DATA SECURITY AND INTEGRITY USING DATA HIDING

Deepak Singla*

ABSTRACT

Steganography is a technique of hiding information in digital media. In contrast to cryptography, it is not to keep others from knowing the hidden information but it is to keep others from thinking that the information even exists. In this paper we propose an advanced system of encrypted data embedded into an image file using random LSB insertion method in which the secret data are spread out among the image data in a seemingly random manner. This is achieved using a secret key. It combined both feature of steganography and cryptography. To enhance security level that. It also provides integrity and message authentication using MAC algorithm.

Keywords: *Authentication, Cryptography, Integrity, Privacy, Security, Secrecy, Steganography, Text stenography.*

*PEC University of Technology, Chandigarh.

I. INTRODUCTION

In the field of Data Communication, security-issues have got the top priority. So, of late the degree of security provided by a security tool has become the main evolutionary criteria of it. Classical cryptography is one of the ways to secure plain text messages. Along with that at the time of data transmission, security is also implemented by introducing the concept of steganography, etc. In this types of combined approach, there exits some drawbacks.

In remote networking, at the time of transmission of hidden encrypted text message, if the eavesdroppers get the track of the hidden text, then they could easily get the encrypted text. Now breaking of encrypted text message can be achieved by applying some brute force technique. So, there remains some probability of snooping of information. So, this type of techniques incurs another level of security which can route the Cryptanalyzer or Steganalyzer in a different direction.

We couple both Encryption-Decryption as well as hashing. The problem with Encryption-Decryption algorithms is that they provide secrecy, or confidentiality, but not integrity. To preserve the integrity of the message, both message and fingerprint are required. This fingerprint is in the form of a message authentication code. Thus, we propose a project with combines both cryptography and steganography. In the process, it takes advantage of all the positives of these respective fields. Steganography provides the ability to hide information in plain sight. Cryptography ensures the basic components of security, namely: Confidentiality and Integrity.

II. RELATED WORK

A. Basic Overview of Cryptography

A large number of cryptography algorithms have been created till date with the primary objective of converting information into unreadable ciphers. We will discuss the two basic and most commonly used algorithms – The joint key cryptography and the public key cryptography.

1) Joint Key Cryptography: It uses a common key for encryption and decryption of the message. This key is shared privately by the sender and the receiver. The sender encrypts the data using joint key and then sends it to the receiver who decrypts the data using same key to retrieve the original message. Joint key cipher algorithms are less complex and execute faster as compared to other forms of cryptography but have an additional need to securely share the key. In this type of cryptography the security of data is equal to the security of the key. In

other words it serves the purpose of hiding a smaller key instead of the huge chunk of message data.

2) Public Key Cryptography: It is a technique that uses a different key for encryption as the one used for decryption. Public key systems require each user to have two keys a public key and a private key. The sender of the data encrypts the message using the receiver's public key. The receiver then decrypts this message using the private key. This technique eliminates the need to privately share a key in case of symmetric key cipher. Asymmetric cryptography is a fundamental and most widely used technique, and is the approach which underlines Internet standards such as Transport Layer Security (TLS). The most common algorithm used for secret key systems us the Data Encryption Algorithm (DEA) defined by the Data Encryption Standard (DES).

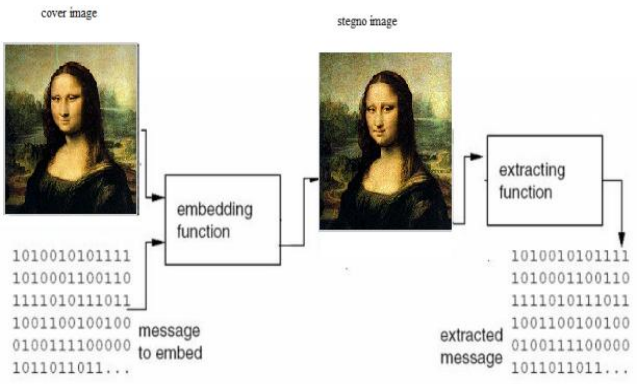
B. Basic Overview of Steganography

Steganography is the art of hiding information in plain sight. Looking at data in transmission it is very easy to detect if its encrypted or not. Thus, the focus of cryptography is not to hide the data but to have a strong enough algorithm so that it is difficult to break cipher text. On the other hand, steganography focuses on hiding the secret message. No algorithm is applied on the information. As soon as it is detected that the career medium has hidden information, the goal of steganography is defeated. Several methods have been proposed to hide information in a better way. We have concentrated on some techniques and methods which are described below.

