A REVIEW OF DIGITAL STEGANOGRAPHY METHODS

Asha Durafe, Assistant Professor,
Shah and Anchor Kutchhi Engineering College (India)

Abstract: Security in data communication is a very important concern today. It is used in almost every region like e-commerce, education, and industry and data warehouse. Steganography literally means secret writing. The technique has been used in various forms for 2500 years or long but in the last two decades Steganography has been introduced to digital media.

The goal of steganography is to hide information by concealing it into some other medium like image, audio or video so that only sender and intended recipient knows existence of information in communication. So the main focus of this paper is to study and discuss the trends which are already been proposed in the direction of cryptography and steganography.

The main focus of this paper is on emerging concepts and recent techniques like Parity Coding, Least Significant Bit Coding (LSB), Phase Coding, Echo Data Hiding, Spread Spectrum about audio steganography. The performance of recent audio steganography techniques on the basis of strengths, weaknesses and hiding rate is evaluated. Some advancement like Generic Algorithm based Audio Steganography and steganalysis concepts are also discussed.

Keywords: Steganography, data hiding, parity, coding, security measures, LSB, cryptography, steganalysis

1. INTRODUCTION

Steganography is the art of invisible communication. Its purpose is to hide the very presence of communication by embedding messages into innocuous-looking cover objects. Each steganographic communication system consists of an embedding algorithm and an extraction algorithm. To accommodate a secret message in a digital cover, the original cover is slightly modified by the embedding algorithm. The result is modified cover object that contains the secret message and it is called stego object. The important requirement for a steganographic system is its detectability by an attacker with probability not better than random guessing, given the full knowledge of the embedding algorithm, including the statistical properties of the source of cover object. Of course the stego key is not revealed.

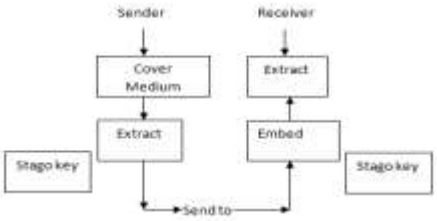

Figure1. Block diagram of Steganography scheme

Sustainable steganography can be described as a method of hiding in such a way that no color bit is altered. Today various digital data formats are used in steganography. Most popular among them are bmp, doc, gif, jpeg, mp3, txt and wav because of the relative ease by which redundant or noisy data can be removed from them and replaced with a hidden message. BMP image is found to be more suitable because of presence of some redundant bytes at the quad word boundary.

2. S/W MEANS OF PROTECTION

From ancient times, several methods are used for protecting data/information.

2.1 Cryptography:

Cryptography is the science using for secure communication. It converts the plaintext to cipher text for preventing unauthorized person to access the message/information. Several algorithms are developed in the intention of protecting data. In those, some famous algorithms are DES, TDES, AES, Blowfish, RSA etc. AES is the commonly used standard by the government of USA.

2.2Steganography:

It is a step forward from Cryptography. It protects the information inside an image. Since nobody suspects about the concealing of information, many steganography techniques are established to hide the data effectively. Some famous steganography algorithms are LSB, RGB, PVD, etc. Apart

from these techniques Water marking, Visual cryptography etc are other techniques used for protection of data.

As already discussed, most of the existing methods have many disadvantageous even though they provide a basic level of protection.

Most of the hardware based protection methods are very difficult to implement, as it requires additional hardware like a dongle.

H/W based methods are not cost effective as the additional hardware usually implements proprietary algorithms and hardware

Easy for crackers to identify the existence of additional hardware and hence they would be able to write additional backdoor programs which can clone the functionality of the hardware and there by bypassing the protection method also.

CD based protection of games can be easily manipulated using cloning software

Serial number based protection method can be replicated to other systems and software can be pirated.

Secure id based protection using RSA is cumbersome to maintain and expensive.

