STEGANOGRAPHY (THE ART OF HIDING INFORMATION)

Megha Goyal*, Maninder Kaur

Doaba Institute of Engineering and Technology, Kharar

ABSTRACT

Steganography is a process that involves hiding a message in an appropriate carrier for example an image or an audio file. The carrier can then be sent to a receiver without anyone else knowing that it contains a hidden message. This is a process, which can be used for example by civil rights organisations in repressive states to communicate their message to the outside world without their own government being aware of it. Less virtuously it can be used by terrorists to communicate with one another without anyone else's knowledge. In both cases the objective is not to make it difficult to read the message as cryptography does, it is to hide the existence of the message in the first place possibly to protect the courier. The initial aim of this study was to investigate steganography and how it is implemented. Based on this work a number of common methods of steganography could then be implemented and evaluated. The strengths and weaknesses of the chosen methods can then be analysed. To provide a common frame of reference all of the steganography methods implemented and analysed used GIF images. Seven steganography methods were implemented. The methods were chosen for their different strengths in terms of resistance to different types of steganalysis or their ability to maximise the size of the message they could store. All of the methods used were based on the manipulation of the least significant bits of pixel values or the rearrangement of colours to create least significant bit or parity patterns, which correspond to the message being hidden.

Keywords:-.Introduction, Steganography Vs Cryptography, Uses of Steganography, Steganography under various Media, Steganalysis, Conclusion, References.

I. INTRODUCTION

The word steganography comes from the Greek *Steganos*, which mean covered or secret and –*graphy* mean writing or drawing. Steganography is the art of concealing the existence of information within seemingly innocuous carriers. In broad sense, term Steganography is used for hiding message within an image. More precisely,

"the goal of steganography is to hide messages inside other harmless messages in a way that does not allow any enemy to even detect that there is a second secret message present."

Steganography (literally meaning *covered writing*) dates back to ancient Greece, where

common practices consisted of etching 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. In contrast to cryptography, where the "enemy" is allowed to detect, intercept and modify messages without being able to violate certain security premises guaranteed by a cryptosystem, the goal of steganography is to hide messages inside other "harmless" messages in a way that does not allow any "enemy" to even detect that there is a second secret message present.

The basic model of steganography consists of *Carrier*, *Message* and *Password*. Carrier is also known as *cover-object*, which the message is embedded and serves to hide the presence of the message. Basically, the model for steganography is shown on Figure 1. Message is the data that the sender wishes to remain it confidential. It can be plain text, ciphertext, other image, or anything that can be embedded in a bit stream such as a copyright mark, a covert communication, or a serial number. Password is known as *stego-key*, which ensures that only recipient who know the corresponding decoding key will be able to extract the message from a *cover-object*. The *cover-object* with the secretly embedded message is then called the *stego-object.*

Recovering message from a *stego-object* requires the *cover-object* itself and a corresponding decoding key if a *stego-key* was used during the encoding process. The original image may or may not be required in most applications to extract the message.

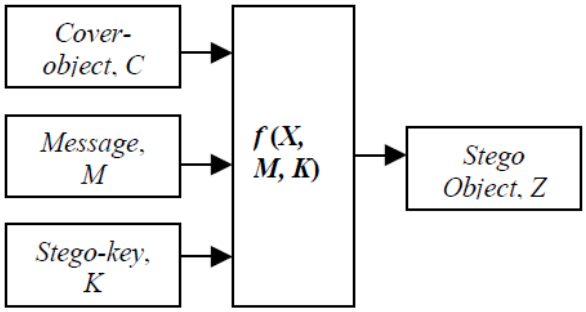

Figure 1 Basic Steganographic Model

Recovering message from a *stego-object* requires the *cover-object* itself and a corresponding decoding key if a *stego-key* was used during the encoding process. The original image may or may not be required in most applications to extract the message.There are several suitable carriers below to be the *cover-object* :

- Network Protocols such as TCP, IP and UDP
- Audio that using digital audio formats such as wav, midi, avi, mpeg, mpi and voc

- File and Disk that can hides and append files by using the slack space
- Text such as null characters, just alike morse code including html and java
- Images file such as bmp, gif and jpg, where they can be both color and gray-scale.

