International Journal of Engineering, Applied and Management Sciences Paradigms, Vol. 02, Issue 01, April 2013
ISSN (Online): 2320-6608
www.ijeam.com

Audio Steganography using Matlab

Rashmi Mishra[1], Rupal chauhan[2], Poonam Jhariya[3], Priya Mishra[4], Vaishali Yadav[5], Tazeem A Hazra[6]

[1 to 6]BE Students (7th Sem) Department of electronics and communication, Takshshila Institute of Engineering and Technology, Sharda chowk, Madan Mahal, Jabalpur, M.P. (India)

Abstract

The paper entitled Audio Steganography using matlab is the application developed to embed a Text data in to another audio signal. It is concerned with embedding information in an innocuous cover Speech in a secure and robust manner. This system makes the Files more secure by using the concepts Steganography. In a computer-based audio Steganography system, secret messages are embedded in digital sound. The secret message is embedded by slightly altering the binary sequence of a sound file. Embedding secret messages in digital sound is usually a more difficult process than embedding messages in other media, such as digital images. In order to conceal secret messages successfully, a variety of methods for embedding information in digital audio have been introduced. These methods range from rather simple algorithms that insert information in the form of signal noise to more powerful methods that exploit sophisticated signal processing techniques to hide information. The next section discusses these methods in greater detail.

Keywords: *Audio Steganography, Matlab, Techniques for Data Hiding.*

1. Introduction

As the need of security increases only encryption is not sufficient. So steganography is the supplementary to encryption. It is not the replacement of encryption. But Steganography along with encryption gives more security to data. The word steganography is of Greek origin and means "concealed writing" from the Greek words stegnos meaning "covered or protected", and graphei meaning "writing". Steganography is the technique to hide the information in some media so that third party can't recognize that information is hidden into the cover media.. That media may be text, image, audio or video. The information that to be hidden is called stego and the media in which the information is hidden is called host. The stego object can be text, image, audio or video. When the information is hidden into the audio then it is called Audio steganography. Steganography is often confused with cryptology because the two are similar in the way that they both are used to protect important information. The difference between the two is that Steganography involves hiding information so it appears that no information is hidden at all. If a person or persons views the object that the information is hidden inside of he or she will have no idea that there is any hidden information, therefore the person will not attempt to decrypt the information.

The process of Steganography is as shown in Figure1.

Hiding information into a media requires following elements

• The cover media(C) that will hold the hidden data
• The secret message (M), may be plain text, cipher text or any type of data
• The stego function (Fe) and its inverse (Fe-1)
• An optional stego-key (K) or password may be used to hide and unhide the message.

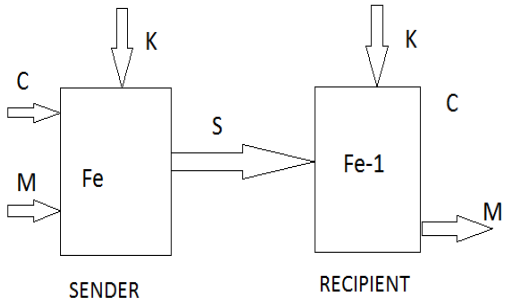

Figure (1) The Steganography Operation

The stego process fE embeds the secret message M in the cover media C. The exact position (S) where M will be embedded is dependence on the key K. The result of the stego function is slightly modified version of C: the stego data C'. After the recipient has received C' he starts the extracting process fE-1 with the stego data C' and the key K as parameters. If the key that is supplied by the recipient is the same as the key used by the sender to embed the secret message and if the stego data the recipient uses as input is the same data the sender has produces (i.e., it has not been modified by an adversary), then the extracting function will produce the original secret message M.

Because the size of the information is generally quite small compared to the size of the data in which it must be hidden (the cover text), electronic media is much easier to manipulate in order to hide data and extract messages.

Secondly, extraction itself can be automated when the data is electronic, since computers can efficiently manipulate the data and execute the algorithms necessary to retrieve the messages. Because degradation in the perceptual quality of the cover object may leads to a noticeable change in the cover object which may leads to the failure of objective of steganography.

If in one application we want to achieve confidentiality, than we have two alternatives: encryption or steganographic techniques for protection against detection

2. Audio Steganography

Like the document images, the sound files may be modified in such a way that they contain hidden information, like copyright information; those modifications must be done in such a way that it should be impossible for a pirate to remove it, at least not without destroying the original signal. The methods that embeds data in sound files use the properties of the Human Auditory System (HAS). The HAS perceives the additive random noise and also the perturbations in a sound file can also be detected. But there are some "holes" we can exploit. While the HAS have a large dynamic range, it has a fairly small differential range.

