Novel Steganography Method for Secure Data Transmission

Krishnadeva K¹, Sripa Vimukthi Bannakkotuwa¹, Gayani Sandamali Wijesinghe¹, Lakmal Rupasinghe¹, Jenny Krishara¹, Chamodi De Silva¹, Thilini Weerasooriya¹, Hasini Perera¹

¹ Department of Information Systems Engineering, Faculty of Computing, Sri Lanka Institute of Information Technology, New Kandy Rd, Malabe, Sri Lanka

Abstract— Steganography is the art of hiding information to prevent the detection of hidden messages. Cryptography can be defined as the conversion of plain text into a scrambled code. Both are two general ways of sending vital information in a secret way. A message in scrambled code aka cipher text, might arouse suspicion on the part of the recipient while an invisible message created with steganographic methods will not. But many steganographic systems are weak against visual and statistical attacks. Systems without these weaknesses offer only a relatively small capacity for a steganographic message. In this paper, we present a new steganographic technique to hide data inside an image. The newly planned technique withstands visual and statistical attacks, yet it still offers a large steganographic capacity. This method does not modify the image. Message to be hidden will be converted into a bit stream and divided into several blocks. The bit pattern of each block is identified inside the image and their location is stored in an index table. This index table represents the locations of data inside the image. Later this table will be encrypted and transferred along with the image.

Keywords— Steganography, Cryptography, Steganalysis, Algorithm, Encryption, Decryption, Stego image, Data hiding, LSB.

I. INTRODUCTION

Data is the raw form of information, stored in columns and rows in databases, network servers, and personal computers. This information is generally from personal files and intellectual property for market analysis which are confidential. With the rapid development of the internet and the technology, personal and sensitive information is at risk. Hence, data security has become one of the most significant factors in information communication technology and data from unauthorized disclosure and modification. Cryptography and Steganography are studies focused on data security. Cryptography enhances security by providing confidentiality, integrity, authentication and non-repudiation for sensitive information. Cryptography uses encryption algorithms which jumble information by using a key which is known to both parties involved in the communication. Any third party cannot interpret the message due to data encryption. Yet, there is a possibility of a third party getting suspicious that there is a message embedded inside the cover medium. Steganography plays a major role in avoiding such situations. Steganography is the mechanism used to hide the existence of the message in the covered medium [1]. The common modern technique of steganography exploits the property of the media itself to convey a message. Plaintext, still imagery, audio and video, and IP datagram can be used for digitally embedding the message. Digital images are the most favored because of their frequency on the Internet [2]. This research is concerned with providing a high-level security mechanism for data transmission considering images. Both cryptography and steganography are combined to achieve this purpose.

II. SIMILAR TECHNOLOGIES

A. Random Pixel Selection

In this algorithm, data is hidden randomly. Data is hidden in some randomly selected pixel. Random pixel is generated by using Fibonacci algorithm. A novel data embedding method using random pixels selection, this embedded data method uses a multilevel histogram shifting technique. In the embedding process, random pixels in natural images were selected to be used in hiding data. The result showed that the embedded data was distributed in a more irregular manner and can better evade the detection of Statistical Steganalysis tool. This method obtained better Stego image quality. In comparison to another similar work, this approach provided better security while offering low distortion [9].

Replacing Intermediate Bits Using this technique, any bit or any intermediate bit from a given byte (of a pixel value) can be substituted by the bit of data to be hidden. Bits are replaced according to any random sequence from LSB to MSB position. Raster Scan Principle Pixels from alternate horizontal lines are used for replacing the secret information. A simple LSB scheme can be used for pixels of the first horizontal line. Then the second line is skipped. Again the third line is used to hide secret information and so on.

A.1 Random Scan Principle

The sequence, in which pixels are drawn, they are used to hide secret information. Again any simple data hiding algorithm like LSB can be used to hide secret information. By this method, data can be hidden in random pixels in an image.

A.2 Color Based Data Hiding

In this scheme, one fixed color is used to hide secret data. Intensity values of this fixed color are converted into binary format, and the secret information is hidden in this binary data.

A.3 Shape Based Data Hiding

Any shape can be taken to hide the data in an image. Confidential information can be hidden only in the pixels which are available in shape, instead of hiding secret

IJERTV6IS010307
(This work is licensed under a Creative Commons Attribution 4.0 International License.)
information in the whole image. We can use any shape having any dimensions.

