Data Hiding Techniques for an Efficient and Secure Communication

Ms.Surekha Shrivastava[1] Mr. Gajendra Singh Chandel[2] Mr. Kailash Patidar[3]
[1]M.Tech. (CTA) Student [2]Head of Department & Professor [3]Assistant Professor
[1,2,3]Department of Computer Science & Engineering
[1,2,3]SSSIST, Sehore

Abstract— Steganography has been in use since the presence of secret messages and this use has gained popularity as the internet became well-liked. Steganography works by replacing bits of useless or unused data in regular computer files (such as graphics, sound, text, HTML, or even floppy disks) with bits of different, invisible information. This hidden information can be plain text, cipher text, or even images. Audio steganography is a young branch of this discipline. An encoding mechanism is used for embedding the message into the audio file. This paper intends to give an overview of audio steganography, its uses and techniques. It also attempts to identify the requirements of a good steganographic algorithm and briefly reflects on which steganographic techniques are more suitable for which applications.

Key word: Image steganography Techniques, Text Steganography Techniques, Spread Spectrum

I. INTRODUCTION

In steganography does not alter the structure of the secret message, but hides it inside a coverimage so that it cannot be seen. A message in a cipher text, for instance, might arouse suspicionon the part of the recipient while an "invisible" message created with steganographic methodswill not. In other word, steganography prevents an unintended recipient from suspecting that thedata exists. In addition, the security of classical steganography system relies on secrecy of thedata encoding system. Once the encoding system is known, the steganography system is defeated.

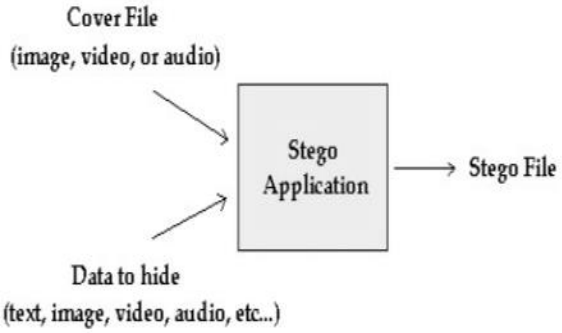


Fig. 1: Steganography Application Scenario

The steganography application hides different types of data within a cover file. The resultingstego also contains hidden information, although it is virtually identical to the cover file. WhatSteganography essentially does is exploit human perception; human senses are not trained to look for files that have information hidden inside of them, although there are programs available that can do what is called Steganalysis (Detecting use of Steganography.)

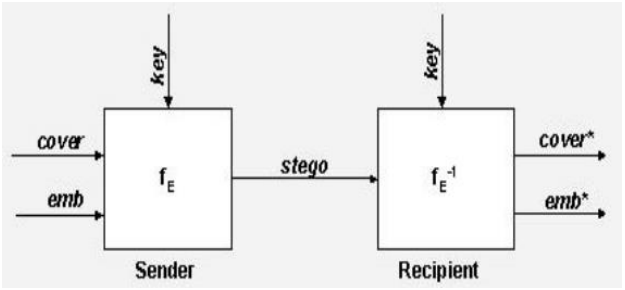
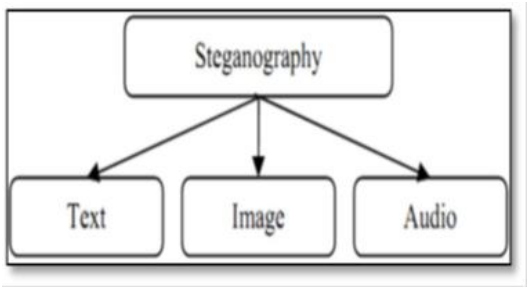


Fig. 2: A generic Steganography System

The components of steganographic system are:

- Emb : The message to be embedded.
- Cover : The data in which emb will be embedded.
- Stego : A modified version of cover that contains the embedded message emb.
- Key :Additional secret data that is needed for the embedding and extracting processes and must be known to both, the sender and the recipient.
- f_E :A steganographic function that has cover, emb and key as parameters and produces stego as output.
- f_E^{-1} : A steganographic function that has stego and key as parameters and produces emb as output. F_E^{-1} is the inverse function of f_E in the sense that the result of the extracting process f_E^{-1} is identical to the input E of the embedding process f_E.