2.2.1TYPES OF STEGANOGRAPHY

- Steganography in Images: Image steganography is very effective, efficient and can serve a variety of purposes included authentication, concealing of messages etc. The hidden messages will change the last bit of byte in an image in Least significant bit method [3]. So by doing this, there will be relatively no change within the carrier image.
- Steganography in Audio: In audio Steganography system, secret messages are embedded into digitized audio signal which results into altering binary sequence of corresponding audio files.
- Steganography in Video: Steganography in Videos basically deals with hiding of information in each frame of video. This type reveals more information instead of hiding.
- Steganography in Text: Encoding secret messages in text can be a very challenging task. This is because text files contain redundant data to replace with a secret message. Another cons is the ease of which text based Steganography can be altered by an unwanted parties by just changing the text itself or reformatting the text to some other form (from .TXT to .PDF, etc.).

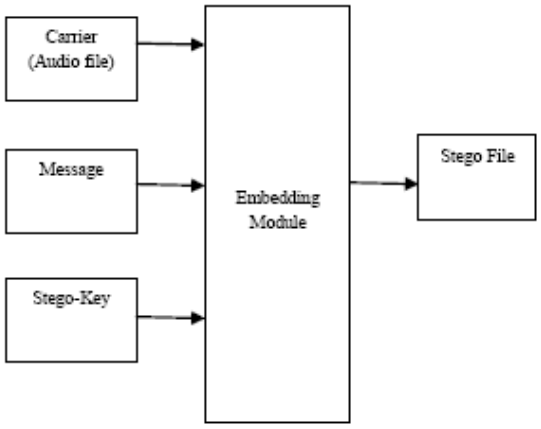

Fig.2: Basic Audio Steganographic Model

3. AUDIO STEGANOGRAPHY

The technique which embeds secret messages into digital sound is called Audio Steganography. The process is usually a more difficult than embedding messages in other media [4]. Audio Steganography methods can embed any messages in WAV, and even MP3 sound files. In Audio Steganography system, secret messages are embedded into digitized audio signal which results into altering binary sequence of corresponding audio files. Audio steganography techniques are lesser prone to malicious attacks because many attacks which are malicious against image steganography algorithms like geometrical distortion, spatial scaling cannot be implemented against audio steganography schemes. Audio steganography in particular addresses key issues brought about by the the P2P software, MP3 format, and the need for a secure broadcasting scheme that can maintain the secrecy of the transmitted information, even when passing through insecure channels.

AUDIO STEGANOGRAPHY TECHNIQUES

A. Least Significant Bit Coding: The most commonly used steganographic method is the Least Significant Bit (LSB) replacement. It works by embedding message bits as the LSBs of randomly selected pixels. A secret stego key shared by the communicating parties usually determines pixel selection.

The popularity of LSB embedding is due to its simplicity as well as the false belief that modifications of LSBs in randomly selected pixels are undetectable because of the noise commonly present in digital images of natural scenes.

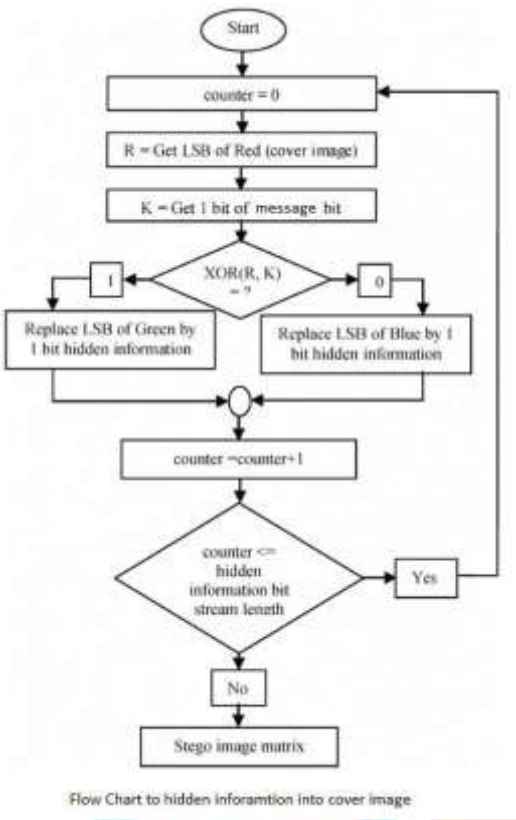

Fig.3 LSB Coding

Phase Coding : Phase coding addresses the disadvantages of the noise-inducing methods of audio Steganography. Phase Coding works by substituting the phase of an initial audio segment with a reference phase that represents the data[7]. This technique relies on the fact that the phase components of sound are not as distinguishable to the human ear as noise is. Rather than introducing perturbations, the technique encodes the message bits as phase shifts in the phase spectrum of a digital signal, achieving an inaudible encoding in terms of signal-to-perceived noise ratio (SPNR).