B. Pixel Value Differencing Steganography
The pixel-value differencing (PVD) scheme provides high imperceptibility to the Stego image by selecting two consecutive pixels and designs a quantization range table to determine the payload by the difference value between the consecutive pixels. Besides, it offers the advantage of conveying a large number of payloads, while still maintaining the consistency of an image characteristic after data embedding [10]. PVD is designed in such a way that the pixel modification does not violate gray scale range interval. The selection of the range intervals is based on the characteristics of human vision sensitivity to gray value (0-255) varies from smoothness to contrast. It provides an way to produce a more imperceptible result than simple LSB replacement methods. The embedded secret message can be extracted from the resulting Stego image without referencing the original cover image. Moreover, to achieve secrecy protection of hidden data, a pseudo-random mechanism may be used. If secret data is stored randomly it is difficult to understand by the intruder. PVD embedding is used for edged areas to increase image quality. It is also used to hide message into grayscale as well as in color image [11].

C. Intensity Based Steganography
In this method, all the three-color planes will be converted into binary values. For each pixel in the image, the plane which has the minimum number of ones in its MSB will act as index plane and the other two color planes are considered as data planes. Compared to method 1 and method 2 in the existing work, this approach will help us to embed number of message bits in the cover medium [13].
RGB is the most common and simplest model. The technique is more secure. The third party cannot easily detect the presence of hidden data. One of the main advantages is its capacity because it embeds a large amount of data as compared to previous techniques. The main benefit of the proposed algorithm is if we utilize all the bytes of cover-image to hide the data bits, then the algorithm has a very high capacity of data hiding. Average number of bytes changed in image for embedding data and Average number of bits changed per pixel

D. LSB Based Steganography Method
Least Significant Bit method is the most popular as well as the easiest amongst the steganography methods. In this LSB technique, the secret message can be hidden within a cover image. The main concept of this method is to get the bit representation of the secret message and the bit pattern of the cover image separately and then replace the least significant bit (last bit) of the cover image by the bits of the message. The number of bits from the secret message that can be stored within a cover image will be differing according to the type of the image. When a 24-bit image is used as the cover image of this process, there are three components such as Red, Green, and Blue and therefore three bits from the secret message can be stored in each pixel. At the same time, if it’s an 8-bit cover image, only one bit can be modified. But there’s a huge difference in those two cases since the change in the 24-bit image can’t be easily detected by the human while the change of the 8-bit image can be easily detected due to the color changes [16].

III. RESEARCH OBJECTIVE
Most of the existing stenographic methods apply the Least Significant Bit (LSB) as a technique of hiding a message inside an image. In this method, the binary representation of the message overwrites the LSB of each byte in the cover image. It is important that the Stego-image does not contain any detectable artifacts. Least significant bits are the less important bits in a byte. Therefore changing these bits has no high impact in modifying the original image. It is indiscernible to the human eye. Modern steganalysis techniques are capable of detecting the embedded message by performing analysis statically or by identifying the signature of stenographic tools. View presents in this paper are to provide an advanced solution for steganography that can be used for the high reliable communication when transferring images. The primary concern of this method is to use an image for data communication instead of embedding the data inside the image. This method will not modify the image. It generates an index table to represents the locations of data inside the image.

This research is focused on introducing an advanced steganography method for transferring images (JPEG, BMP, GIF, and PNG). An image can be illustrated as an array of numbers which represents light intensities at various points (pixels). Each pixel is derived from three primary colors: Red, Green, and Blue, and each primary color is represented using 1 byte. 24-bit images use, 24 bits (3 bytes) per pixel to represent a color value. A message can be embedded by modifying these bits in an image.

IV. SYSTEM OVERVIEW AND DESIGN
A standard approach for hiding information in digital images include LSB insertion, masking and filtering algorithms and transformations. Each of these techniques can be applied, with varying degrees of success to different image files. Most stenographic techniques involved in changing properties of the cover source, but these changes can be detected. Statistical analysis is the widely-used method for this purpose to identify the difference between random values and real image values. Using this technique, it is also possible to detect messages hidden inside JPEG files with the DCT method, since it also involves bit modifications [3]. In contrast, the image is not modified in the proposed method, but it provides high security by standing against visual and statistical attacks. Instead of modifying the image, an index table will be generated to point out the locations of data inside the image. The encrypted image and the encrypted Index table are transmitted separately. Hence, intercepting one transmission is insufficient to access the data.