II. DIFFERENT KINDS OF STEGANOGRAPHY



A. Text Steganography:

It hides the text behind some other text file. It is the a difficult form of steganography as the redundant amount of text to hide the secret message is scarce in text files.

1) Text Steganography Techniques:

1) Selective hiding:This hides the characters in the first (or any specific location) characters of the words. Concatenating those characters help extracting the text. But this technique requires huge amount of plain text.
2) HTML web pages**:** This may hide text using the fact that attributes of HTML tags are case insensitive. Those characters can then be used to retrieve the original text.

Text steganography can use HTML pages to hide the text behind them because of two factors:

3) Web pages are present in a vast amount and detecting which one is containing hidden information is next to impossible.
4) The order of tags used for formatting the appearance of a web page does not matter and this can help to hide one bit of text behind the tags.
5) Hiding using Whitespace**:** Fewer number of whitespaces may specify a 0 and more number of whitespaces between words may determine a 1.
6) Semantic Hiding: Uses synonyms to hide the message.

2) Image Steganography:

It is one of the most commonly used technique because of the imitation of the Human visual System(HVS). Human eye cannot detect the vast range of colors and an insignificant change in the quality of an image that results from steganography.

B. Image Steganography Techniques:

1) Patchwork

Patchwork is a statistical technique that uses redundant pattern encoding to embed a message in an image . The algorithm adds redundancy to the hidden information and then scatters it throughout the image .A pseudorandom generator is used to select two areas of the image (or patches), patch A and patch B. All the pixels in patch A is lightened while the pixels in patch B is darkened. In other words the intensities of the pixels in the one patch are increased by a constant value, while the pixels of the other patch are decreased with the same constant value . The contrast changes in this patch subset encodes one bit and the changes are typically small and imperceptible, while not changing the average luminosity.

A disadvantage of the patchwork approach is that only one bit is embedded. One can embed more bits by first dividing the image into sub-images and applying the embedding to each of them. The advantage of using this technique is that the secret message is distributed over the entire image, so should one patch be destroyed, the others may still survive. This however, depends on the message size, since the message can only be repeated throughout the image if it is small enough. If the message is too big, it can only be embedded once.

2) Spread Spectrum:

In spread spectrum techniques, hidden data is spread throughout the cover-image making it harder to detect . A system proposed by Marvel et al. combines spread spectrum communication, error control coding and image processing to hide information in images.

Spread spectrum communication can be defined as the process of spreading the bandwidth of a narrowband signal across a wide band of frequencies. This can be accomplished by adjusting the narrowband waveform with a wideband waveform, such as white noise. After spreading, the energy of the narrowband signal in any one frequency band is low and therefore difficult to detect. In spread spectrum image steganography the message is embedded in noise and then combined with the cover image to produce the stego image. Since the power of the embedded signal is much lower than the power of the cover image, the embedded image is not perceptible to the human eye or by computer analysis without access to the original image.

3) Audio Steganography:

It is also a difficult form of steganography as humans are able to detect a minute change in the quality of audio. To hide information in audio files similar techniques are used as for image files. One different technique unique to audio steganography is masking, which exploits the properties of the human ear to hide information unnoticeably. A faint, but audible, sound becomes inaudible in the presence of another louder audible sound.