1) Methods of Digital Steganography: There are many techniques available to the digital steganographer. The most common technique is to exploit the lenient constraints of popular file formats. In a normal file format there is a header and a data part. Many fields in the header are not used. In digital steganography, such fields of file formats are used to images as Carriers store data. In this way, no impact is made on the original media. Many publicly available software packages use this technique on a variety of media.

2) Images as Carriers*:* Information can be hidden in many different ways in images. To hide information, straight message insertion may encode every bit of information in the image or selectively embed the message in noisy areas that draw less attention those areas where there is a great deal of natural color variation. The message may also be scattered randomly throughout the image. Redundant pattern encoding wallpapers the cover image with the message. A number of ways exist to hide information in digital images. Common approaches include least significant bit insertion.

3) Image Files: To a computer, an image is an array of numbers that represent light intensities at various points (pixels). These pixels make up the image's raster data. A common image size is 640 × 480 pixels and 256 colors (or 8 bits per pixel). Such an image could contain about 300 kilobits of data. Digital images are typically stored in either 24-bit or 8-bit files. A 24-bit image provides the most space for hiding information; however, it can be quite large (with the exception of JPEG images). All color variations for the pixels are derived from three primary colors: red, green, and blue. Each primary color is represented by 1 byte; 24-bit images use 3 bytes per pixel to represent a color value. These 3 bytes can be represented as hexadecimal, decimal, and binary values. In many Web pages, the background color is represented by a six-digit hexadecimal number—actually three pairs representing red, green, and blue. A white background would have the value FFFFFF: 100 percent red (FF), 100 percent green (FF), and 100 percent blue (FF). Its decimal value is 255, 255, 255,and its binary value is 11111111, 11111111, 11111111, which are the three bytes making up white.

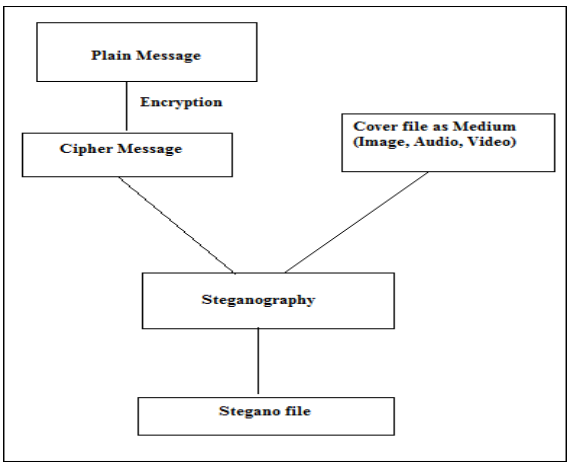


This definition of a white background is analogous to the color definition of a single pixel in an image. Pixel representation contributes to file size. For example, suppose we have a 24-bit image 1,024 pixels wide by 768 pixels high—a common resolution for high-resolution graphics. Such an image has more than two million pixels, each having such a definition, which would produce a file exceeding 2Mbytes. Because such24-bit images are still relatively uncommon on the Internet, their size would attract attention during transmission. File compression would thus be beneficial, if not necessary, to transmit such a file.

4) File Compression: Two kinds of compression are lossless and lossy. Both methods save storage space but have different results, interfering with the hidden information, when the information is uncompressed. Lossless compression lets us reconstruct the original message exactly; therefore, it is preferred when the original information must remain intact (as with steganographic images). Lossless compression is typical of images saved as GIF (Graphic Interchange Format) and bit BMP (a Microsoft Windows and OS/2 bitmap file)

Lossy compression is a data encoding method that compresses data by discarding (losing) some of it. The procedure aims to minimise the amount of data that need to be held, handled, and/or transmitted by a computer. Thus, lossy compression, on the other hand, saves space but may not maintain the original images integrity. This method typifies images saved as JPEG.

5) Embedding data: Embedding data, which is to be hidden, into an image requires two files. The first is the innocent-looking image that will hold the hidden information, called the cover image. The second le is the message the information to be hidden. A message may be plaintext, ciphertext, other images, or anything that can be embedded in a bit stream. When combined, the cover image and the embedded message make a stego image.

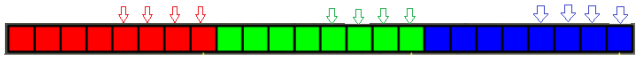


A stego-key (a type of password) may also be used to hide, then later decode, the message. Most steganography software neither supports nor recommends using JPEG images, but recommends instead the use of lossless 24-bit images such as BMP. The next-best alternative to 24-bit images is 256-color or gray-scale images. The most common of these found on the Internet are GIF files.