3. Technique for Data Hiding in Audio

There are four techniques for hiding data in Audio as following:

(i) Least Significant Bit (LSB) Encoding:

When files are created there are usually some bytes in the file that aren't really needed, or at least aren't that important. These areas of the file can be replaced with the information that is to be hidden, without significantly altering the file or damaging it. This allows a person to hide information in the file and make sure that no human could detect the change in the file. The LSB method works best in Picture files that have a high resolution and use many different colors, and with Audio files that have many different sounds and that are of a high bit rate. The LSB method usually does not increase the file size, but depending on the size of the information that is to be hidden inside the file, the file can become noticeably distorted. The main advantage of the LSB coding method is a very high watermark channel bit rate; use of only one LSB of the host audio sample gives capacity of 44.1 kbps (sampling rate 44 kHz, all samples used for data hiding) and a low computational complexity. The obvious disadvantage is considerably low robustness, due to fact that simple random changes of the LSBs destroy the coded watermark.

A very popular methodology is the LSB (Least Significant Bit) algorithm, which replaces the least significant bit in some bytes of the cover file to hide a sequence of bytes containing the hidden data. That's usually an effective technique in cases where the LSB substitution doesn't cause significant quality degradation, such as in 24-bit bitmaps. In computing, the least significant bit (LSB) is the bit position in a binary integer giving the units value, that is, determining whether the number is even or odd. The LSB is sometimes referred to as the right-most bit, due to the convention in positional notation of writing less significant digit further to the right. It is analogous to the least significant digit of a decimal integer, which is the digit in the ones (right-most) position.

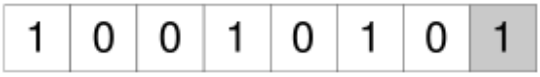

Figure (2) Binary Representation of Decimal 149

The binary representation of decimal 149, with the LSB highlighted. The MSB in an 8-bit binary number represents a value of 128 decimal. The LSB represents a value of 1. For example, to

hide the letter "a" (ASCII code 97, which is 01100001) inside eight bytes of a cover, you can set the
LSB of each byte like this:
10010010
01010011
10011011
11010010
10001010
00000010
01110010
00101011
The application decoding the cover reads the eight Least Significant Bits of those bytes to recreate the hidden byte—that is 0110001—the letter "a." As you may realize, using this technique let you hide a byte every eight bytes of the cover. Note that there's a fifty percent chance that the bit you're replacing is the same as its replacement, in other words, half the time, the bit doesn't change, which helps to minimize quality degradation.

(ii) 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. Two versions of SS can be used in audio steganography: the direct-sequence and frequency-hopping schemes. In direct sequence SS, the secret message is spread out by a constant called the chip rate and then modulated with a pseudorandom signal. It is then interleaved with the cover-signal. I frequency-hopping SS, the audio file's frequency spectrum is altered so that it hops rapidly between frequencies.
The SS method has the potential to perform better in some areas than LSB coding techniques in that it offers a moderate data transmission rate while also maintaining a high level of robustness against removal techniques. However, the SS method shares a disadvantage with LSB in that it can introduce noise into a sound file.The fig (3) shows the spread spectrum.