III. AUDIO STEGANOGRAPHY TECHNIQUES

A. LSB Coding:

An audio steganography technique could be grouped into two assemblies dependent upon the area of operation. Steganography is implementing using audio as host and image or text as watermark data or secret data. Least significant bit (LSB) is used for watermarking process on the audio. Few work done by researchers on this technique, which is one of the common techniques employed in signal processing applications. It is based on the substitution of the LSB of the carrier signal with the bit pattern from the stego noise. The robustness depends on the number of bits that are being replaced in the host signal. This type of technique is commonly used because, each frame is represented as an integer hence it will be easy to replace the bits. The audio signal has real values as samples, if converted to an integer will degrade the quality of the signal to a great extent. The operation of the 2-bit LSB coding is shown in following figure.

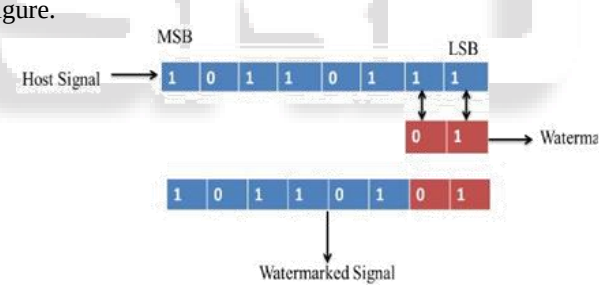


Fig. 3: 2-Bit LSB modification

B. Related Work:

In this section we will look into the review of digital watermarks used for images. It describes the previous work which had been done on digital watermarking by using LSB technique and other techniques, including the analysis of various watermarking schemes and their results.

Gaurav Bhatnagar et al, presented a semi-blind reference watermarking scheme based on discrete wavelet transform (DWT) and singular value decomposition (SVD) for copyright protection and authenticity. Their watermark was a gray scale logo image. For watermark embedding, their algorithm transformed the original formed using directive contrast and wavelet coefficients. Then, their algorithm embedded the watermark into reference image by modifying the singular values of reference image using the singular values of the watermark.

Hao Luo et al, proposed a self-embedding watermarking scheme for digital images. In their proposed algorithm they used the cover image as a watermark. It generates the watermark by halftoning the host image into a

halftone image. Then, the watermark is permuted and embedded in the LSB of the host image. The watermark is retrieved from the LSB of the suspicious image and inverse permuted.

Wen-Chao Yang et al,used the PKI (Public-Key Infrastructure), Public-Key Cryptography and watermark techniques to design a novel testing and verifying method of digital images. The main idea of their paper is to embed encryption watermarks in the least significant bit (LSB) of cover images.

Hong Jie He et al, proposed a wavelet-based fragile watermarking scheme for secure image authentication. In their proposed scheme, they generated the embedded watermark using the discrete wavelet transform (DWT), and then they elaborated security watermark by scrambling encryption is embedded into the least significant bit (LSB) of the host image.

Sung-ChealByun et al, proposed a fragile watermarking scheme for authentication of images. They used singular values of singular value decomposition (SVD) of images to check the integrity of images. In order to make authentication data, the singular values are changed to the binary bits using modular arithmetic. Then, they inserted the binary bits into the least significant bits (LSBs) of the original image. The pixels to be changed are randomly selected in the original image.

Gil-Je Lee et al presented a new LSB digital watermarking scheme by using random mapping function. The idea of their proposed algorithm is embedding watermark randomly in the coordinates of the image by using random function to be more robust than the traditional LSB technique.

SaeidFazli et al presented trade-off between imperceptibility and robustness of LSB watermarking using SSIM Quality Metrics. In their algorithm, they put significant bit-planes of the watermark image instead of lower bit-planes of the asset picture.

DebjyotiBasu et al proposed Bit Plane Index Modulation (BPIM) based fragile watermarking scheme for authenticating RGB color image. By embedding R, G, B component of watermarking image in the R, G, B component of original image, embedding distortion is minimized by adopting least significant bit (LSB) alteration scheme. Their proposed method consists of encoding and decoding methods that can provide public detection capabilities in the absences of original host image and watermark image.

