Different Medias of Steganography- An Emerging Field of Network Security

M. Indra Sena Reddy^{#1}, M. Purushotham Reddy^{*2}, K. Subba Reddy^{#3}

^{1,3} R. G. M. College of Enginereing & Technology, Nandyal, A.P, India. ² VBIT College of Engineering, Proddatur.

Abstract— Steganography, literally meaning "secret writing", involves hiding a data file in another innocuous-looking file. From the time of Herodotus in Greece, to the defense mechanisms of today, steganography has been used to deny one's adversaries the knowledge of message traffic.

Steganography takes one piece of information and hides it within another. A "container file" holds the secret message in it, in such a way that the existence of the message cannot be suspected. This could be done in several possible ways, with newer methods being discovered with each passing day.

In this paper, a detailed analysis of steganography is made. The history of steganography is briefly dealt with. How steganography works is examined, keeping in mind Bender's specifications.

Data hiding is implemented in three different media; text, audio and image files. Each offers challenges and solutions to these challenges are analysed. How they is implemented using steganographic tools is also seen. The main characteristics of steganographic software are discussed, together with the various forms of steganographic methods. Steganalysis, the science of detecting steganography is touched upon. The weaknesses of steganography are also described, together with measures for improvement. The paper concludes by taking a look at the potential of steganography and the changes it can bring about as the future of network security.

Keywords— steganography, Steganographic tools, Network security.

I. INTRODUCTION

Due to advances in ICT, most of information is kept electronically. Consequently, the security of information has become a fundamental issue. Besides cryptography, steganography can be employed to secure information. Steganography is a technique of hiding information in digital media. In contrast to cryptography, the message or encrypted message is embedded in a digital host before passing it through the network, thus the existence of the message is unknown. Besides hiding data for confidentiality, this approach of information hiding can be extended to copyright protection for digital media: audio, video, and images.

The growing possibilities of modern communications need the special means of security especially on computer network. The network security is becoming more important as the number of data being exchanged on the Internet increases. Therefore, the confidentiality and data integrity are requires to protect against unauthorized access and use. This has resulted in an explosive growth of the field of information hiding. In addition, the rapid growth of publishing and broadcasting technology also require an alternative solution in hiding information. The copyright such as audio, video and other source available in digital form may lead to large-scale unauthorized copying. This is because the digital formats make possible to provide high image quality even under multi-

copying. Therefore, the special part of invisible information is fixed in every image that could not be easily extracted without specialized technique saving image quality simultaneously [1]. All this is of great concern to the music, film, book and software publishing industries.

Information hiding is an emerging research area, which encompasses applications such as copyright protection for digital media, watermarking, fingerprinting, and steganography [2]. All these applications of information hiding are quite diverse [2].

- (i) In watermarking applications, the message contains information such as owner identification and a digital time stamp, which usually applied for copyright protection.
- (ii) Fingerprint, the owner of the data set embeds a serial number that uniquely identifies the user of the data set. This adds to copyright information to makes it possible to trace any unauthorized used of the data set back to the user.
- (iii) Steganography hide the secret message within the host data set and presence imperceptible.

In those applications, information is hidden within a host data set and is to be reliably communicated to a receiver. The host data set is purposely corrupted, but in a covert way, designed to be invisible to an informal analysis. However, this paper will only focus on information hiding using steganography approach. In this section, we give an overview about steganography in detail in order to avoid confusion with cryptography. The introduction of steganography is usually given as a synonym for cryptography but it is not normally used in other way. The section also discusses several information hiding methods useable for steganographic communication. some design issues and comparative studies of the methods employed in steganography are discussed in the paper. The survey also includes the limitations imposed by the technique on a range of steganography applications. Finally, section 4 will outline the summary of the overall information hiding technique using steganography in order to guarantee the confidentiality and data integrity.