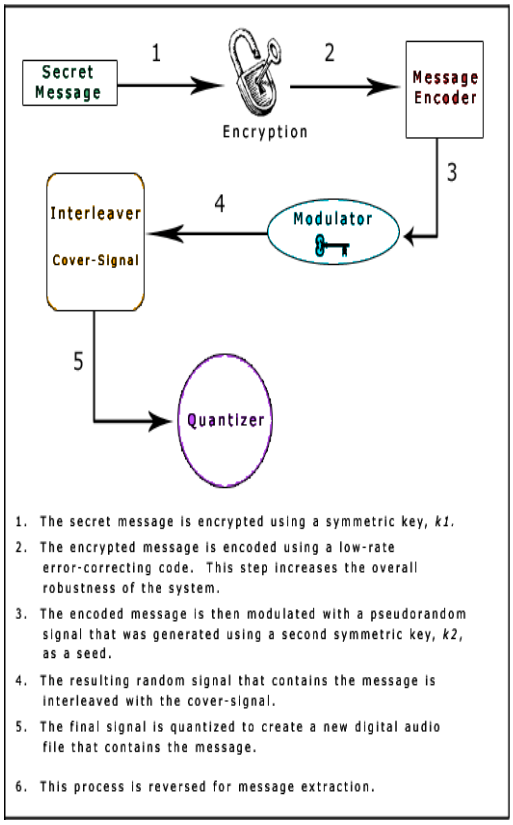

Figure (3) Spread Spectrum

(iii) Echo Hiding

Echo data hiding is yet another method of hiding information inside an audio file. This method uses the echoes in sound files in order to try and hide information. By simply adding extra sound to an echo inside an audio file, information can be concealed. The thing that makes this method of concealing information inside of audio files better than other methods is that it can actually improve the sound of the audio inside an audio file.

4. Proposed Technique

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 lsb hiding 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. lsb 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:

a. Covert the audio file into bit stream.

b. Convert each character in the secret information into bit stream.

c. 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.

Experiment

This Steganpgraphy is implemented in Matlab 7. To measure the performance of proposed method , MOS (Mean Opinion Score strategy is used. Mean value is calculated by asking people about the difference in the original wav file and embedded wav file.This rating is done on 5 point scale. The LSB algorithm is tested for 1, 2, 3, 4, 5, 6 , 7 LSB Position.

Generally, a steganographic message will appear to be something else: a special music.The detection of steganographically-encoded packages is called steganalysis. The simplest method to detect modified files, however, is to compare them to the originals. To detect information being moved through the sound on a website, for example, an analyst can maintain known-clean copies of these materials and compare them against the current contents of the site. The differences (assuming the carrier is the same) will compose the payload. The message shall be composed of some lines written by any human and they will then be converted to voice for transferring to any other system. The primary reason for converting the text in voice into wave is that no machines will be able to verify the lines of code. Furthermore these will be saved in the database in the form of binary data. The data should also be accompanied by the wave that shall be used to break the code through or else receiver of the message will not be able to get any. There are many mechanisms by which the data can be made secure but then there are intruders that can retrieve them too. To send a message, a source text, a wave in which the text should be embedded. The information should be hidden in the wave. To receive a message, a source wave file containing the information is required. The result will appear in the text tab after decoding. This solution would be eventually implemented on a windows application.

MATLAB

MATLAB stands for "MATrix LABoratory" and is a numerical computing environment and fourth-generation programming language. Developed by The Math Works, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, and Fortran In 2004, Math Works claimed that MATLAB was used by more than one million people across the industry and the academic world. MATLAB users come from various backgrounds of engineering, science, and economics. Among these users are academic and research institutions such as Massachusetts Institute of Technology, NASA, Max Planck Society, and RWTH Aachen University as well as industrial enterprises such as ABB Group, Boeing, Ford Motor, Halliburton, Lockheed Martin, Motorola, Novartis, Pfizer, Philips, Toyota, and UniCredit Bank.

Data Flow Diagram

Figure (4) 0 Level DFD

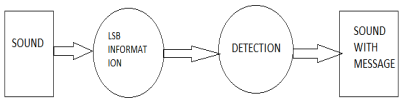

Figure (5) 1 Level DFD

Implementation

H = AUDIOSTEGANO returns the handle to a new AUDIOSTEGANO or the handle to the existing singleton*.AUDIOSTEGANO('CALLBACK',h Object, eventData, handles,...) calls the local function named CALLBACK in AUDIOSTEGANO.M with the given input arguments.AUDIOSTEGANO('Property','Value', ...) creates a new AUDIOSTEGANO or raises the existing singleton*. Starting from the left, property value pairs are applied to the GUI before audiostegano_Opening Function gets called. An unrecognized property name or invalid value makes property application stop. All inputs are passed to audiostegano_OpeningFcn via varargin.