C. Phase Coding:

This section presents some common methods used for hiding secret information in audio. Many software implementations of these methods are available on the Web and are listed in the relatives section.

Phase coding, when it can be used, is one of the most effective coding methods in terms of the signal-to perceived noise ratio. When the phase relation between each frequency component is dramatically changed, noticeable phase dispersion will occur. However, as long as the modification of the phase is sufficiently small (sufficiently small depends on the observer; professionals in broadcast radio can detect modifications that are imperceivable to an average observer), an inaudible coding can be achieved . Phase coding relies on the fact that the phase components of sound are not as perceptible 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. Phase coding method is used when only a small amount of data needs to be considered.
Phase coding is explained in the following procedure:

- The original sound signal is broken up into smaller segments whose lengths equal the size of the message to be encoded.
- A Discrete Fourier Transform (DFT) is applied to each segment to create a matrix of the phases and Fourier transform magnitudes.
- Phase differences between adjacent segments are calculated.
- Phase shifts between consecutive segments are easily detected. In other words, the absolute phases of the segments can be changed but the relative phase differences between adjacent segments must be preserved.

D. Related Work:

The phase coding method works by substituting the phase of an initial audio segment with a reference phase that represents the data. The phase of subsequent segments is adjusted in order to preserve the relative phase between segments. Using the new phase matrix and original magnitude matrix, the sound signal is reconstructed by applying the inverse DFT and then concatenating the sound segments back together. To extract the secret message from the sound file, the receiver must know the segment length. The receiver can then use the DFT to get the phases and extract the information (consider Figure 4 for phase cosing procedure).

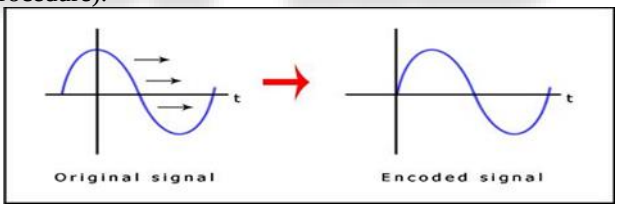


Fig. 4: The signals before and after Phase coding procedure

E. Spread Spectrum:

The proposed stego system uses Spread Spectrum technique which is appliedin spatial domain together with error correction coding. These are used to increase the security and robustness of the system. Random location selection within the cover image pixels is also proposed in the work. Improvement has been achieved in robustness on the expense of reducing the capacity of hiding. The imperceptibility of the stego image is assessed by using peak signal-to-noise ratio (PSNR) measure. Attacks in the form of lossy compression and additive noise are considered. The performance of the proposed system has shown good immunity to moderate levels of channel noise and lossy compression ratios.

F. Related work:

A lot of work has been done in the field of information hiding in the context of the spread spectrum technology. In a DSSS system, a signal of low bandwidth is spread over a

broad frequency range. Hence the power of the signal is decreased and thus the signal vanishes in the noise of the cover media. To extract an embedded signal from the cover, the receiver needs knowledge about the spreading process. This knowledge can therefore be described as a kind of secret key, needed as input to the system.

For embedding we used many technique. For example As explained by Rizky M. Nugraha et. Al. Image steganography has widely developed. There are also many algorithm developed for it. Meanwhile, the interest in using audio data as cover object in steganography can be spelled out late emergence than image data. This paper discusses the implementation of steganography in audio data using Direct Sequence Spread Spectrum method. Spread Spectrum method is often used to send hidden message through radio waves. This message is transmitted through noise-like wave. The same method can be applied to embed message in audio data. The embedded audio data will be heard as noise. The Spred Spectrum method used in this paper is Direct Sequence Spread Spectrum. A key is needed to embed messages into noise, this key is used to generate pseudo-noise wave. The information to be embedded must first modulated using the pseudo-noise.

