International Journal of Advanced Research and Development
ISSN: 2455-4030
Impact Factor: RJIF 5.24
www.advancedjournal.com
Volume 3; Issue 1; January 2018; Page No. 460-467

Through Stegnalysis Identification of Extreme Data in Social Media Through WhatsApp Transmitted

[1] Nishtha Pateriya, *[2] Dr. Tripti Arjariya
[1] M-Tech, Bhabha Group of Institute Bhopal, Madhya Pradesh, India
[2] HOD of (CSE), Bhabha Group of Institute Bhopal, Madhya Pradesh, India

Abstract
Today WhatsApp is a wide user base social media so numbers of attacker use it to send hidden data through it. It is possible to design an audio steganalyzer based on audio quality measures. Several, seemingly redundant, features need to be used. Apparently these diverse features probe different aspects of the watermarked signals in order to differentiate between clear-objects and stego-objects. In this research paper region-adaptive steganography algorithm which will be used for the novel technic to detect steganography attacks. We propose the major advantages of the proposed steganography detection technique is Spread Spectrum. A distortion metric based on Signal spectrum was designed specifically to detect modifications and additions to audio media. We used the Signal spectrum to measure the distortion. The distortion measurement was obtained at various wavelet decomposition levels from which we derived high-order statistics as features for a classifier to determine the presence of hidden information in an audio signal. In this study, an audio Steganalysis technique is proposed and tested. The objective audio quality measures, giving clues to the presence of hidden messages, are searched thoroughly. In this paper a new algorithm proposed for detection hidden data from the original Audio File transmitted through WhatsApp and has little relation with secret message file. It was providing more security to the information. After analysis our algorithm success rate varies from to 80% for combined active warden and passive warden problems to 100% for certain single methods.

Keywords: Echo Hiding discrete wavelet transform and least significant bit (LSB) coding technique, Audio File steganography, Audio File encryption, Linear Classifier, Message Encryption

1. Introduction
In the watermarking context, some copyright or copy control information is embedded into the cover/host audio signals in order to prove the ownership of the cover object or preserve unauthorized copying of it. An audio steganography technique can be classified into two groups based on the domain of operation. One type is time domain technique and the other is transformation based method.
In addition to this, the watermarking can also be used for various other applications mentioned above.
Detection of the hidden information by untrusted parties and reliable and/or correct watermark extraction are two major problem areas in watermarking. Spread spectrum watermarking has been proposed as a solution to the latter problem. In spread spectrum watermarking, the embedded message is spread over very many samples or frequency bins so that the energy in one sample or bin is very small. In this system even missing some embedded samples one can still reconstruct the embedded message. Besides reliable detections, this also causes small modifications of host samples so that the distortions will be imperceptible. The former problem is that, the hidden message should not be detected or revealed by untrusted parties or adversaries. That brings forward to the security property of watermarking. The current research issues in secure watermarking methods based on key dependent embedding. Thus the embedded signal depends on a secret key as the threat model, a malicious adversary, could not reveal the watermark content or invalidate it. In the watermarking context, we always assume that the adversary knows that the content is watermarked and, in principle, also knows the exact technique used for watermarking.
The only thing she does not know is the secret key which, for example, can be used to disperse the watermark locations in an image. Besides this unauthorized detection, unauthorized embedding is another security issue in watermarking. The adversary can embed some fake watermarks or extract the watermark from a marked object and embed it to other objects in order to fool the system. Key-dependent watermarking could be a solution of fake embedding but cannot solve problem of copying the embedded watermark into other objects. Therefore content dependent keying or watermarking has been studying as a solution of unauthorized copying of watermarks. One of the concerns of this thesis is designing such a tool.
In steganography, the very existence of the message is secret. It ideally suited for covert communication. In this context, the host object is used in order to mask the very existence of the communicating secret information. Therefore the adversary does not and should not know that there is a secret message embedded in the content. Ideally, the information should be embedded in a way that, the distortions on cover object should not be perceivable by human sensory systems. In fact, the modern formulation of steganography goes by the name of the prisoner's problem. Here Alice and Bob are in prison, and a warden, Wendy, who will punish them at the first hint of any

suspicious communication, examines all communication between them. Hence, Alice and Bob must trade seemingly inconspicuous messages that actually contain hidden messages.