II. OVERVIEW STEGANOGRAPHY

The word steganography comes from the Greek *Steganos*, which mean covered or secret and -graphy mean writing or drawing. Therefore, steganography means, literally, covered writing. Steganography is the art and science of hiding information such that its presence cannot be detected [3] and a communication is happening [4, 5]. A secret information is encoding in a manner such that the very existence of the information is concealed. Paired with existing communication

methods, steganography can be used to carry out hidden exchanges.

The main goal of steganography is to communicate securely in a completely undetectable manner [6] and to avoid drawing suspicion to the transmission of a hidden data [7]. It is not to keep others from knowing the hidden information, but it is to keep others from thinking that the information even exists. If a steganography method causes someone to suspect the carrier medium, then the method has failed [8]. Until recently, information hiding techniques received very much less attention from the research community and from industry than cryptography. This situation is, however, changing rapidly and the first academic conference on this topic was organized in 1996. There has been a rapid growth of interest in steganography for two main reasons [9]:

- (i) The publishing and broadcasting industries have become interested in techniques for hiding encrypted copyright marks and serial numbers in digital films, audio recordings, books and multimedia products.
- (ii) Moves by various governments to restrict the availability of encryption services have motivated people to study methods by which private messages can be embedded in seemingly innocuous cover messages.

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 [10]. 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 *coverobject*. 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.

There are several suitable carriers below to be the *cover-object* [11]:

(i) Network Protocols such as TCP, IP and UDP

- (ii) Audio that using digital audio formats such as wav, midi, avi, mpeg, mpi and voc
- (iii) File and Disk that can hides and append files by using the slack space
- (iv) Text such as null characters, just alike morse code including html and java
- (v) 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 [12, 13].

- (i) 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*.
- (ii) 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. Steganography vs. Cryptography

Basically, the purpose of cryptography and steganography is to provide secret communication. However, steganography is not the same as cryptography. Cryptography hides the contents of a secret message from a malicious people, whereas steganography even conceals the existence of the message. Steganography must not be confused with cryptography, where we transform the message so as to make it meaning obscure to a malicious people who intercept it. Therefore, the definition of breaking the system is different [14]. 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.

In cryptography, the structure of a message is scrambled to make it meaningless and unintelligible unless the decryption key is available. It makes no attempt to disguise or hide the encoded message. Basically, cryptography offers the ability of transmitting information between persons in a way that prevents a third party from reading it. Cryptography can also provide authentication for verifying the identity of someone or something. In contras, steganography does not alter the structure of the secret message, but hides it inside a coverimage so it cannot be seen. A message in ciphertext, for instance, might arouse suspicion on the part of the recipient while an "invisible" message created with steganographic methods will not. In other word, steganography prevents an unintended recipient from suspecting that the data exists. In addition, the security of classical steganography system relies on secrecy of the data encoding system [12]. Once the encoding system is known, the steganography system is defeated.

It is possible to combine the techniques by encrypting message using cryptography and then hiding the encrypted message using steganography. The resulting *stego-image* can be transmitted without revealing that secret information is being exchanged. Furthermore, even if an attacker were to defeat the steganographic technique and detect the message from the *stego-object*, he would still require the cryptographic decoding key to decipher the encrypted message [10].

III. HOW STEGANOGRAPHY WORKS

Hidden data is often encrypted. This is done because of the Kerckhoff principle, which states 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 key. Most of steganographic techniques meet this principle.

Computer files (images, sound recordings, etc) contain unused or insignificant areas of data. Steganography takes advantage of these areas, replacing them with information that may include encrypted mail, functions, etc. The files can then be encrypted without anyone really knowing about what lies within them.

The advantage of steganography is that it can be used to secretly transmit messages without the fact of the transmission being discovered. Using encryption can identify the sender and the receiver. Thus, steganography has a double layer of protection: first, the file itself is hidden and second, the data in it is encrypted.