G. Echo Hiding:

Audio watermarking or audio steganography started consider later as attractive area that have viable applications and space for development (Zhang et al., 2010a, b; Abdulfetah et al., 2010a, b). In the past few years, several techniques for data hidden in audio sequences have been presented. All of the developed techniques take benefit of the perceptual properties of the human auditory system (HAS).

The main challenge in digital audio watermarking and steganography is that if the perceptual transparency parameter is fixed, the design of a watermark system cannot obtain high robustness and a high watermark data rate at the same time (Cvejic, 2004; Yang et al., 2009). To achieve any of data hidden goals, we need to select a proper cover, domain, and take into the account the challenges of data hidden approaches.

Arnold (2000) has tried to improve the performance of the original patchwork algorithm. Arnold's algorithm is a landmark in the area of watermarking research, especially for patchwork algorithm. Moreover, the performance of this algorithm in terms of inaudibility and robustness has been shown to be satisfactory by many researchers such as (Yeo and Kim, 2003). They have derived mathematical formulations that help to improve robustness. The core idea of the improved scheme is called the Modified Patchwork Algorithm (MPA) which can enhance the power of the original patchwork algorithm considerably.

Large work has been carried out in audio watermarking using spread spectrum technology and is presented in several key publications like (Bender et al., 1996), (Cox et al., 2002) and (Cvejic, 2004). The first method of spread spectrum into watermarking was in (Cox et al., 1997). Xu et al. (1999) proposed a multiple echo technique. Rather than embedding one large echo into the host audio signal, they use multiple echoes with different offsets. Oh et al. (2001) introduced the positive-negative echo hiding scheme. Their echo kernels comprise positive and negative echoes at nearby locations. Since the frequency response of a negative echo is the inversed shape having similar ripples as that of a positive echo, the frequency response of the positive and negative echoes has the smooth shape in the low frequency band. By employing positive and negative echoes, one can thus embed multiple echoes to allow that the host audio quality is not apparently deteriorated. Kim and Choi (2003) presented an echo hiding scheme with backward and forward kernels. The theoretically-derived results show that the amplitude of the cepstrum coefficient at the echo position from the backward and forward kernels is bigger than that from the backward kernel only when the embedded echoes are symmetric. Therefore, the backward and forward kernels can improve the robustness of echo hiding scheme.

Ko et al. (2005) went further to propose the time-spread echo kernel. With the use of pseudo-noise sequence, an echo is spread out as numerous little echoes in a time region. When the embedded data of watermarked audio signals are extracted, the pseudo-noise sequence functions like a secret key. Without obtaining the pseudo-noise sequence used in the embedding process, extracting the embedded data would be harder.

In order to add a watermark into a host signal in a perceptually transparent manner, a wide range of embedding techniques are proposed going from simple least significant bits (LSB) scheme or Low-bit encoding, Phase coding, Spread spectrum, Patchwork coding, Echo coding and noise gate technique.

H. Related work:

Echo-hiding method for audio watermarking. The method is quite different from previous echo-hiding methods since it presents a new echo kernel which introduces a forward kernel as well as a backward kernel. The new kernel, a combination of the backward and forward kernels, can enhance considerably rate. Thus, it is possible to reduce echo amplitude. The echo hiding scheme in which the analysis-by-synthesis approach, interlaced kernels, and frequency hopping are adopted to achieve high robustness, security, and perceptual quality. The amplitudes of the embedded echoes are adequately adapted during the embedding process by considering not only the characteristics of the host signals, but also cases in which the watermarked audio signals have suffered various attacks. Additionally, the interlaced kernels are introduced such that the echo positions of the interlaced kernels for embedding "zero" and "one" are interchanged alternately to minimize the influence of host signals and various attacks on the watermarked data. Frequency hopping is employed to increase the robustness and security of the proposed echo hiding scheme in which each audio segment for watermarking is established by combining the fractions selected from all frequency bands based on a pseudo noise sequence as a secret key. Experimental results indicate that the proposed analysis-by-synthesis echo hiding scheme is superior to the conventional schemes in terms of robustness, security, and perceptual quality.