B. Multiple LSB Algorithm:

Multiple LSB algorithm is a specialized form of the standard LSB algorithm. In this novel approaches of LSBs of audio samples for data hiding are used. These methods check the MSBs of the samples, and then number of LSBs for data hiding is decided. In this way, multiple and variable LSBs are used for embedding secret data. These proposed methods remarkably increase the capacity for data hiding as compared to standard LSB without causing any noticeable distortion to the data. This method considers the value of the MSB of the digitized samples of cover audio for data hiding.

Steps for data embedding:-

1) Read the cover audio signal

2) Write the text in a file to be embedded.

3) Generate public and private key by using RSA algorithm

4) Apply encryption function on plaintext to produce cipher text.

5) Convert it into a sequence of binary bits

6) Every message bit from step 5 is embedded into the variable and multiple LSBs of the samples of the digitized cover audio cover

7) For embedding purpose, the MSB of the cover sample is checked

8) If MSB is “0” then use 6 LSBs for data embedding

9) If MSB is “1” then use 7 LSBs for data embedding

10) The modified cover audio samples are then written to the file forming the stego object

Steps for data extraction:-

1) Read the stego object

2) Retrieval of message bits is done by checking the MSB of the samples

a) If MSB is “0” then use 6 LSBs for data retrieve

b) If MSB is “1” then use 7 LSBs for data retrieve

3) Recover the cipher text from stego object file

4) After every such 16 messages bits retrieved, they are converted into their decimal equivalents to get ascii values of the secret message

5) Finally apply RSA decryption to recover plaintext

C. RSA Algorithm:-

RSA algorithm was designed by Ron Rivest, Adi Shamir and Leonard Adleman. This algorithm was specially designed for encryption of data. RSA algorithm is one of the technique use to implement cryptography. RSA algorithm is based on public key cryptography. It generates public and private key. Public key is used for encryption purpose and private key is used for decryption purpose.

The basic algorithm is as follows :-

1) Choose 2 prime numbers p and q

2) Compute n=p q

3) Compute m(Euler's totient function)=(p-1)(q-1)

4) Select e such that GCD(e,m)=1

5) Where (e,n) is the public key

6) Find d such that 1=d e mod m

7) 1 = km + de

8) Using the extended Euclid Algorithm we see that 1= - km + de

9) where (d,n) is the private key

10) Encryption function y = x^e mod n

11) Decryption function x = y^d mod n

12) Here x is the number to be encrypted and y is its encrypted form

Parity Coding [7]: Parity coding is one of the robust audio Steganographic techniques. Instead of breaking a signal into individual samples, this method breaks an original signal into separate samples and embeds each bit of the secret message from a parity bit. If the parity bit of a selected region does not match the secret bit to be encoded, the process inverts the LSB of one of the samples in the region. Thus, the sender has more of a choice in encoding the secret bit.

Echo data hiding [6]: Echo data hiding deals with embedding of text (or data) in audio file by introducing an echo to the original signal . The data then hidden by varying three parameters of the echo:

I. Initial amplitude,

II. Decay rate, and

III. Offset

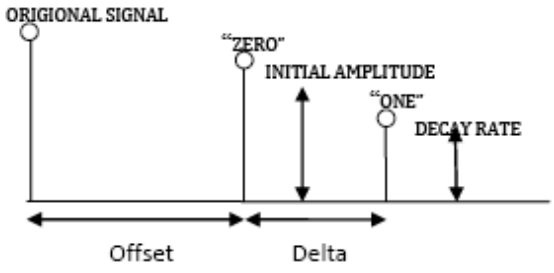

Fig.4: Adjustable parameters

Spread Spectrum: In audio steganography, the basic spread spectrum (SS) method attempts to spread secret information across the frequency spectrum of the audio signal using a code which is independent of the actual signal [6]. Two versions of Spread Spectrum can be used in audio Steganography:

Direct-sequence: Direct-sequence SS attempts to spread out the secret message by a constant called the chip rate and then modulated with a pseudorandom signal and interleaved with the cover-signal[7]. Frequency-hopping schemes. In frequency-hopping SS, the frequency spectrum of audio files is changed so that it hops rapidly between frequencies. Steps of spread spectrum are given below

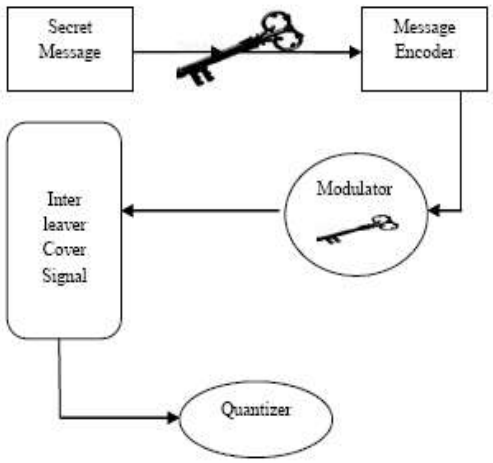

Fig 5: Spread spectrum Steps

The figure 5 demonstrates the steps of spread spectrum as:

a. The secret message is encrypted using a symmetric key k1.

b. Then encode encrypted message using a low rate error correcting code that increase overall robustness of the system.

c. The encoded message is then modulated with a pseudorandom signal that was generated using a second symmetric key.

d. The resulting random signal that contains the message is interleaved with cover signal.

e. The final signal is quantized to create a new digital audio file that contains the message.

f. This process is reversed for message extraction.

ADVANTAGES

1) Secrecy in terms of message hiding.

2) High Capacity of the audio file.

3) Accurate Extraction.

4) Resistance from external attacks.

DISADVANTAGES

1) Small alteration in the audio file.

2) Complications in recovering original audio file.

4. COMPARISON OF AUDIO STEGANOGRAPHY TEQNIQUES

Method	Embedding Techniques	Strengths	Weakness	Hiding Rate
Least Significant Bit	LSB of each sample in the audio is replaced by one bit of hidden information	Simple and easy way of hiding Information with high bit rate	Easy to extract and to destroy	16 Kbps
Echo Hiding	Embeds data by introducing echo in the cover signal	Resilient to lossy data compression algorithms	Low security and capacity	40-50 Bps
Phase Coding	Modulate the phase of the cover signal	Robust against signal processing manipulation and data retrieval needs the original signal	Low capacity	333 Bps
Parity Coding	Break the signal into separate samples and embeds each bit from secret message in sample region parity bit	Sender has more of a choice in encoding the secret bit.	Not Robust	320bps
Spread Spectrum	Spread the data over all signal frequencies	Provide better robustness	Vulnerable to time scale modification	20 Bps

5. STEGANALYSIS

Discussion of steganography necessarily leads to discussion of steganalysis. Steganalysis is the art of discovering the presence or transmission of stego content in the communication channel. From information security point of view it is crucial to thwart any attempt of stego communication. It is a probabilistic science. A steganalysis algorithm generally output a number within a given range to indicate the likelihood that stego was actually used on the input file. Steganalysis is a rapidly advancing science, and will continue to develop as long as steganographic algorithms are being created and used.

There are two approaches to the problem of steganalysis, one is specific to a particular steganographic algorithm. The other is developing techniques completely independent of the steganographic algorithm to be analyzed. Each of the two approaches has its own advantages and disadvantages. A steganalysis technique specific to an embedding method would give very good results when tested only on that embedding method and in all possibility it fails on other steganographic algorithm. However an independent steganalysis method may perform less accurately but it is capable to provide acceptable result on any new embedding algorithm. When the embedding algorithm is based on a graph theoretic approach that either does not change/replace any color bits of cover or changes very few colors bit, a general steganalysis algorithm cannot work. In fact associations in the maximum bipartite matching takes care of the embedding process.