A. RULES OF STEGANOGRAPHY

- The data in the container should not be significantly degraded by the embedded data.
- Embedded data should be as imperceptible as possible.
- Embedded data should directly be encoded into the media, to maintain data consistency.
- Embedded data should be as immune as possible to modification, manipulation and attacks.
- Error correcting codes should be used in the embedded data.
- The embedded data should be self-clocking.
- Even if only parts of the cover data is available, the entire embedded data should be recoverable



Fig.1 Technical methods of steganography

Most of steganography is used in protection against detection. This can be done by hiding info in user data or volatile data. The latter model is called data hiding in network model architecture. For example, in the OSI reference model, data is sent through packets. Covert channels can be established using the control data to send info that is hidden. At the receiver, the information is stripped off. Information files stay on the hard drive unless specifically deleted. Data hiding is done by compressing data into a message. There is an innocent 'container file' that will hold the message and a secret 'message file' that holds the information to be hidden. Compression is done by one of 2 methods--lossless and lossy techniques. Lossless uses GMP or BMP format. Here, the receiver reconstructs the original information exactly. Lossy is used by JPEG. It saves space, but may not maintain 100% message integrity.

B. WATERMARKING

Steganography can be used to place a hidden trademark in music, images and software using a technique called "watermarking". Watermarking techniques are more integrated into the image, so they can be applied without fear of destruction due to lossy compression. Watermarking extends image information and becomes an attribute of the cover information, providing copyright details.

IV. STEGANOGRAPHY IN DIFFERENT MEDIA.

Steganography is used to conceal files in various forms of data. This is done in three different media: text, images and audio signals. Steganography can and is being used widely in these media.

A. STEGANOGRAPHY IN TEXT

Steganography in text is done through three different techniques: line-shift coding, word-shift coding and feature coding.

1) Line-shift coding:

Features are stamped into the text by shifting the lines, so as to make identification harder. Text lines are vertically shifted for encoding. This can be applied to either a format file or to the bitmap of a page. By moving every second line of the document 1/300th of an inch up or down, line-shift encoding ensures that documents can be encoded and decoded properly. A couple of disadvantages include the fact that it is difficult to remove encoding and also that it is the most visible form of coding.

2) Word-shift coding:

Here, word spacing is done in a fixed pattern, so that even the fact that a message is present is unknown. Words are coded into a document by shifting the horizontal locations of words within text lines, while maintaining a natural spacing appearance. One example of a code: 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 increased by the same amount. Thus, line length is maintained and change is not visible. But this method is not foolproof either, as if the spaces between words are measured, the encoded data would be revealed.

3). Feature coding:

Characteristics of the letters themselves are altered. The specific ways in which certain letters are written forms a code by itself. For example, the endlines of d,h,b etc. could be modified. This is largely indiscernible and can be directly used on the image file. The only way in which this code can be attacked is by adjusting each endline to a fixed value. However, this is time-consuming and painstaking and is usually not done.

Other interesting alternative methods are available, but are still in their infant stages. These include:

- Using grammar to encode data.
- Syntactic encoding.
- Semantic encoding.

Software to hide data in text (textual steganography) comes in various versions like *Texthide*, *Webstego* and *Steganos*. All these allow the communication of encrypted sensitive data through plain text files in a format not identifiable by a third party.

B. IMAGE STEGANOGRAPHY:

Image steganography has truly advanced with the invention of fast, powerful computers. Software is easily available for processing and hiding of data images. Images can also be retrieved very easily. There are three main methods of information hiding in images. These are:

1) Least Significant Bit Insertion:

This is the most well-known image steganography technique. It is simple, easy to create and also easy to apply. Unfortunately, it is extremely vulnerable to attack. A simple conversion of formats can destroy all hidden information.

In this technique, in each byte of a 24-bit image, we can encode three bits in each pixel. Any change in pixel bytes is indiscernible to the eye. For example:

8-bit images are not as easy to convert as 24-bit ones, as a small change in bit values can outright lead to a change in colour. Thus, care needs to be taken here. The examples of an image steganography as shown in figure 2-4.



Fig 2.Image without embedded picture



Fig 3.Image with embedded steganographic picture.



Fig4. Embedded picture in image 2.