IV. PROPOSED WORK

Here we will discuss the disadvantages of the previous procedure and how those are different with present method. The main disadvantages associated with the use of existing methods like echo hiding, spread spectrum and parity coding are, human ear is very sensitive to noise and it can often detect even the slightest bit of noise introduced into a sound file and another problem is robustness.

Phase coding has main disadvantage of low data transmission rate because of the fact that the secret message is encoded only in the first signal segment. Hence this method is used only when a small amount of data needs to be transferred.

Among different information hiding techniques proposed to embed secret information within audio file, Least Significant Bit (LSB) coding method is the simplest way to embed secret information in a digital audio file by replacing the least significant bit of audio file with a binary message. Hence LSB method allows large amount of secret information to be encoded in an audio file.

Steps to hide secret information using LSB are:

1) Covert the audio file into bit stream.
2) Convert each character in the secret information into bit stream.
3) Replace the LSB bit of audio file with the LSB bit of character in the secret information.

This proposed method provides greater security and it is an efficient method for hiding the secret information from hackers and sent to the destination in a safe and undetectable manner. This proposed system also ensures that the size of the file is not changed even after encoding and it is also suitable for any type of audio file format.

V. CONCLUSION

In this paper we have introduced a robust method of imperceptible audio data hiding. Thus we conclude that audio data hiding techniques can be used for a number of purposes other than covert communication or deniable data storage, information tracing and finger printing, tamper detection. As the sky is not limit so is not for the development. Man is now pushing away its own boundaries to make every thought possible. So similarly these operations described above can be further modified as it is in the world of Information Technology.

REFERENCES

[1] Mazdak Zamani, AzizahBt Abdul Manaf, RabiahBt Ahmad, FarhangJaryani, "A secure audio steganography approach", International Conference for Internet Technology and Secured Transactions, Page(s): 1 – 6, 2009.
[2] Kaliappan Gopalan., "Audio steganography using bit modification", 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, Page(s): II - 421-4 vol.2.
[3] ZaidoonKh. AL-Ani, A.A.Zaidan, B.B.Zaidan and Hamdan. O. Alanazi, Overview: Main Fundamentals for Steganography, journal of computing, volume 2, issue 3, march 2010, issn 2151-9617.
[4] Sos S. Agaian, David Akopian and Sunil A. D'Souza" TWO ALGORITHMS IN DIGITAL AUDIO STEGANOGRAPHY USING QUANTIZED FREQUENCY DOMAIN EMBEDDING AND REVERSIBLE INTEGER TRANSFORMS" 1Non-linear Signal Processing Lab, University of Texas at San Antonio, Texas 78249, USA.
[5] Ashwini Mane, GajananGalshetwar, AmuthaJeyakumar, "DATA HIDING TECHNIQUE: AUDIO STEGANOGRAPHYUSING LSB TECHNIQUE" International Journal of Engineering Research and Applications (IJERA), Vol. 2, Issue 3, May-Jun 2012, pp.1123-1125
[6] Neil Jenkins, Jean Everson Martina ," Steganography in Audio" Anais do IX SimpósioBrasileiroemSegurança da Informação e de SistemasComputacionais page: 269-278,2007
[7] Westfeld, A. (2003). Detecting low embedding rates. In Petitcolas, F. A., editor, Information Hiding: 5th International Workshop. Springer.
[8] I. J. Cox, J. Kilian, F. T. Leighton and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia," IEEE Trans. Signal Processing, vol. 6, no. 12, pp. 1673- 1687, December 1997.
[9] B. Chen and G. W. Wornell, "Digital Watermarking and Information Embedding using Dither Modulation," Multimedia Signal Processing, 1998 IEEE Second Workshop, pp: 273-278, December 1998