In general, the information hiding process extracts redundant bits from *cover-object*. The process consists of two steps

- Identification of redundant bits in a *cover-object*. Redundant bits are those bits that can be modified without corrupting the quality or destroying the integrity of the *cover-object*.
- The embedding process then selects the subset of the redundant bits to be replaced with data from a secret message. The *stego-object* is created by replacing the selected redundant bits with message bits. A possible formula of the process may be represented as: *cover medium + embedded message + stego key = stego-medium*

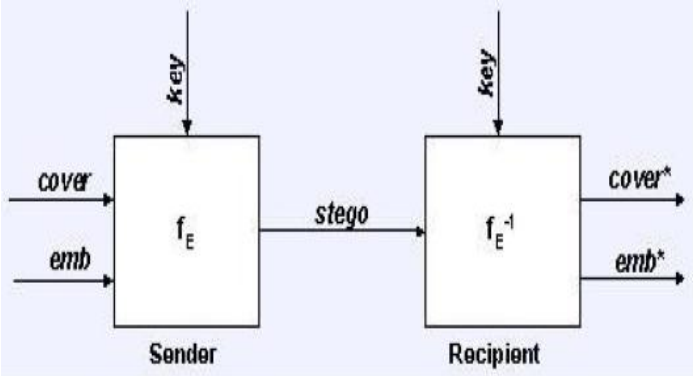

Figure 2 Graphical Version of the Steganographic Process
f_E : steganographic function "embedding"
f_{E-1} : steganographic function "extracting"
cover: cover data in which *emb* will be hidden
emb: message to be hidden
stego: cover data with the hidden message

For example:

II. STEGANOGRAPHY VS CRYPTOGRAPHY

Steganography can be viewed as akin to cryptography. Both have been used throughout recorded history as means to protect information. At times these two technologies seem to converge while the objectives of the two differ. Cryptographic techniques "scramble" messages so if intercepted, the messages cannot be understood. Steganography, an essence, "camouflages" a message to hide its existence and make it seem "invisible" thus concealing the fact that a message is being sent altogether. An encrypted message may draw suspicion while an invisible message will not.

In an ideal world we would all be able to openly send encrypted email or files to each other with no fear of reprisals. However there are often cases when this is not possible, either because you are working for a company that does not allow encrypted email or perhaps the local government does not approve of encrypted communication. This is where steganography can come into play.

A good steganography system should fulfill the same requirements posed by the "Kerckhoff principle" in cryptography. This means that the security of the system has to be based on the assumption that the "enemy" has full knowledge of the design and implementation details of the steganographic system. The only missing information for the "enemy" is a short easily exchangeable random number sequence, the secret key, and without the secret key, the "enemy" should not have the slightest chance of even becoming suspicious that on an observed communication channel hidden communication might take place. In cryptography, the system is broken when the attacker can read the secret message. Breaking a steganographic system need the attacker to detect that steganography has been used and he is able to read the embedded message. Steganography cannot be detected. Therefore, it 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 is often confused with cryptology because the two are similar in the way that they both are used to protect important information.

Steganography	Cryptography
Unknown message passing	Known message passing
Little known technology	Common technology
Technology still being developed for certain formats	Most algorithms known to government departments
Once detected message is known	Strong algorithm are currently resistant to brute force attack Large expensive computing power required for cracking Technology increase reduces strength
Many Carrier formats	

III. USES OF STEGANOGRAPHY

1. Steganography can be a solution which makes it possible to send news and information without being censored and without the fear of the messages being intercepted and traced back to us.

2. It is also possible to simply use steganography to store information on a location. For example, several information sources like our private banking information, some military secrets, can be stored in a cover source. When we are required to unhide the secret information in our cover source, we can easily reveal our banking data and it will be impossible to prove the existence of the military secrets inside.