Specifically, in the general model for steganography, we have Alice wishing to send a secret message m to Bob. In order to do so, she "embeds" m into a cover-object c, to obtain the stego-object s. The stego-object s is then sent through the public channel.

The general requirements for data hiding are robustness, imperceptibility and security. Robustness means that the hidden data should survive after standard data manipulations and intentional attacks. Security means that detecting or removing the hidden data is impossible even when the exact algorithms for embedding and extracting of the watermark are known. Using a private key for watermark generation enables security.

However, one of the most significant problems, which affect the commerce of digital media, is how to protect copyright and ownership. Digital steganography, one of the popular approaches considered as a tool for providing the copyright protection, is a technique based on embedding a specific mark or signature into the digital products. While several steganography algorithms have been proposed [1], transform domain schemes, such as discrete wavelet transform (DWT) based steganography have shown more advantages and provide higher performance than others. Steganalysis is the discovery of the existence of hidden information; therefore, like cryptography and cryptanalysis, the goal of Steganalysis is to discover hidden information and to break the security of its carriers [2].

Using of transformation based techniques provides additional information about the signal. In general, the time domain techniques provide least robustness as a simple low pass filtering can remove the watermark. Hence time domain techniques are not sensible for the applications such as copyright protection and airline traffic monitoring; however, it can be used in applications like proving ownership and medical applications [3].

Spreads spectrum technology to improve the robustness, simultaneous reduces the performance of anti-synchronization attack. In the field of digital audio steganography, the idea is to use the stable feature points of the audio to mark the embedded position of the steganography, and use the stable performance of these feature points anti-synchronized attacks to improve the ability of the steganography anti-synchronization attack. Feature points should have the feature such as stability, more uniform distribution and the ability to accommodate the steganography [4].

Progress in this area has been steady as can be seen from a healthy number of publications in the field and the sheer number of institutes around the world that deal with the issue [5]. In the more specific field of digital Audio File steganography, one of the most notable techniques is region-based Audio File steganography [6]. The paper described a method for embedding and detecting chaotic steganography's in large Audio Files. An adaptive clustering technique is employed in order to derive a robust region representation of the original Audio File. The robust regions are approximated by ellipsoids, whose bounding rectangles are chosen as the embedded area for the steganography. The drawback of this technique is due to limited number of suitable regions for storing the steganography the steganography storing capacity can be low.

Related work

In this Research work [7] a hidden message is information that is not immediately noticeable, and that must be discovered or uncovered and interpreted before it can be known. A cover audio object can be converted into a stego-audio object via steganography methods. In this study we present statistical method to detect the presence of hidden messages in audio signals. The basic idea is that, the distribution of various statistical distance measures, calculated on cover audio signals and on stego-audio signals are statistically different. Distortion measure plays an important role in audio Steganalysis the analysis and classification method of determining if an audio medium is carrying hidden information The design of audio steganalyzer relies on the choice of these audio quality measures and the construction of two-class classifier. In this paper, we propose distortion metric proposed technique can be used to detect the presence of hidden messages in digital audio data.

In this Research work [8] the digital information revolution has brought important changes in our society and life. Nowadays, large amount of data is transmitted over the network and if the data that is being transmitted is important, one should use secure technique like steganography to transmit it. Steganography is a method of hiding a secret message in a cover media such as text, image, audio etc. in a way that hides the existence of the secret data. This paper introduces new method for audio steganography. The proposed method works on the basis of low bit blind encoding scheme which is used to embed secret data into non-silent samples of wav audio file. Robustness and performance of the proposed scheme is investigated by performing experiments on different audio signals.

In this Research work [9] Steganography is the art and science of hiding the existence of information. In computer based steganography, several forms of digital media may be used as "cover" for hidden information. Photos, documents, web pages, and even MP3 music files may all serve as innocuous-looking hosts for secret messages. Multimedia documents are very easy to copy and distribute in an illicit manner. Copyright labeling is a process that may help to reduce their illicit copying. If this document is copied the copy will also contain the label. This label (or watermark) should be robust enough to withstand normal image processing activities (like image compression, transforming to different format) that do not significantly alter the image appearance. In this paper we investigate the relevant concepts and terminology. Information hiding and its applications, and image compression using proposed efficient encoding technique with the main focus being on hiding in the spatial domain. Three information hiding methods are proposed, which are based on the encoding technique, are tested and the results are analyzed.