6) Least significant bit insertion: Least significant bit (LSB) insertion is a common, simple approach to embedding information in a cover file. Unfortunately, it is vulnerable to even a slight image manipulation. For 24-bit images, to hide an image in the LSBs of each byte of a 24-bit image, you can store 3 bits in each pixel. A 1,024 * 768 image has the potential to hide a total of 2,359,296 bits (294,912 bytes) of information. If you compress the message to be hidden before you embed it, you can hide a large amount of information. To the human eye, the resulting stego-image will look identical to the cover image.

There are two approaches are used in least significant bit insertion:

4LSB:- When the second half (i.e four bits) of a color byte are used for hiding the message.



For example, the letter Z can be hidden in three pixels (assuming no compression). The original raster data for 3 pixels (9 bytes) may be

(00100111 11101001 11001000)

(00100111 11001000 11101001)

(11001000 00100111 11101001)

The binary value for Z is 01011011. Inserting the binary value for A in the three pixels would result in

(00100101 11101011 11001000)

The underlined bits are actually changed in the 2 bytes used.

1LSB:- To hide an image in the LSBs of each byte of a 24-bit image, you can store 3 bits in each pixel.

For example, the letter Z can be hidden in three pixels (assuming no compression). The original raster data for 3 pixels (9 bytes) may be

(00100111 11101001 11001000)

(00100111 11001000 11101001)

(11001000 00100111 11101001)

The binary value for Z is 01011011. Inserting the binary value for A in the three pixels would result in

(00100110 11101001 11001000)

(00100111 11001001 11101000)

(11001001 00100111 11101001)

The underlined bits are actually changed in the 8 bytes used. On average, LSB requires that only half the bits in an image be changed. You can hide data in the least and second least significant bits and still the human eye would not be able to discern it.

Apply both 4LSB and 1LSB technique on this image

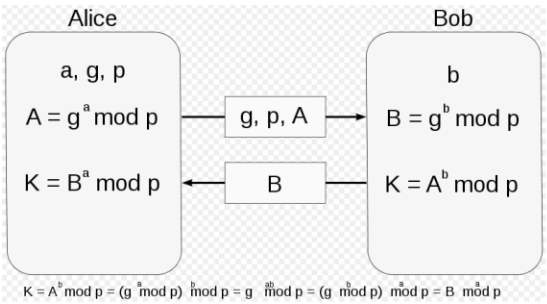
Visualization Effects By Hiding by 1 bit and 4 bits.

7) 8-bit images: 8-bit images are not as forgiving to LSB manipulation because of color limitations. Steganography software authors have devised several approaches—some more successful than others—to hide information in 8-bit images. First, the cover image must be more care-fully selected so that the stego-image will not broadcast the existence of an embedded message. When information is inserted into the LSBs of the raster data, the pointers to the color entries in the palette are changed. In an abbreviated example, a simple four-color palette of white, red, blue, and green has corresponding palette position entries of 0 (00), 1 (01), 2 (10), and 3 (11), respectively. The raster values of four adjacent pixels of white, white, blue, and blue are 00 00 10 10. Hiding the binary value 1010 for the number 10 changes the raster data to 01 00 11 10, which is red, white, green, blue. These gross changes in the image are visible and clearly highlight the weakness of using 8-bit images.

8) Audio File Carriers: The carrier is a digitized audio information. In case of digital sound CDs the sampling frequency standard is 44 kHz and frequency standard is 44 kHz and the quantizing resolution is at least 12 bits. That yields an amplitude scale of 4096 levels, in which changing the LSB means a 0.025% relative modification which is essentially a distortion. Such an extremely little distortion cannot be perceived by human ear. In case of silent parts of a sound track it still may be overheard.

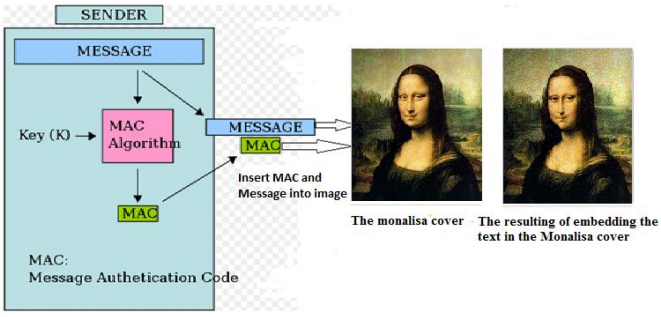
III. PROPOSED WORK

Our work deals with the security of text messages at the time of sending it over the network. In our algorithm, we have used symmetric key cryptography which means same keys are needed to encrypt and decrypt the data. Key are exchanged by Diffie-Hellman Cryptosystem.