3. Steganography can also be used to implement watermarking.Although the concept of watermarking is not necessarily steganography, there are several steganographic techniques that are being used to store watermarks in data. The main difference is on intent, while the purpose of steganography is hiding information, watermarking is merely extending the cover source with extra information. Since people will not accept noticeable changes in images, audio or video files because of a watermark, steganographic methods can be used to hide this.

4. E-commerce allows for an interesting use of steganography. In current e-commerce transactions, most users are protected by a username and password, with no real method of verifying that the user is the actual card holder. Biometric finger print scanning,combined with unique session IDs embedded into the fingerprint images via steganography, allow for a very secure option to open ecommerce transaction verification.

5. Paired with existing communication methods, steganography can be used to carry out hidden exchanges. Governments are interested in two types of hidden communications: those that support national security and those that do not. Digital steganography provides vast potential for both types. Businesses may have similar concerns regarding trade secrets or new product information.

6. The transportation of sensitive data is another key use of steganography. A potential problem with cryptography is that eavesdroppers know they have an encrypted message when they see one. Steganography allows to transport of sensitive data past eavesdroppers without them knowing any sensitive data has passed them. The idea of using steganography in data transportation can be applied to just about any data transportation method, from E-Mail to images on Internet websites.

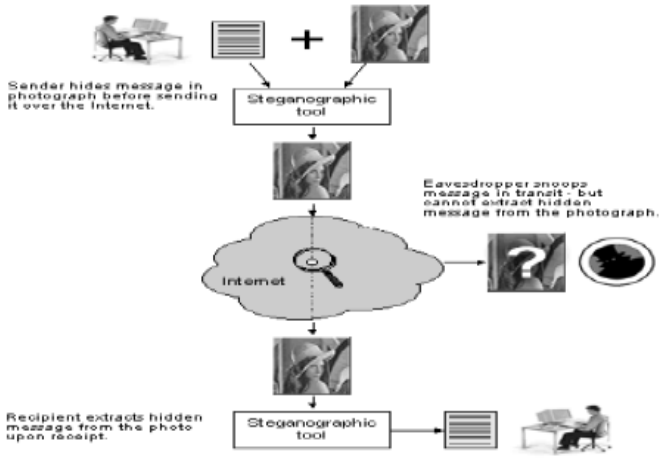

Figure 3 Steganography on the Internet

IV. STEGANOGRAPHY UNDER VARIOUS MEDIA

A. Steganography in Text:

The illegal distribution of documents through modern electronic means, such as electronic mail, means such as this allow infringers to make identical copies of documents without paying royalties or revenues to the original author. A method of marking printable documents with a unique codeword that is indiscernible to readers, but can be used to identify the intended recipient of a document just by examination of a recovered document. The techniques they propose are intended to be used in conjunction with standard security measures. For example, documents should still be encrypted prior to transmission across a network. Primarily, their techniques are ntended for use after a document has been decrypted, once it is readable to all.An added advantage of their system is that it is not prone to distortion by methods such as photocopying, and can thus be used to trace paper copies back to their source. Three features are described in the following subsections**:**

1. Line Shift Coding:

In this method, text lines are vertically shifted to encode the document uniquely. Encoding and decoding can generally be applied either to the format file of a document, or the bitmap of a page image. By moving every second line of document either 1/300 of an inch up or down, it was found that line-shift coding worked particularly well, and documents could still be completely decoded, even after the tenth photocopy. However, this method is probably the most visible text coding technique to the reader. Also, line-shift encoding can be defeated by manual or automatic measurement of the number of pixels between text baselines. Random or uniform respacing of the lines can damage any attempts to decode the codeword. However, if a

document is marked with line-shift coding, it is particularly difficult to remove the encoding if the document is in paper format. Each page will need to be rescanned, altered, and reprinted. This is complicated even further if the printed document is a photocopy, as it will then suffer from effects such as blurring, and salt-and-pepper noise.