In general, the number of unmatched vertices is an upper bound on the number of changes to first-order statistics. An experimental result has shown that sufficiently good matching (< 3% unmatched) can be reached for natural cover data. This makes a graph theoretic approach practically undetectable by tests that look only at first-order statistics. It would be interesting to run a blind steganalysis scheme against this implementation to compare its detectability to the other tested algorithms. Adding a restriction to the set of edges could easily extend this graph theoretic approach and make it more secure against any general steganalysis algorithm. However in the approach presented the steganography totally depends on the key pairs consisting of size of message and data unit size. To the third party none of the information is known and hence given the arbitrariness of message size and data unit size for any message, it is simply Herculean for anybody to guess and extract the message.

Today, a fairly large numbers of downloadable steganographic programs are available on the Internet that is based on the Least Significant Bit (LSB) embedding technique. A few examples are: Steganos II, S-Tools 4.0, Steghide 0.3, Contraband Hell Edition, Web Stego 3.5, EncryptPic 1.3, StegoDos, Winstorm, Invisible Secrets Pro, and many others.

CONCLUSION:

In this paper several aspects of cryptography and steganography with their working approached and enhancements are discussed. Based on the analysis and observations it would be better to hybrid different encryption technique. The increasing size of key with random attribute is also a better and powerful security improvement.

There are many steganographic algorithms available like JSteg, F5 and LSB algorithms. Least Significant Bit algorithm is widely used in designing the steganographic application because LSB algorithm works efficiently when we consider bit map images .bmp files. The speed of embedding is also high when using LSB compared to the JSteg algorithm.

REFERENCES:

[1] S. Katzenbeisser, F.A.P. Petitcolas, "Information Hiding Techniques for Steganography and Digital Watermarking", Artech House, Norwood, MA, 2000.

[2] G. J. Simmons, "The prisoners' problem and the subliminal channel" in Proc. Advances in Cryptology (CRYPTO '83), pp. 51-67.Berglund, J.F.and K.H. Hofmann, 1967. Compact semitopological semigroups and weakly almost periodic functions. Lecture Notes in Mathematics, No. 42, Springer-Verlag, Berlin-New York.

[3] Masoud Nosrati, Ronak Karimi, Mehdi Hariri, "An introduction to steganography methods", World Applied Programming, Vol (1), No (3),August 2011. 191-195.

[4] Min Wu, Bede Liu, "Multimedia Data hiding", Springer-Verlag New York, 2003

[5] Bender W, Gruhl D & Morimoto N (1996) "Techniques for data hiding". IBM Systems Journal 35(3): p 313–336.

[6] M.Ram Mohan Reddy S.S Divya. Hiding text in audio using multiple lsb steganography and provide security using cryptography. International Journal of Scientic Technology Research, 1(6):68-70, 2012 Shobha lokhande. 2012.

[7] Tanmay Bhowmik Pramatha Nath Basu. On embedding of text in audio a case of steganography.IEEE publication. pages 203-206, 2010.

[8] K. Gopalan. Audio steganography using bit modication. Proc. IEEE Int.Conf. Acoustics, Speech, and Signal Processing, 2(2):421-424, 2009.

[9] Teoh Suk Kuan Rosziati Ibrahim.Steganography algorithm to hide secret message inside an image.Computer Technology and Application, 2:102-108, 2011.

[10] Min Wu, Bede Liu, "Multimedia Data hiding", Springer-Verlag New York, 2003

[11] Mehdi Hussain and Mureed Hussain, "A Survey of Image Steganography Techniques" , International Journal of Advanced Science and Technology Vol. 54, May, 2013

[12] Trivedi S, Chandramouli R. Secret key estimation in sequential steganography. IEEE Transactions on Signal Processing. 2005; 53(2):746-57.