Steganography and Its Applications in Security

Ronak Doshi, [1] Pratik Jain, [2] Lalit Gupta[3]
1, 2,3Department of Electronics and telecommunication, Pune University, India

ABSTRACT: *Steganography is the dark cousin of cryptography, the use of codes. While cryptography provides privacy, steganography is intended to provide secrecy. Steganography is a method of covertly communicating. 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 organizations in repressive states to communicate their message to the outside world without their own government being aware of it. In this article we have tried to elucidate the different approaches towards implementation of Steganography using 'multimedia' file (text, static image, audio and video). Steganalysis is a newly emerging branch of data processing that seeks the identification of steganographic covers, and if possible message extraction. It is similar to cryptanalysis in cryptography. The technique is ancient emerging monster that have gained immutable notice as it have newly penetrated the world of digital communication security. Objective is not only to prevent the message being read but also to hide its existence.*

Keywords: *Carrier, Privacy, Secrecy, Steganalysis, Steganography*

I. INTRODUCTION

The word steganography is of Greek origin and means "covered, or hidden writing". It is the science of hiding information. Whereas the goal of cryptography is to make data unreadable by a third party, the goal of steganography is to hide the data from a third party. In steganography information can be hidden in carriers such as images, audio files, text files, and video and data transmissions. When message is hidden in the carrier a stego carrier is formed for example a stego-image. It will be perceived to be as close as possible to the original carrier or cover image by the human senses. Steganography and cryptography are closely related. Cryptography scrambles messages so they cannot be understood. Steganography, on the other hand, will hide the message so that there is no knowledge of the existence of the message in the first place. Steganography includes the concealment of information within computer files. In digital steganography, electronic communications may include steganographic coding inside of a transport layer, such as a document file, image file, program or protocol.[1]

Steganography today, however, is significantly more sophisticated, allowing a user to hide large amounts of information within image and audio files. These forms of steganography often are used in conjunction with cryptography so that the information is doubly protected; first it is encrypted and then hidden so that an adversary has to first find the information (an often difficult task in and of itself) and *then* decrypt it.

In this paper, a security thesis is proposed which imposes the concept of secrecy over privacy for messages in various formats.

II. HISTORY

Steganographic techniques have been used for centuries. Steganography has been widely used in historical times, especially before cryptographic systems were developed.

The first known application dates back to the ancient Greek times, when messengers tattooed messages on their shaved heads and then let their hair grow so the message remained unseen.

A different method from that time used wax tables as a cover source. Text was written on the underlying wood and the message was covered with a new wax layer. The tablets appeared to be blank so they passed inspection without question.

During World War 2 invisible ink was used to write information on pieces of paper so that the paper appeared to the average person as just being blank pieces of paper. Liquids such as urine, milk, vinegar and fruit juices were used, because when each one of these substances is heated they darken and become visible to the human eye.

Another clever invention in *Steganographia* was the "Ave Maria" cipher. The book contains a series of tables, each of which has a list of words, one per letter. To code a message, the message letters are replaced by the corresponding words. If the tables are used in order, one table per letter, then the coded message will appear to be an innocent prayer.

All of these approaches to steganography have one thing in common -- they hide the secret message in the physical object which is sent. The cover message is merely a distraction, and could be anything. Of the innumerable variations on this theme, none will work for electronic communications because only the pure information of the cover message is transmitted. Nevertheless, there is plenty of room to hide secret information in a not-so-secret message. It just takes ingenuity.

III. IMPLEMENTATION OF STEGANOGRAPHY

Secrets can be hidden inside all sorts of cover information. The following formula provides a very generic description of the pieces of the steganographic process:

$$\text{cover\_medium} + \text{hidden\_data} + \text{stego\_key} = \text{stego\_medium} \qquad (1)$$

In this context, the cover_medium is the file in which we will hide the hidden_data, which may also be encrypted using the stego_key. The resultant file is the stego_medium (which will, of course be the same type of file as the cover_medium).
There are four ways to implement steganography:

1. Using text.
2. Using images.
3. Using audio files.
4. Using video files.[2]

3.1 Text Steganography:

Text steganography can be classified in three basic categories **-** format-based, random and statistical generation and linguistic method. Format-based methods used physical text formatting of text as a place in which to hide information. Generally, this method modifies existing text in order to hide the steganographic text. Insertion of spaces, deliberate misspellings distributed throughout the text, resizing the fonts are some of the many format-based methods being used in text steganography. However, Bennett has stated that those format-based methods managed to trick most of the human eyes but it cannot trick once computer systems have been used. Random and statistical generation is generating cover text according to the statistical properties. This method is based on character sequences and words sequences. The hiding of information within character sequences is embedding the information to be appeared in random sequence of characters. This sequence must appear to be random to anyone who intercepts the message.[5]