2. **Word-Shift Coding:**

In word-shift coding, codewords are coded into a document by shifting the horizontal locations of words within text lines, while maintaining a natural spacing appearance. This encoding can also be applied to either the format file or the page image bitmap. The method, of course, is only applicable to documents with variable spacing between adjacent words, such as in documents that have been text-justified. As a result of this variable spacing, it is necessary to have the original image, or to at least know the spacing between words in the unencoded document. The following is a simple example of how word-shifting might work. For each text-line, the largest and smallest spaces between words are found. To code a line, the largest spacing is reduced by a certain amount, and the smallest is extended by the same amount. This maintains the line length, and produces little visible change to the text. Word-shift coding should be less visible to the reader than line-shift coding, since the spacing between adjacent words on a line is often shifted to support text justification.

B. Steganography in Images

1. Image Compression:

Image compression offers a solution to large image files. Two kinds of image compression are *lossless* and *lossy* compression. Both methods save storage space but have differing effects on any uncompressed hidden data in the image. Lossy compression, as typified by JPEG(Joint Photographic Experts Group) format files, offers high compression, but may not maintain the original image's integrity. This can impact negatively on any hidden data in the image. This is due to the lossy compression algorithm, which may ``lose" unnecessary image data, providing a close approximation to high-quality digital images, but not an exact duplicate. Hence, the term ``lossy" compression. Lossy compression is frequently used on true-color images, as it offers high compression rates. Lossless compression maintains the original image data exactly; hence it is preferred when the original information must remain intact. It is thus more favored by steganographic techniques. Unfortunately, lossless compression does not offer such high compression rates as lossy compression. Typical examples of lossless compression formats are CompuServe's GIF(Graphics Interchange Format) and Microsoft's BMP (Bitmap) format.

2. Image Encoding Techniques:

Information can be hidden many different ways in images. Straight message insertion can be done, which will simply encode every bit of information in the image. More complex encoding can be done to embed the message only in ``noisy'' areas of the image, that will attract less attention. The message may also be scattered randomly throughout the cover image. The most common approaches to information hiding in images are:

- Least significant bit (LSB) insertion
- Masking and filtering techniques
- Algorithms and transformations

2.1 Least Significant bit insertion:

The least significant bit insertion method is probably the most well known image steganography technique. A simple conversion from a GIF or BMP format to a lossy compression format such as JPEG can destroy the hidden information in the image. When applying LSB techniques to each byte of a 24-bit image, three bits can be encoded into each pixel. (As each pixel is represented by three bytes) Any changes in the pixel bits will be indiscernible to the human eye. For example, the letter A can be hidden in three pixels. Assume the original three pixels are represented by the three 24-bit words below:

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

The binary value for the letter A is (10000011). Inserting the binary value of A into the three pixels, starting from the top left byte, would result in:

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

The emphasized bits are the only bits that actually changed. The main advantage of LSB insertion is that data can be hidden in the least and second to least bits and still the human eye would be unable to notice it.When using LSB techniques on 8-bit images, more care needs to be taken, as 8-bit formats are not as forgiving to data changes as 24-bit formats are. Care needs to be taken in the selection of the cover image, so that changes to the data will not be visible in the stego-image. When modifying the LSB bits in 8-bit images, the pointers to entries in the palette are changed. It is important to remember that a change of even one bit could mean the difference between a shade of red and a shade of blue. Such a change would be immediately noticeable on the displayed image, and is thus unacceptable. For this reason, data-hiding experts recommend using grey-scale palettes, where the differences between shades are not as pronounced.

Alternatively, images consisting mostly of one color, such as the so-called Renoir palette, named
because it comes from a 256 color version of Renoir's ``Le Moulin de la Galette".

2.2 Masking and Filtering:

Masking and filtering techniques hide information by marking an image in a manner similar to paper watermarks. Because watermarking techniques are more integrated into the image, they may be applied without fear of image destruction from lossy compression. By covering, or masking a faint but perceptible signal with another to make the first non-perceptible, we exploit the fact that the human visual system cannot detect slight changes in certain temporal domains of the image. Technically, watermarking is not a steganographic form. Masking techniques are more suitable for use in lossy JPEG images than LSB insertion because of their relative immunity to image operations such as compression and cropping.