2) Masking and filtering:

Masking and filtering hide information by marking an image in a manner similar to paper watermarks. By masking a faint image with another in order to make the first non-perceptible, we exploit the fact that the human eye cannot detect faint changes in a visual image. Masking techniques are more suitable for use in lossy JPEG images than in LSB insertion because of their relative immunity to compression and cropping.

3) Algorithms and transformation:

JPEG images are very popular, as they provide high quality pictures and the ability to hide information in them. JPEG uses the discrete cosine transform (DCT) to achieve compression. DCT is a lossy transform, as cosine values cannot be calculated exactly and rounding errors often take place. Variances in data depend on the methods and values used. Images can also be processed using Fourier transformations and wavelet transformations. Spread spectrum is also used, as a narrow bandwidth may be spread over a larger one in such a way that the signal's spectral density looks like noise. Directsequence and frequency-hopping spread-spectrum techniques are used. Direct-sequencing operates by phase-modulating the data with a pseudorandom number sequence that both the sender and receiver are aware of. Frequency-hopping deals with dividing the bandwidth into multiple channels and then hopping between them.

Other techniques encrypt and scatter the message throughout the image in some pre-determined manner. It is assumed that even if the message is discovered, it is useless without the algorithm and the key. Unfortunately, none of these algorithms are immune to destruction of data through image manipulation.

C. STEGANOGRAPHY IN AUDIO:

This is a very risky and challenging approach, as the human auditory system can detect even very minor changes in sound in a wide range. Random noise can be sensed easily. An audio environment is determined by two considerations—one, its digital representation and two, its transmission media. <u>Digital</u> <u>audio files</u> have two main characteristics:

1). Sample quantisation rate: This is a 16-bit linear quantisation and represents high-quality digital audio, such as those used by WAV files.

2). *Temporal sampling rate*: This puts an upper bound on the usable portion of the frequency range. The most popular ones include 8 kHz, 9.6 kHz and so on, up to 44.1 kHz.

The <u>transmission medium</u> of an audio signal refers to the environments the signal might go through, on its way from encoder to decoder. This could be digital *end-end*, *analog transmission*, *increasing-decreasing resampling* or "over-theair" environments.

Audio data hiding (steganography in audio) uses these considerations to present an effective way of hiding data. The four primary methods are:

- ◆ <u>Low-bit encoding</u>: Binary data can be stored in the Least Significant Bits of the sound files (similar to the image files). For example, channel capacity is 1kb per second per Hz. Therefore, if we have a 8kHz sequence, the capacity is 8kbps. But this method introduces audible noise. this has very poor immunity to manipulation. Factors like resampling and channel noise can easily damage the signal. However, if the amplitude is slightly modified, such that it does not produce any perceptible difference, the implementation offers high robustness to MPEG compression and other forms of signal manipulation like filtering, resampling and requantization.
- Phase coding: This works by substituting the phase of an audio segment with a reference phase that represents data. Here, the original sound sequence is broken up into a series of N short segments. A DFT (discrete Fourier Transform) is applied to each segment and the phase difference is calculated. New phase frames are created for all segments. The phase and original magnitude are combined to create a new segment. All new segments are concatenated for the desired encoded output. At the receiver end, the segment length and DFT are known and the values are extracted.
- <u>Spread spectrum:</u> The encoded data is spread as much as possible over the frequency spectrum. In Direct Sequence

Spread Spectrum, the signal is spread by multiplying it by a certain maximal length pseudorandom sequence, called a chip. The sampling rate for the host signal is used as the chip rate for coding. The start and end quanta for the phase locking purpose is taken care of by the discrete, sampled nature of the host signal. A higher chip rate leads to higher amount of associated data. The only negative factor here is random noise, introduced by the DSSS.