In this Research work [10] Psychoacoustic models are routinely used in audio watermarking algorithms to adjust the changes induced by the watermarking process to the sensitivity of the ear. The performances of such models in audio watermarking

applications are tightly related to the determination of tonal and noise-like components. In this paper, we present improved tonality estimation and its integration into a psychoacoustic model. Instead of conventional binary classification, we exploit bi-modal prediction for more precise tonality estimation. Experimental results show improved robustness of the considered audio watermarking algorithm integrating the new tonality estimation, while preserving the high quality of the audio track.

Proposed algorithm

The proposed steganography attack detection scheme requires the certain threshold of original Audio Files and steganography Audio File.

Steganography is an information hiding technique where secret message is embedded into unsuspicious cover signal. An effective audio stenographic scheme should possess the following three characteristics: Inaudibility of distortion (Perceptual Transparency), Data Rate (Capacity) and Robustness. These characteristics (requirements) are called the magic triangle for data hiding. We have presented an audio Steganalysis algorithm based on the generalized moments of the denoising residuals of speech and audio signals.

If the embedding method is known ahead, the steganalyzer yields very satisfactory detection results, that is, the average success rate ranges. Finally, in the absence of any knowledge, that is if we are uncertain which of the eight watermarking or steganography methods has been used, the correct detection probability becomes. Some content dependency has been observed; in fact, the steganalyzer is more successful with speech cover material as compared to the tested music varieties. Finally, there is a critical strength threshold below, which Steganalysis of watermarking methods is not possible and there is a critical capacity threshold below which Steganalysis of steganography method is not possible.

Low-bit Encoding

In Low-bit encoding [11], the binary version of the secret data message is substituted with the least significant bit (LSB) of each sample of the audio cover file. Though this method is simple and can be used to embed larger messages, the method cannot protect the hidden message from small modifications that can arise as a result of format conversion or damage compression.

Phase Coding

Phase coding [12] is based on the fact that the phase components of sound are not as perceptible to the human ear as noise is. Message bits are encoded as phase shifts in the phase spectrum of a digital signal. This leads to inaudible encoding in terms of the Signal-to-Perceived Noise Ratio (SPNR) and the secret message gets camouflaged in the audio signal, not detectable by the Steganalysis methods based on SPNR. Thus, phase coding addresses the disadvantages of the noise-inducing methods of audio steganography.

The sequence of steps involved in phase coding is as follows:

The original audio signal is decomposed into smaller segments such that their length equals the size of the message that needs to be encoded.

A Discrete Fourier Transform (DCT) is then applied to each segment in order to create a phase matrix.

Phase differences between every pair of consecutive segments are computed.

Phase shifts between adjacent segments are identified. Although, the absolute phases of the segments can be altered, the relative phase differences between consecutive segments must be unchanged.

The new phase matrix is created using the new phase of the signal's first segment and the set of original phase differences.

Based on the new phase matrix and the original magnitude matrix, the sound signal is regenerated by using inverse DFT and then by joining the sound segments together.

The receiver is mandated to know the message length in order to use DFT and extract the embedded message from the cover signal.

A characteristic feature of phase coding is the low data transmission rate owing to the fact that the secret message is encoded only in the first segment of the audio signal. On the contrary, an increase in the length of the segment would have a ripple effect by altering the phase relations between the frequency components of the segment; thereby making detection easier. Hence, the phase coding method is normally used only when a small amount of data (e.g., watermark needs to be masked).

Echo Hiding

With echo hiding (e.g. [13]), information is embedded by introducing an echo into the discrete audio signal. Like SS coding, echo hiding allows for a higher data transmission rate and provides superior robustness when compared to the noise-inducing methods. To successfully hide the data, three parameters of the echo need to be altered: amplitude, decay rate and offset (delay time) from the original signal. The echo is not easily resolved as all the three parameters are set below the human audible threshold limit. Also, the offset is altered to represent the binary message to be hidden. The first offset value represents a one (binary), and the second offset value represents a zero (binary).