A second approach for character generation is to take the statistical properties of word-length and letter frequency in order to create "words" (without lexical value) which will appear to have the same statistical properties as actual words in a given language. The hiding of information within word sequences, the actual dictionary items can be used to encode one or more bits of information per word using a codebook of mappings between lexical items and bit sequences, or words themselves can encode the hidden information.

The final category is linguistic method which specifically considers the linguistic properties of generated and modified text, frequently uses linguistic structure as a place for hidden messages. In fact, steganographic data can be hidden within the syntactic structure itself.
Example:

Sender sends a series of integer number (Key) to the recipient with a prior agreement that the secret message is hidden within the respective position of subsequent words of the cover text. For example the series is '1, 1, 2, 3, 4, 2, 4'**and the cover text is "A team of five men joined today".** So the hidden message is **"Atfvoa".** A "0" in the number series will indicate a blank space in the recovered message. The word in the received cover text will be skipped if the number of characters in that word is less than the respective number in the series (Key) which shall also be skipped during the process of message unhide.

3.2 Image Steganography:

The most widely used technique today is hiding of secret messages into a digital image. This steganography technique exploits the weakness of the human visual system (HVS). HVS cannot detect the variation in luminance of color vectors at collection of color pixels. The individual pixels can be represented by their optical higher frequency side of the visual spectrum. A picture can be represented by a characteristics like 'brightness', 'chroma' etc. Each of these characteristics can be digitally expressed in terms of 1s and 0s.

For example: a 24-bit bitmap will have 8 bits, representing each of the three color values (red, green, and blue) at each pixel. If we consider just the blue there will be 2 different values of blue. The difference between 11111111 and 11111110 in the value for blue intensity is likely to be undetectable by the human eye. Hence, if the terminal recipient of the data is nothing but human visual system (HVS) then the Least Significant Bit (LSB) can be used for something else other than color information.

3.2.1 LSB Coding

The simplest approach to hiding data within an image file is called least significant bit (LSB) insertion. In this method, we can take the binary representation of the hidden_data and overwrite the LSB of each byte within the cover_image. If we are using 24-bit color, the amount of change will be minimal and indiscernible to the human eye. As an example, suppose that we have three adjacent pixels (nine bytes) with the following RGB encoding:

10010101	00001101	11001001
10010110	00001111	11001010
10011111	00010000	11001011

Now suppose we want to "hide" the following 9 bits of data (the hidden data is usually compressed prior to being hidden): 101101101. If we overlay these 9 bits over the LSB of the 9 bytes above, we get the following (where bits in **bold** have been changed):

10010101	00001100	11001001
10010111	00001110	11001011
10011111	00010000	11001011

Note that we have successfully hidden 9 bits but at a cost of only changing 4, or roughly 50%, of the LSBs.

Fig. 1.Original image

Fig. 2. Embedded image

Example of still imagery steganography. Left hand side image is the original cover image, whereas right hand side does embedding a text file into the cover image make the stego image

3.2.2 Masking And Filtering:

Masking and filtering techniques are mostly used on 24 bit and grey scale images. They hide information in a way similar to watermarks on actual paper and are sometimes used as digital watermarks. Masking images entails changing the luminance of the masked area. The smaller the luminance change, the less of a chance that it can be detected.

Masking is more robust than LSB insertion with respect to compression, cropping, and some image processing. Masking techniques embed information in significant areas so that the hidden message is more integral to the cover image than just hiding it in the "noise" level. This makes it more suitable than LSB with, for instance, lossy JPEG images.

3.3 Audio Steganography:

Steganography, in general, relies on the imperfection of the human auditory and visual systems. Audio steganography takes advantage of the psychoacoustical masking phenomenon of the human auditory system [HAS]. Psychoacoustical or auditory masking property renders a weak tone imperceptible in the presence of a strong tone in its temporal or spectral neighborhood. This property arises because of the low differential range of the HAS even though the dynamic range covers 80 dB below ambient level. Frequency masking occurs when human ear cannot perceive frequencies at lower power level if these frequencies are present in the vicinity of tone- or noise-like frequencies at higher level.