This overall concept has been achieved using several phases. In this section, we mainly focus on 3 main phases: encryption phase, transmission phase, and decryption phase. In the Encryption phase, the inputs are given: sensitive data, secret key and career image. Delivering appropriate index table is the
main responsibility of this phase. In Transmission phase, the data is sent to the receiver using transmission media: email, the web or portable storage. In the decryption phase, the data is retrieved using the career image and image table using the correct key.

V. METHODOLOGY

In the proposed algorithm an image is used to provide high security to the message, but the message is not embedded inside the image. Instead, it generates an index table to represents the message inside the image. This index table contains locations of each byte in a message inside the image. Later, index table has to be transmitted along with the image to the receiver for retrieval of the message. Implementation of the proposed steganographic system consists of 3 major modules:

- Encryption Module
- Decryption Module
- Recovery Module

Cover image and the data file are the inputs of the Encryption module. It will accept any images regardless of its type. Then it converts the data file into index table which will represent the locations of each byte of the data file in the image. In the process of generating the index includes the following procedure.

First, the availability of the inputs are verified: cover image and data file. Second, the career image is converted to a byte array and check all the possible 8-bit patterns are contained in the image. If not available, an appropriate image handler module is used to add missing patterns randomly. Image handler module is a sub module of the encryption module. The purpose of this module is to ensure the suitability of the career image. Third, the data file is converted to a byte array, and get the hash value of the data file. Then search the locations of each element in the data array, within the image byte array. Store the matching coordinates in a separate array. Furthermore, file name, file type and the hash value are added to the index table. Finally, the index table is encrypted using a proper key and compressed.

Decryption module retrieves data using the index table and career image. Career image plays a huge role in this process. If the correct career image is not selected, the data cannot be retrieved. Hash verification is done for verification of data transmission. In this module, following operations are performed. First, the availability of all inputs are verified: career image and encrypted index table. Second, the index table is decompressed and decrypted. Then the hash value and the file name is retrieved from the index table. Simultaneously, the image is converted into a byte array and retrieve the data in the locations mentioned in index table and reconstruct the data file. Finally, the hash value is verified and generate the data file.

In addition to the Encryption and decryption modules, recovery module is used for higher reliability. Encryption is an optional module, which is used to retrieve data when the image is corrupted during the transmission. The image is divided into two parts and generate two individual arrays of the index table. Next, a single index table is generated by combining them. Finally, a parameter is added to the index table to indicate the recovery option is selected. If the image is modified during the transmission, data is retrieved by the Recovery module using both index tables.

First, the normal retrieval operation is performed. If fails, the second index table is used. In this case, a small portion of the image is used to retrieve data. Alternatively, if the portion used by the recovery module is damaged, data cannot be retrieved. The proposed method uses new algorithms shown in figure 1, to generate the index table.

A. Index table generating algorithm
B. Pattern checker algorithm

![Figure 1: Index table generating and pattern checker](image-url)

A. Index table generating algorithm
The purpose of this algorithm is to generate index table. The initial step is to convert the message into a byte array. Then a search operation is performed to identify the same bit pattern of each element in byte array inside the image. Identified locations are stored sequentially in another array. This newly created array is converted into a string and is written to a text file.

B. Pattern checker algorithm
Availability of possible 8 bits patterns in the image is checked using this algorithm. It converts the image into a byte array and searches for all patterns to identify any missing patterns. Then it inserts missing patterns randomly. The same procedure is done at the receivers end to generate same image array. This module is to ensure that the system supports all possible characters in the message.
VI. RESULTS AND DISCUSSION

Based on the testing results, the functionality and the performance of the proposed method are identified. To evaluate the system performance on encryption and decryption, the duration was recorded respect to the various file sizes. A comparison was carried out using the existing steganography tools and the results are shown in Table 1.

Table 1: Test results in comparison with existing Steganography tools

<table>
<thead>
<tr>
<th>Tools</th>
<th>File Size</th>
<th>Encryption</th>
<th>Decryption</th>
<th>Encryption</th>
<th>Decryption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed System</td>
<td>10.012</td>
<td>1.8</td>
<td>45</td>
<td>1.8</td>
<td>170</td>
</tr>
<tr>
<td>S-tool</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>can't do</td>
</tr>
<tr>
<td>Opennuff</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>can't do</td>
</tr>
</tbody>
</table>