Spread Spectrum Coding

The basic Spread Spectrum (SS) coding method [14] randomly spreads the bits of the secret data message across the frequency spectrum of the audio signal. However, unlike LSB coding, the SS coding method spreads the secret message using a code that is independent of the actual cover signal. The SS coding method can perform better than LSB coding and phase coding techniques by virtue of a moderate data transmission rate coupled with a high level of robustness against Steganalysis techniques. However, like the LSB coding method, the SS method can introduce noise to the audio file. This vulnerability can be tapped for Steganalysis.

In the proposed algorithm have presented a high capacity and high stego-signal quality audio steganography scheme based on samples comparison in Echo Hiding discrete wavelet transform and least significant bit (LSB) coding technique of a segment are compared with pre-determined threshold value T and based on comparison bits are embedded. The strength of our algorithm is depend on the segment size and their strength are enabled the algorithm to achieve very high embedding capacity for different data type that can reach up to 25% from

the input audio file size with lest of 35 dB SNR for the output stego signal.

Experimentation and results

We are take a plain audio file and check it signal spectrogram, frequency response, pole-zero, impulse response and step response. We also plot graph between Time and amplitude. We are not get any mixed sound or distraction.

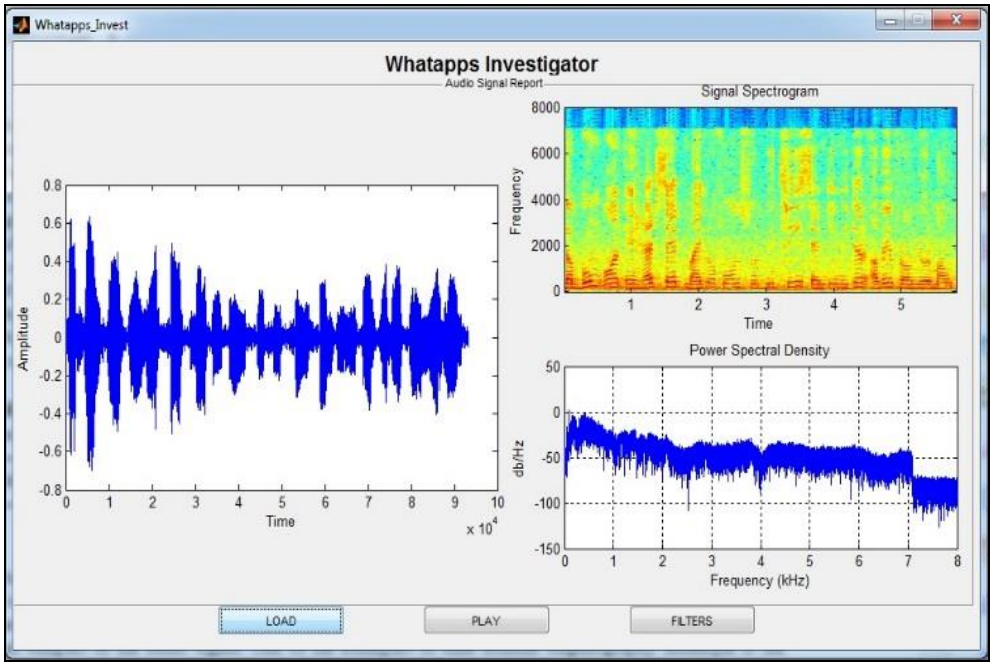

Fig 1: A signal spectrogram graph plot of original and plain audio file

First we have taken original audio file without any hidden massage. And apply different method to check hidden file.