How it works

#The details of wav file format is in the web page.
#Description of audiostegano.m

Hide text

The code to hide text starts from line no.173 to line no.232 First a wav file is opened using fopen. Then the header part (first 40 bytes) of wav file is copied to variable "header". Byte no. 41 contains 32 bit data size of file this is copied to variable data_size.Then data samples starts from byte no.44 to the end. Copy the data samples to variable "dta" using fread and close the file. Convert the text message and the length of text message to binary. Hide three content in data samples indentifier, length of text message and text message itself. Indentifier helps in the recovery of text. If identifier is not in the file then code stop execution assuming that the wav file has no hidden text message. Here the identifier is binary 10101010. Lsb of first 8 data samples have the indentifier. Lsb of next 10 data samples have length of text message. Lsb of next 10 data samples have width of text message. Then all the lsb of data samples after this have bits of binary text message. Open a new wav file in write mode,copy the header original wav file to this file,then copy the 32 bit data_size to the new file. At last the copy "dta" the data samples containing the hidden text message bits to the new file.Close the file. The new wav file is in the current directory.

Recovery Test

The code to recover text starts from line no.118 to line no.153 The procedure to recover text is reverse of hidding the text. Open the wav file. Then directly copy the data samples starting from 44th byte to the end of file. Take out the lsb of first 8 data samples and check for the identifier 10101010.If identifier is not present then the file has no hidden text. Take out next 10 lsb which is the length of text message and next 10 lsb which is the width of text message. Now take out the lsb of data samples upto the length of message.Convert to text and then reshape them.

Control Flow Diagram

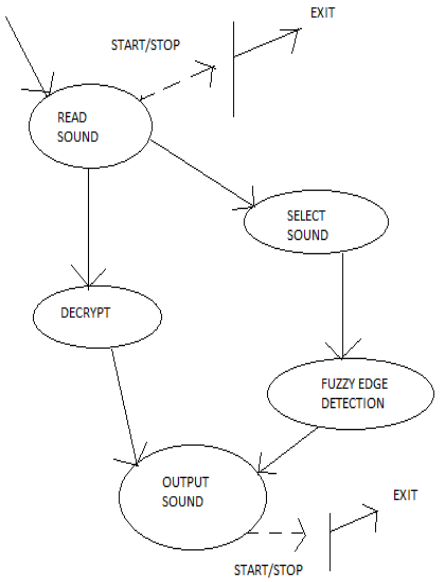

Figure (6) Control Flow Diagram

5. Conclusions

In this paper, I have introduced a robust method of imperceptible audio data hiding. This system is to provide a good, efficient method for hiding the data from hackers and sent to the destination in a safe manner. This proposed system will not change the size of the file even after encoding and also suitable for wav type of audio file format. 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. Audio Steganography is a really interesting subject and outside of the mainstream cryptography and system administration that most of us deal with day after day. But it is also quite real; this is not just something that's used in the lab or an arcane subject of study in academia. Stego may, in fact, be all too real — there have been several reports that the terrorist organization attacks in New York City, Washington, D.C., and outside of Pittsburgh used audio steganography as one of their means of communication.

Future Enhancement

Though it is well modulated system, it has been limited to certain restrictions. The quality of sound depends on the size of the audio which the user selects and length of the message.
The quality of the sound in the encoded audio file can be increased. There are number of ways that this paper could be extended. Its performance can be upgraded to higher levels in practical conditions. There are also other weighting algorithms like spread spectrum, echo data hiding etc., and those can be implemented.
Instead of having common secret key to encode and decode, a public–private key pairs will be introduced.

References

[1] A Study of Steganography and the Art of Hiding Information Alain C. Brainos II East Carolina University
[2] Audio Steganography http://en.wikIPedia.org/wiki/ Audio _Steganography
[3] Efficient Method of Audio Steganography by Modified Lsb Algorithm and Strong Encryption Key with Enhanced Security 1r Sridevi, 2dr. A Damodaram, 3dr. Svl.Narasimham Assoc. Prof., Department Of Computer Science And Engineering, Jntuceh, Hyderabad Prof., Department Of Computer Science And Engineering, Jntuceh, Hyderaba Prof., School Of Information Technology, Jntuh, Hyderabad.
[4] Steganography FAQ Aelphaeis Mangarae [Zone-H.Org] March 18th 2006
[5] Study on Information Hiding Algorithm Based on RBF and LSB
[6] Kaili Zhou1,2, Taifan Quan1, Yaohong Kang2, Research Inst. of Electronic Engineering; Harbin Inst. of Tech.; Harbin; 150001; China College of Information Science & Tech., Hainan Univ., Haikou 570228, China
[7] Data Hiding in Audio signal – A review International journal of database theory and application, 2 June 2009
[8] MATLAB http://www.mathworks.com/academia/student_center/tutorials/ http://www.mathworks.com/access/helpdesk_r13/help/techdoc/ref/ref.html