Table 2: Comparison between the LSB technique and proposed method

<table>
<thead>
<tr>
<th>#</th>
<th>Features</th>
<th>LSB</th>
<th>Proposed Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Image Types</td>
<td>Only selected image types</td>
<td>All image types</td>
</tr>
<tr>
<td>2</td>
<td>Data Size</td>
<td>Depend on image size</td>
<td>Independent of image size</td>
</tr>
<tr>
<td>3</td>
<td>Hiding Time</td>
<td>Fast</td>
<td>Slow</td>
</tr>
<tr>
<td>4</td>
<td>Retrieval Time</td>
<td>Fast</td>
<td>Fast</td>
</tr>
<tr>
<td>5</td>
<td>Statistical attack</td>
<td>High Possibilities</td>
<td>No impact</td>
</tr>
<tr>
<td>6</td>
<td>Complexity</td>
<td>High</td>
<td>No impact</td>
</tr>
</tbody>
</table>

In other steganographic techniques, the career image size should be larger than the data file size [4], but in the proposed method the file size is independent of the image size. A comparison is done between the proposed method and the LSB technique on important characteristics. The results are shown in Table 2. The researcher also compares the proposed method with the S-tool which is considered as one of the leading Steganography tools in the market. The comparison results are shown in Table 3.

Table 3: Comparison of features and facility: S-Tools and proposed method

<table>
<thead>
<tr>
<th>#</th>
<th>Features</th>
<th>S-Tools</th>
<th>Proposed Tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Image Types (compatible)</td>
<td>bmp, gif</td>
<td>All image types</td>
</tr>
<tr>
<td>2</td>
<td>Image Size (maximum)</td>
<td>Unlimited</td>
<td>Unlimited</td>
</tr>
<tr>
<td>3</td>
<td>Data Size</td>
<td>Depend on image size</td>
<td>Independent of image size</td>
</tr>
<tr>
<td>4</td>
<td>Hiding Time</td>
<td>Fast</td>
<td>Slow</td>
</tr>
<tr>
<td>5</td>
<td>Retrieval Time</td>
<td>Fast</td>
<td>Fast</td>
</tr>
<tr>
<td>6</td>
<td>Statistical attack</td>
<td>High Possibilities</td>
<td>No impact</td>
</tr>
</tbody>
</table>

Table 4: Comparison of attacks/ vulnerabilities: S-Tools and proposed method

<table>
<thead>
<tr>
<th>#</th>
<th>Attacks</th>
<th>S-Tool</th>
<th>Proposed Tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Stego-only attack</td>
<td>Less impact</td>
<td>No impact</td>
</tr>
<tr>
<td>2</td>
<td>Known cover attack</td>
<td>High impact</td>
<td>No impact</td>
</tr>
<tr>
<td>3</td>
<td>Known message attack</td>
<td>High impact</td>
<td>No impact</td>
</tr>
<tr>
<td>4</td>
<td>Chosen stego attack</td>
<td>High impact</td>
<td>Less impact</td>
</tr>
<tr>
<td>5</td>
<td>Chosen message attack</td>
<td>High impact</td>
<td>High impact</td>
</tr>
<tr>
<td>6</td>
<td>Known stego attack</td>
<td>High impact</td>
<td>Less impact</td>
</tr>
</tbody>
</table>

A. Encryption Process

Figure 2: Test results for decryption process for difference data sizes

B. Decryption Process

The proposed technique is comparatively slow in encryption and fasts in decryption as shown in figure 2 and figure 1 respectively. The delay in the encryption is due to the search algorithm which generates index table. Following records shown in figure 2 and figure 1 are calculated using an image of 75KB size.
VII. STEGANALYSIS ADVANCEMENT

Steganalysis is the discovery of the existence of hidden information. The goal of steganalysis is to discover hidden information and to break the security of its carriers [5]. LSB method is widely used for image steganography [6]. In LSB method, least significant bits of the cover pixels are overwritten by the secret bit stream and steganalysis program needs to read back the LSBs of the image. Paris Analysis, a higher-order steganalysis method implemented [7], detects secret messages embedded in digital images. This method is used to estimate the length of the secret message. Other steganalysis methods [8] use different approaches to work against any steganographic embedding algorithm. All steganalysis methods are based on a common characteristic. They detect secret message by using the modification done to the image during the embedding process. In this proposed method it avoids the image modification which eliminates the steganalysis attack.

There are two approaches for steganalysis, steganalysis method specific to a particular steganographic algorithm and steganalysis technique works with any steganographic embedding algorithm, even with an unknown algorithm. The proposed system generates an index table instead of embedding data into the image. Both index table and image are transmitted separately, intercepting one communication is insufficient for an attacker to get the message. Selection of image which both parties are having, will avoid the image transmission and completely eliminate steganalysis attacks because there is no way that the attacker can get the image.

REFERENCES