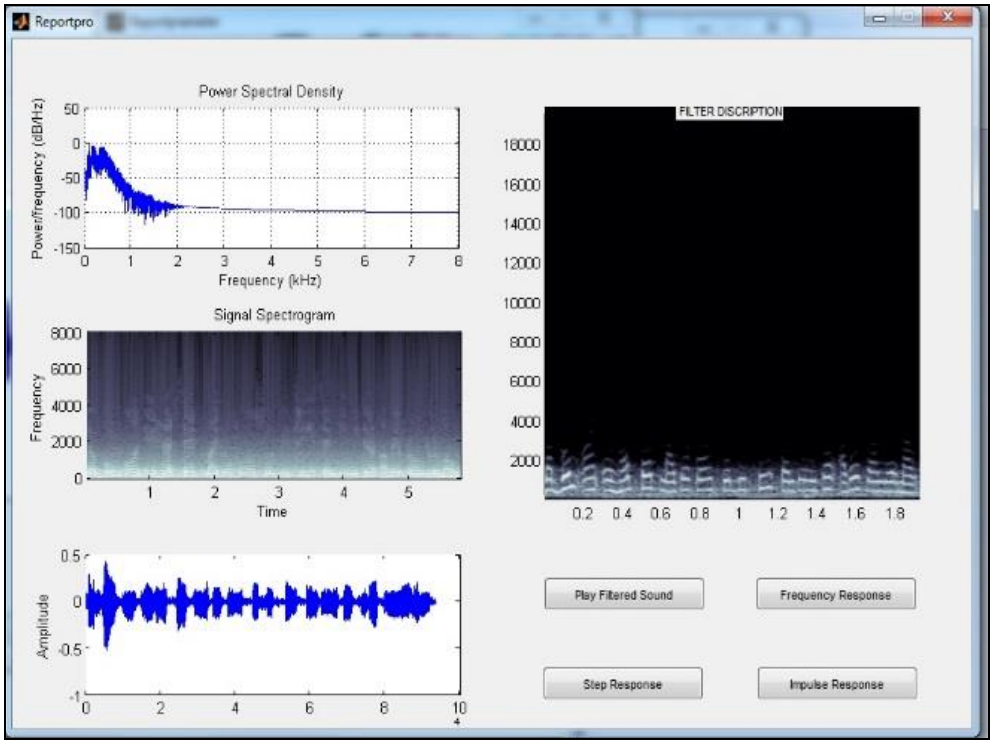

Fig 2: A complete report plot of original and plain audio file

We are taken audio file with the including of hidden massage and check it signal spectrogram, frequency response, pole-zero, impulse response and step response. We also plot graph between Time and amplitude. We got sound distraction.

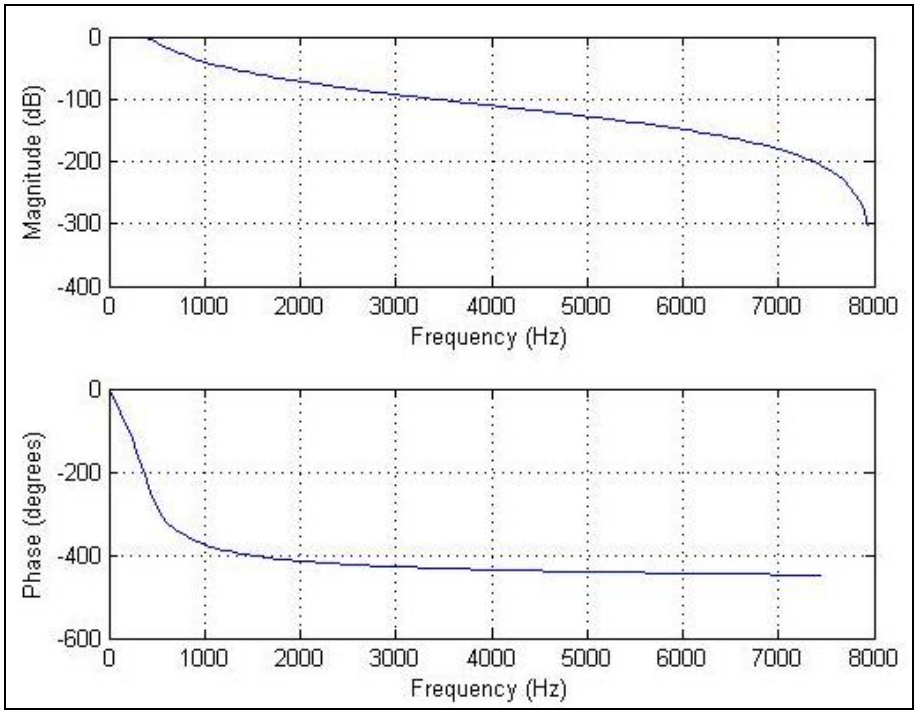

Fig 3: A frequency response graph plot of original and plain audio file

We have taken audio file with hidden massage. Check its audio quality for finding of hidden massage.