Additionally, a weak pure tone is masked by wide-band noise if the tone occurs within a critical band. This property of inaudibility of weaker sounds is used in different ways for embedding information. Embedding of data by inserting inaudible tones in cover audio signal has been presented recently. [3]

In audio steganography, secret message is embedded into digitized audio signal which result slight altering of binary sequence of the corresponding audio file. The list of methods that are commonly used for audio steganography are listed and discussed below.

- LSB coding
- Parity coding
- Phase coding
- Spread spectrum
- Echo hiding

3.3.1 LSB Coding:

Sampling technique followed by Quantization converts analog audio signal to digital binary sequence.

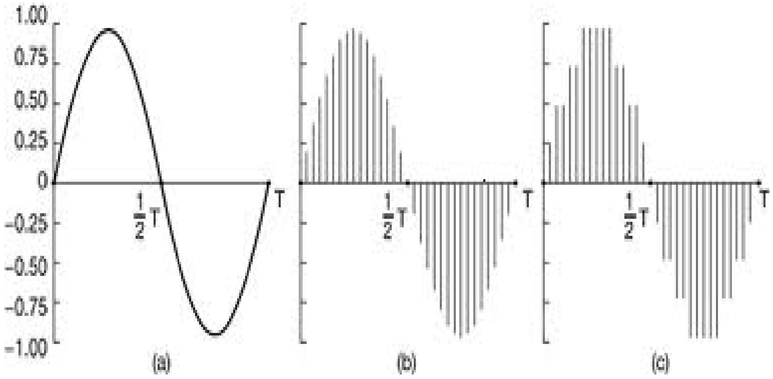

Fig. 3. Sampling of the Sine Wave followed by Quantization process.

In this technique LSB of binary sequence of each sample of digitized audio file is replaced with binary equivalent of secret message.

3.3.2 Parity Coding:

Instead of breaking a signal down into individual samples, the parity coding method breaks a signal down into separate regions of samples and encodes each bit from the secret message in a sample region's parity bit. If the parity bit of a selected region does not match the secret bit to be encoded, the process flips the LSB of one of the samples in the region. Thus, the sender has more of a choice in encoding the secret bit, and the signal can be changed in a more unobtrusive fashion.

3.3.3 Phase coding:

Human Auditory System (HAS) can't recognize the phase change in audio signal as easy it can recognize noise in the signal. The phase coding method exploits this fact. This technique encodes the secret message bits as phase shifts in the phase spectrum of a digital signal, achieving an inaudible encoding in terms of signal-to- noise ratio.

3.3.4 Spread spectrum:

In the context of audio steganography, the basic spread spectrum (SS) method attempts to spread secret information across the audio signal's frequency spectrum as much as possible. This is analogous to a system using an implementation of the LSB coding that randomly spreads the message bits over the entire sound file. However, unlike LSB coding, the SS method spreads the secret message over the sound file's frequency spectrum, using a code that is independent of the actual signal. As a result, the final signal occupies a bandwidth in excess of what is actually required for transmission.

3.3.5 Echo Hiding:

In echo hiding, information is embedded in a sound file by introducing an echo into the discrete signal. Like the spread spectrum method, it too provides advantages in that it allows for a high data transmission rate and provides superior robustness when compared to the noise inducing methods. If only one echo was produced from the original signal, only one bit of information could be encoded. Therefore, the original signal is broken down into blocks before the encoding process begins. Once the encoding process is completed, the blocks are concatenated back together to create the final signal.

3.4 Video Steganography:

Video files are generally a collection of images and sounds, so most of the presented techniques on images and audio can be applied to video files too. When information is hidden inside video the program or person hiding the information will usually use the DCT (Discrete Cosine Transform) method. DCT works by slightly changing each of the images in the video, only so much that it is not noticeable by the human eye. To be more precise about how DCT works, DCT alters values of certain parts of the images, it usually rounds them up. For example, if part of an image has a value of 6.667 it will round it up to 7. [6]

The great advantages of video are the large amount of data that can be hidden inside and the fact that it is a moving stream of images and sounds. Therefore, any small but otherwise noticeable distortions might go by unobserved by humans because of the continuous flow of information.