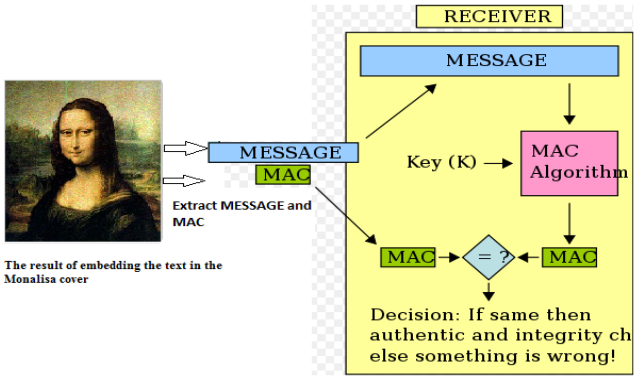


Encryption and decryption provide privacy, but not integrity and message authentication.

To provide message integrity and message authentication, sender use MAC (message authentication code) algorithm that produce MAC, the MAC is created at sender site and is inserted into image with the message. To check the integrity and message authentication, receiver creates the MAC algorithm again and compares the new MAC with the one received. If both are the same, the receiver is sure that the original message has not been changed and sender of the message is definitely Alice.



Sender side



Receiver side

IV. CONCLUSION

As steganography becomes more widely used in computing there are issues that need to be resolved. There are wide varieties of different techniques with their own advantages and disadvantages. Constant improvement and changes need to be made and newer versions released. Steganography has its place in security. It is not intended to replace cryptography but supplement it. Hiding a message with steganography methods reduces the chance message being detected. In our work, we presented Encryption and decryption provides privacy, but not integrity and authentication.

To provide message integrity and message authentication, sender use MAC (message authentication code) algorithm that produce MAC, the MAC is created at sender site and is inserted into image with the message.

Random steganography using LSB with key gives us more security than simple LSB method, where it is difficult to identify the hidden data in the stego image at specific location.

Simple steganography using LSB with more than one bit used for the hidden data gives us more space to store data but here cover image will lose its visual appearance.

REFERENCES

1. C. Kurak and J. McHugh, "A Cautionary Note nImage Downgrading," Proc. IEEE Eighth Ann. Computer Security Applications Conf., IEEE Press, Piscataway, N.J., 1992, pp. 153-159.
2. B.Pfitzmann, "Information Hiding Terminology," Proc.First Int'l Workshop Information Hiding, Lecture Notesin Computer Science No. 1,174, Springer-Verlag, Berlin, 1996, pp. 347-356.

3. T. Aura, "Invisible Communication," EET 1995, technical report, Helsinki Univ. of Technology, Finland, Nov. 1995; http://deadlock.hut.fi/ste/ste_html.html. 4. I. Cox et al., "A Secure, Robust Watermark for Multi-media," Proc. First Int'l Workshop Information Hiding,Lecture Notes in Computer Science No. 1, 174, Springer-Verlag, Berlin, 1996, pp. 185-206.
4. A. Brown, S-Tools for Windows, 1994, ftp://idea.sec.dsi.unimi.it/pub/security/crypt/code/s-tools3.zip.
5. E. Koch, J. Rindfrey, and J. Zhao, "Copyright Protection for Multimedia Data," Proc. Int'l Conf. Digital Media and Electronic Publishing, Leeds, UK, 1994.
6. W. Brown and B.J. Shepherd, Graphics File Formats: Reference and Guide, Manning Publications, Green-wich, Conn., 1995.
7. X-G. Xia, C.G. Boncelet, and G.R. Arce, "A Multiresolution Watermark for Digital Images," IEEE Int'l Conf.Image Processing, IEEE Press, Piscataway, N.J., 1997.
8. Neil F. Johnson and Sushil Jajodia, "Digital Steganography: Hiding Data within Data," IEEE int'l Conf. Image processing IEEE Press, 2004 .
9. William Stallings, "Cryptography and Network Security, Principles and Practice" Edition 3rd. PRENTICE HALL 2003, ISBN 0-13-091429-0
10. Kevin Curran and Karen Bailey "An evaluation of Image based Steganography methods" International Journal of Digital Evidence Fall 2003, Vol-2, Issue-2
11. Dr. K. Duraiswamy and R.Umarani "Security through obscurity"
12. Piyush Marwaha and Paresh Marwaha, "Visual Cryptographic Steganography in Images"
13. M.S. Sutaone and M.V. Khandare, "Image Based Steganography using LSB insertion"
14. Debnath Bhattacharyya, Poulami Das, Samir Kumar Bandyapadhyay and Tai-hoon Kim, "Text Steganography: A novel approach"