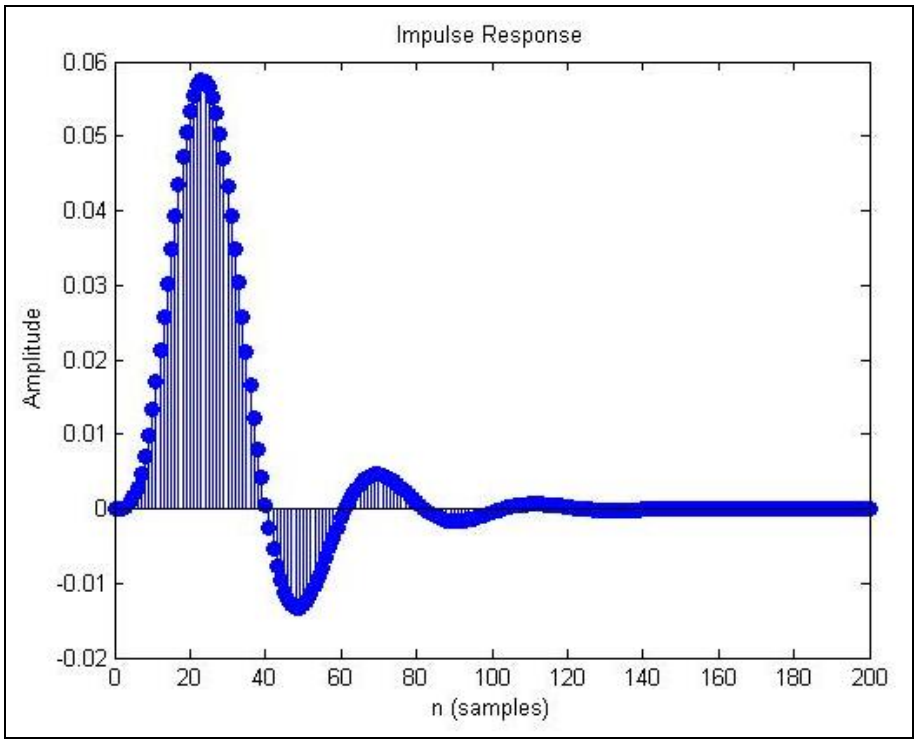

Fig 4: A graph plot of Impulse Response with original and plain audio file

In this paper will analysis of some WhatsApp audio file in which some data are hidden. But we detect the presence of steganography programs, detect suspect carrier files, and disrupt stegano-graphically hidden messages.

The detection of steganography file on a suspect computer is important to the subsequent forensic analysis. As the research shows, many steganography detection algorithm work best when there are clues as to the type of steganography that was employed in the first place.

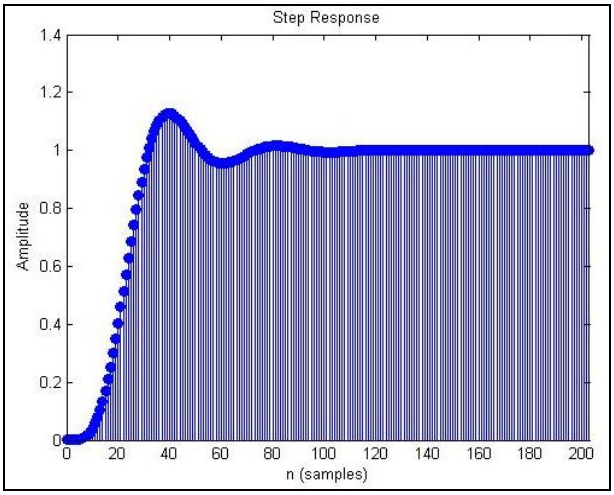

Fig 5: A graph plot of Step Response with original and plain audio file

Finding stegano file on a computer would provide growth to the doubt that there are actually steganography files with secreted messages on the suspect computer. This paper has been tested on 10 audio file. Our algorithms was able to identify the presence of hidden messages with 65 percent precision with a false-positive rate less than 0.001 percent.