IV. STEGANALYSIS

The art of detecting Steganography is referred to as Steganalysis. Steganalysis is the process of identifying steganography by inspecting various parameter of a stego media. The primary step of this process is to identify a suspected stego media. After that steganalysis process determines whether that media contains hidden message or not and then try to recover the message from it. In cryptanalysis, it is clear that the intercepted message is encrypted and it certainly contains the hidden message because the message is scrambled. But in the case of steganalysis this may not be true. The suspected media may or may not be with hidden message. The steganalysis process starts with set of suspected information streams. Then the set is reduced with the help of advance statistical methods.[4]

In the case of Visual detection steganalysis technique, a set of stego images are compared with original cover images and note the visible difference. Signature of the hidden message can be derived by comparing numerous images. Cropping or padding of image also is a visual clue of hidden message because some stego tool is cropping or padding blank spaces to fit the stego image into fixed size. Difference in file size between cover image and stego images, increase or decrease of unique colors in stego images can also be used in the Visual Detection steganalysis technique.

Scientists and researchers are trying new methods to try and discover ways of detecting hidden files and rendering them useless. The U. S. Government has contracted Wetstone Technologies to work with the U.S. Air Force to research algorithms that can be used to discover embedded files in digital, audio and video format. Steganalysis is the technique to detect steganography or defeat steganography. The research to device strong steganographic and steganalysis technique is a continuous process.

V. APPLICATIONS

Steganography can be used anytime you want to hide data. There are many reasons to hide data but they all boil down to the desire to prevent unauthorized persons from becoming aware of the existence of a message. With these new techniques, a hidden message is indistinguishable from white noise. Even if the message is suspected, there is no proof of its existence. In the business world steganography can be used to hide a secret chemical formula or plans for a new invention.

Steganography can also be used for corporate espionage by sending out trade secrets without anyone at the company being any the wiser. [7]

Terrorists can also use steganography to keep their communications secret and to coordinate attacks. All of this sounds fairly nefarious, and in fact the obvious uses of steganography are for things like espionage. But there are a number of peaceful applications. The simplest and oldest are used in map making, where cartographers sometimes add a tiny fictional street to their maps, allowing them to prosecute copycats. A similar trick is to add fictional names to mailing lists as a check against unauthorized resellers.

Most of the newer applications use steganography like a watermark, to protect a copyright on information. Photo collections, sold on CD, often have hidden messages in the photos which allow detection of unauthorized use. The same technique applied to DVDs is even more effective, since the industry builds DVD recorders to detect and disallow copying of protected DVDs.

VI. FUTURE

In today's world, we often listen a popular term "Hacking". Hacking is nothing but an unauthorized access of data which can be collected at the time of data transmission. With respect to steganography this problem is often taken as Steganalysis.

Steganalysis is a process in which a steganalyzer cracks the cover object to get the hidden data. So, whatever be the technique will be developed in future, degree of security related with that has to be kept in mind. It is hoped that Dual Steganography, Steganography along with Cryptography may be some of the future solution for this above mentioned problem.

VII. CONCLUSION

Steganography is a fascinating and effective method of hiding data that has been used throughout history. Methods that can be employed to uncover such devious tactics, but the first step are awareness that such methods even exist. There are many good reasons as well to use this type of data hiding, including watermarking or a more secure central storage method for such things as passwords, or key processes. Regardless, the technology is easy to use and difficult to detect. The more that you know about its features and functionality, the more ahead you will be in the game.

In this paper, different techniques are discussed for embedding data in text, image, and audio/video signals as cover media. I have presented a brief overview of a very exciting and fast paced area of computer security. This technology has many in the security field worried as the possible harm that may be done to both government and private industries. As pc's become more powerful this technology will grow substantially and become much more main stream. There are already hundreds of steganography programs available that can be used on text, audio and graphic files. The government and many private companies are researching ways to best detect the use of steganography on files. As steganalysis becomes more mature it will be implemented as a standard security tool the way firewalls, virus detection software and intrusion detection programs currently are.

REFERENCES

Books:

[1] Compression Algorithms for Real Programmers, Wayner Peter. 2000
[2] Disappearing Cryptography: Being and Nothingness on the Net, Wayner Peter. 1996
[3] Secure Steganography for Audio Signals

Journal Papers:

[4] Steganography and steganalysis-Robert Krenn, Internet Publication, March 2004 http://www.krenn.nl/univ/cry/steg/article.pdf
[5] Steganography, http://searchsecurity.techtarget.com/sDefinition/0,,sid14_gci213717,00.html
[6] Johnson, Neil F., "Steganography", 2000 http://www.jjtc.com/stegdoc/index2.html
[7] The WEPIN Store, "Steganography (Hidden Writing)", 1995, http://www.wepin.com/pgp/stego.html