V. STEGANALYSIS

Whereas the goal of steganography is the avoidance of suspicion to hidden messages in other data, steganalysis aims to discover and render useless such covert messages. Hiding information within electronic media requires alterations of the media properties that may introduce some form of degradation or unusual characteristics. These characteristics may act as signatures that broadcast the existence of the embedded message, thus defeating the purpose of steganography. Attacks and analysis on hidden information may take several forms: detecting, extracting, and disabling or destroying hidden information. An attacker may also embed counter information over the existing hidden information. Here two methods are looked into: detecting messages or their transmission and disabling embedded information. These approaches (attacks)vary depending upon the methods used to embed the information in to the cover media Some amount of distortion and degradation may occur to carriers of hidden messages even though such distortions cannot be detected easily by the human perceptible system. This distortion may be anomalous to the "normal" carrier that when discovered may point to the existence of hidden information. Steganography tools vary in their approaches for hiding information. Without knowing which tool is used and which, if any, stego key is used; detecting the hidden information may become quite complex. However, some of the steganographic approaches have characteristics that act as signatures for the method or tool used.

VI. CONCLUSION

Steganography can be used for hidden communication. We have explored the limits of steganography theory and practice. We pointed out the enhancement of the image

steganographic system using LSB approach to provide a means of secure communication. A *stego-key* has been applied to the system during embedment of the message into the cover-image. In our proposed approach, the message bits are embedded randomly into the cover-image pixels instead of sequentially. Finally, we have shown that steganography that uses a key has a better security than non-key steganography. This is so because without the knowledge of the valid key, it is difficult for a third party or malicious people to recover the embedded message. However there are still some issues need to be tackled to implement LSB on a digital image as a *cover-object* using random pixels.

VII. REFERENCES

1. Cachin, "An Information-Theoretic Model for Steganography", in *proceeding 2nd Information Hiding Workshop*, vol. 1525.

2. D. Artz, "Digital Steganography: Hiding Data within Data", *IEEE Internet Computing*, pp. 75-80, May-Jun 2001.

3. E.T. Lin and E.J. Delp, "A Review of Data Hiding in Digital Images," *in Proceedings of the Image Processing, Image Quality, Image Capture Systems Conference*, PICS '99, Ed., Apr. 1999.

4. Anderson R.J. and Petitcolas F.A.P., "On the Limits of steganography," J. Selected Areas in Comm., vol. 16, no.4, 1998.

5. Bailey, K. and Curran, K. "An evaluation of image-based steganography methods". International Journal of Digital Evidence, Fall 2003.

6. Information Hiding: Steganography and Watermarking Attacks and Countermeasures (Advances in Information Security, Volume 1) Johnson, Neil F. / Doric, Zoran / Jajodia.

7. L. Zheng and I. Cox, JPEG based conditional entropy coding for correlated steganography, in: Proceedings of IEEE International Conference on Multimedia and Expo, Beijing, China, 2-5 July 2007.

8. Information hiding, watermarking and steganography, Public Seminar, Intelligent Systems Research Centre (ISRC), University of Ulster at Magee, Northern Ireland, 28th April 2009.

9. M. J. Atallah, V. Raskin, C. F. Hempelmann, M. Karahan, R. Sion, U. Topkara, and K. E. Triezenberg,"Natural language watermarking and tamperproofing," in Information Hiding: Fifth International Workshop, F. A. P. Petitcolas, ed., Lecture Notes in Computer Science 2578, pp. 196–212, Springer,October 2002.

10. K. Bennett, "Linguistic steganography: Survey, analysis, and robustness concerns for hiding information in text," Tech. Rep. TR 2004-13, Purdue CERIAS, May 2004.

11. M. Topkara, G. Riccardi, D. Hakkani-Tur, and M. J. Atallah, "Natural language watermarking: Challenges in building a practical system," in Proceedings of the SPIE International Conference on Security,Steganography, and Watermarking of Multimedia Contents, January 2006.

12. Grothoff, K. Grothoff, L. Alkhutova, R. Stutsman, and M. Atallah, "Translation-based steganography," in Proceedings of Information Hiding Workshop (IH 2005), pp. 213–233, Springer, 2005.