- Echo data hiding: Echo data hiding embeds data into a signal by using an echo. The data is hidden by varying three parameters of the echo: initial amplitude, decay rate and offset or delay. As the offset increases, the siganl and its echo blend. At a certain point, the human ear cannot distinguish between the two and the echo is heard as added resonance. By using two different delay times, both below the human audible level, we can encode a binary one or zero. The signal is divided into smaller bits, each of which is echoed to encode the desired bit. The final echoed signal is a recombination of all independent echoed portions. This signal works exceptionally well and is the strongest code to date among audio files.
- ٠

V. STEGANOGRAPHY TOOLS (S-TOOLS)

S-Tools (steganography tools) are used to conceal files within various forms of data. These are the tools required to implement steganography. We can work with steganography in sound and pictures with the help of S-Tools. Using S-Tools, we can hide multiple files in one object. The files are first individually compressed and stored with their names. Then, S-Tools may precede the stored information with some random garbage , so as to make decryption difficult. After this, the sender chooses a "passphrase", which is the key to the decryption. According to the passphrase, the whole lot is encrypted. All encryption works in what is called 'Cipher Feedback Mode (CFB)'.

S-Tools are used mainly in digital data, like a scanned image, or sampled sound. Now, digitally sampled sound has the advantage of not needing to be perfectly accurate; using this property, we can hide data in such a way that the change is inaudible to the human ear. Unless the passphrase is present, in spite of knowing a code may be involved, an enemy cannot decrypt it.

♦ <u>S-Tools in sound</u>: Sound samples in Windows WAV files can be either 8 bits (range 0-255) or 16 bits (range 0-65535). S-Tools distribute the bit pattern of the coded message file across the Least Significant Bits of the sample.

VI. CHARACTERISTICS OF STEGANOGRAPHIC SOFTWARE

Steganographic software enables information to be hidden in graphic, sound and apparently blank media. Examples include data sent through images and pictures. Work that is in progress on 24-bit images involves the following: Let us assume the resolution is 1024 * 768. The file that results will be of the size 1024 * 768 * 24/8. This is clearly greater than 2 MB. The RGB triples use 3 bytes per pixel. When we put information into the 24-bit image, the container file looks innocuous and the human eye cannot distinguish the embedded message file. This saves us from problems of file compression, where the storage of information may be interfered with.

Image steganography is most effectively handled by JPEG software. This means the source code is provided in order to compile the code at various platforms. JPEG uses lossy encoding to compress its data. JFIF files are used for output. JFIF consist of both lossy and lossless stages. The information to be passed is hidden between these stages. File compression in JPEG is its greatest advantage. Large images in unlimited colours can be stored in relatively small files. Another example could involve data sent through sound or audio files. Various steganographic software packages available in the market are very recent and include Hide-&-Seek, StegosDos, White Noise, etc. in all versions; the messages are encrypted before being embedded, in order to provide an increased layer of protection.

VII. FORMS OF STEGANOGRAPHY:

Many forms of steganography were devised and implemented. This includes methods like *blindside*, *S-Mail* and *Scramdisk*.

- <u>A)</u> <u>Blindside:</u> This is an application of steganography that allows one to conceal a file or a set of files within a standard computer image. This involves some very easy steps to store the data file, but is subject to the limitation that images only up to 50K can be slipped through. Encrypted passwords are used for authorisation to access data.
- B) <u>S-Mail:</u> This encrypts any data in a very difficult-todecrypt kind of way and then hides it in EXE or DLL files. The EXE file is then sent through the internet, via e-mail, to the recipient.
- <u>C)</u> <u>Scramdisk:</u> this allows the creation and use of a virtual encrypted data drive. On an existing hard drive, first an encrypted password is entered and data files are stored in the virtual drive. The recipient needs to first access the hard drive with the correct passphrase, without which the drive is inaccessible, and then the data can be extracted.

VII. STEGANALYSIS

This is the science of detecting hidden messages. A rising field today, it aims to discover and render useless all covert messages. A public watermark detector has been developed as an oracle to estimate a secret spread watermark. The image is first degraded and then random signals are added to completely wash out the watermark. Steganalysis is getting more and more advanced, in an effort to combat steganography.

