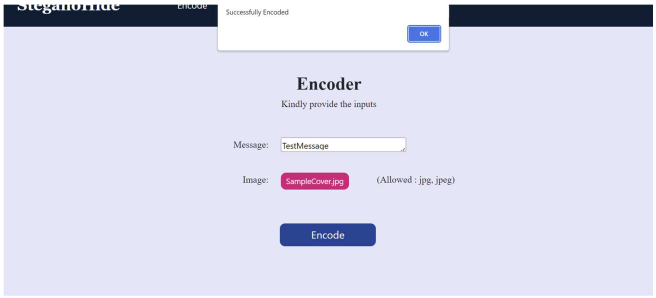

Fig. 6 Encoding performed

Step 5: For decoding, you can return to the home page or by clicking decode in the header bar. The decode page is as depicted in fig. 7.

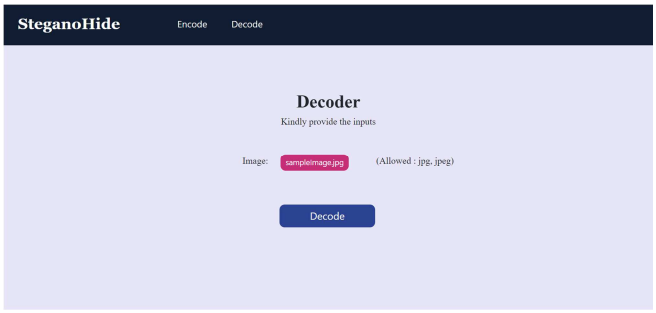

Fig. 7 Decode Page

Step 6: Select the stego image and click on decode button. A pop-up will show that the message is decoded. The decoded message will be shown on the page. The same is depicted in fig. 8.

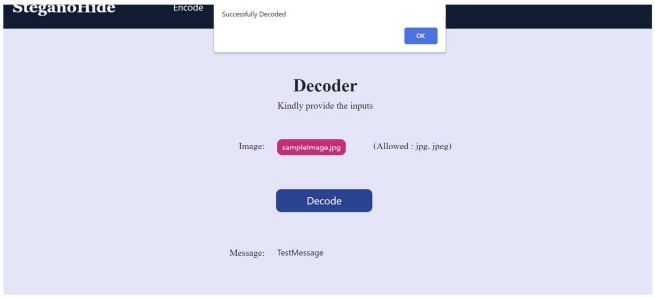

Fig. 8 Decode Process

7. Conclusion and Future Work

This paper discusses the basic concepts of steganography and steganalysis. It also focusses on the existing techniques in the field. A user-friendly interface has been developed for steganography. The application is implemented using a modified version of the LSB algorithm for encryption and decryption. The key advantage of the proposed approach is that it evades the use of secret key during encryption and decryption.

Although a lot of work has been carried out in both the fields of steganography and steganalysis, there is yet scope for future works. Various guidelines govern the development of new schemes in both the fields. Developing a steganography technique that can accomplish a fair trade-off between imperceptibility and security is the biggest challenge in the field. Concentrating on one of these characteristics can cause the other properties to become exposed because they are mutually exclusive.

For steganalysis, blind steganalysis techniques could possibly be used in the interim to uncover even undiscovered steganalysis schemes. Another idea could be to utilize deep learning architectures to construct steganalysis schemes. They can be greatly used to perform blind steganalysis and perform the process of feature extraction and classification in a single algorithm. The classification is also effective in deep learning architectures

References

[1] Martin A, Sapiro G, Seroussi G "Is image steganography natural?", IEEE Transactions on Image processing, 2005

[2] Phan, R.C.W, Ling, H.C "Steganalysis of Random LSB Insertion Using Discrete Logarithms Proposed At Cita03", MMU International Symposium on Information and Communication Tecnologies/M2USIC, Petaling Jaya, Malaysia, 2003

[3] Kodovský J, J. Fridrich "Influence of embedding strategies on security of steganographic methods in the JPEG domain", Proc. SPIE, Electronic Imaging, Security, Forensics, Steganography, and Watermarking of Multimedia Contents X, 2008.

[4] Fridrich J, "Steganography in Digital Media: Principles, Algorithms, and Applications", Cambridge, UK: Cambridge University Press, ISBN: 978-0-521-19019-0, 2010

[5] Cheddad, Abbas, "Digital image steganography: Survey and analysis of current methods." Signal processing 90.3, 2010.

[6] Artz, Donovan. "Digital steganography: hiding data within data." IEEE Internet computing 5.3, 2001

[7] Goel P "Data hiding in digital images: a Steganographic paradigm" Department of Computer Science & Engineering Indian Institute of Technology, Kharagpur, 2008

[8] Gunjal BL, Manthalkar RR "An overview of transform domain robust digital image watermarking algorithms", Journal of Emerging Trends in Computing and Information Sciences, 2010

[9] Chedded A "Steganoflage: a new image steganography algorithm", School of Computing & Intelligent Systems Faculty of Computing & Engineering University of Ulster, 2009

[10] Saha B, Sharma S "Steganographic techniques of data hiding using digital images", Def Sci J 62(1), 2012

[11] L. Rathika, B. Loganathan "Approaches and Methods for Steganalysis – A Survey", International Journal of Advanced Research in Computer and Communication Engineering, 2017

[12] Patil K, Gupta R, Singh G "Digital image steganalysis schemes for breaking steganography" International Conference on Advances in Communication and Computing Technologies, 2012