A). WEAKNESSES OF STEGANOGRAPHY:

- Steganography is not without its disadvantages. However, these can be corrected and once implemented; it can strengthen the core of steganography.
- Most data hiding methods take advantage of human perceptual weaknesses, but they have weaknesses of their own. However, these can be individually rectified.
- One major drawback of steganography is that, unlike cryptography, it requires a lot of overhead to hide relatively few bits of data. Also, once the steganographic system is discovered, it is rendered useless. However, it fares no worse than cryptography and is still the preferred medium.

IX. CONCLUSION

Steganography was in the news lately, as an advanced means of secret communication. It has its obvious advantages, as an almost unbreakable system, and is complemented by cryptography. It can slip important communication without anyone knowing. Soon, we can have artists, musicians and authors using steganography to fight piracy. It can be used to track infringement of copyrights in a digital medium and can work wonders on the internet.

Steganography, if fallen into wrong hands, can create tremendous damage. It was in the news lately, as being the form by which Osama Bin Laden communicated with his associates, via Al-Jazeera television images. The US government recently released a public statement, declaring that "terrorist organisations are hiding maps, photographs of their targets and instructions for terrorist activities on chat boards and other rooms, bulletin websites using steganography". The famous copyright infringement case against Napster, the online music website, was filed after using steganographic methods. Currently in the news, steganography is finding increasing uses. It can be used to protect copyrights, prevent piracy and work in the transfer of top-secret data from place to place. Once its relatively minor disadvantages are rectified, steganography will be found to have amazing potential in the days to come.

REFERENCES

[1] N. Provos, "Probabilistic Methods for Improving Information Hiding", CITI Technical Report 01-1, January 31, 2001.

[2] R A Isbell, "Steganography: Hidden Menace or Hidden Saviour", Steganography White Paper, 10 May 2002.

[3] M. Ramkumar & A.N. Akansu. "Some Design Issues For Robust Data hiding Systems", http://citeseer.nj.nec.com/404009.html

[4] N.F. Johnson, S. Jajodia, "Staganalysis: The Investigation of Hiding Information", IEEE, pp. 113-116, 1998.

[5] R.Popa,"An Analysis of Steganographic

System", The "Politehnica" University of Timisoara, Faculty of Automatics and Computers, Department of Computer Science and Software Engineering, May 25, 1998.

[6] N.F. Johnson & S. Jajodia, "Steganalysis of Images Created Using Current Steganography Software", in Proceeding for the Second Information Hiding Workshop, Portland Oregon, USA, April 1998, pp. 273-289.[7] N.F. Johnson and S. Jajodia, "Exploring Steganography: Seeing the

Unseen", IEEE, pp. 26-34, 1998.

[8] N. Provos, P. Honeyman, "Detecting Steganography Content on the Internet". CITI Technical Report 01-11, 2001.

[9] R.J. Anderson, F.A.P. Petitcolas, "On The Limits of Steganography", IEEE Journal of Selected Area in Communications, pp. 474-481, May 1998.

[10] C. Cachin, "An Information-Theoretic Model for Steganography", in proceeding 2nd Information Hiding Workshop, vol. 1525, pp. 306-318, 1998 [11] S. Tanako, K. Tanaka and T. Sugimura, "Data Hiding via Steganographic

Image Transformation", IEICE Trans. Fundamentals, vol. E83-A, pp. 311-319, February, 2000.

[12] F.A.P Peticolas, R.J. Anderson and M.G. Kuhn, "Information Hiding - A Survey", in proceeding of IEEE, pp. 1062-1078, July 1999.

[13] N.F. Johnson, S. Jajodia, "Staganalysis: The Investigation of Hiding Information", *IEEE*, pp. 113-116, 1998.

[14] M.M. Amin, M. Salleh, S. Ibrahim, et al., "Information Hiding Using

Steganography", 4 National Conference On Telecommunication Technology Proceedings (NCTT2003), Shah Alam, Malaysia, pp. 21-25, January 14-15, 2003.