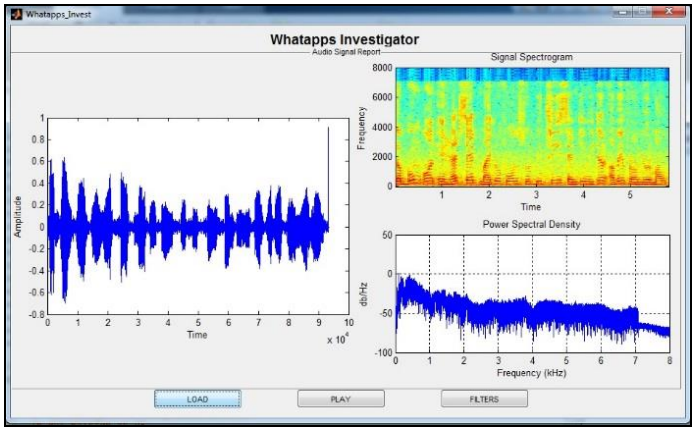

Fig 6: A graph plot of Signal Spectrogram of audio file with hidden massage

Finding stegano message in a file suspected to cover it is relatively easy compared to removing hidden data. Steganography discovery patterns do not directly help in the retrieval of the password. Discovery appropriate clues is where the rest of the examination and computer forensics comes into play.

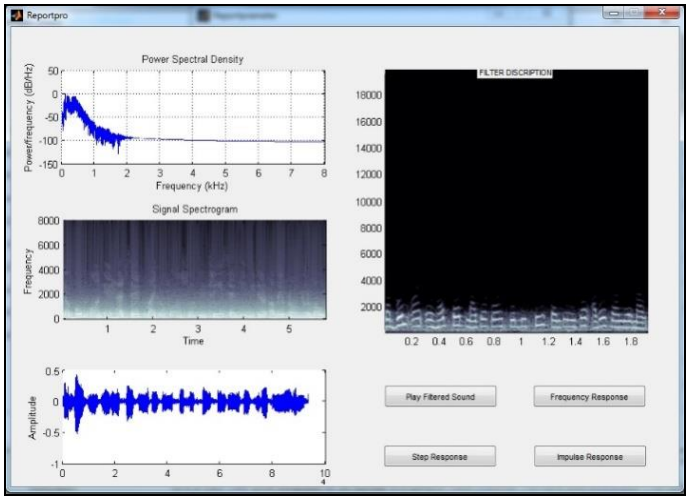

Fig 7: A complete report plot of audio file with hidden massage

This paper looking at evidence in a criminal case probably has no reason to alter any evidence files. However, an examination that is part of an ongoing terrorist surveillance might well want to disrupt the hidden information even if it cannot be recovered. Secreted content, which is steganography and digital watermarks, can be attacked in numerous ways so that it can be removed or changed and there is software exactly intended to attack audio file and send WhatsApp. Such attacks have one of two possible effects they either reduce the steganography carrying capacity of the carrier (necessary to avoid the attack) or fully disable the capability of the carrier as a steganography medium.

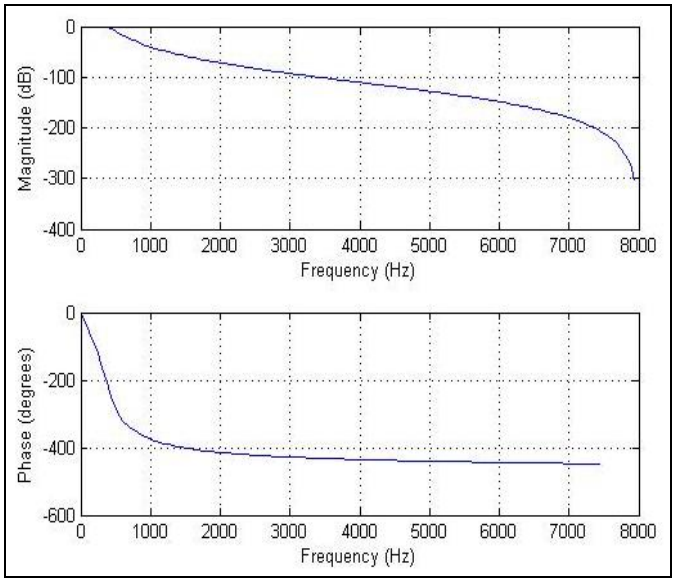

Fig 8: A graph plot of Frequency Response of audio file with hidden massage

In Figure 1, the impacts of some attacks on original wave sound are presented. In the figure the attacks can be deduced from the figure that the attacks generate visible distortions and the distortions on the wave shapes can easily be observed.

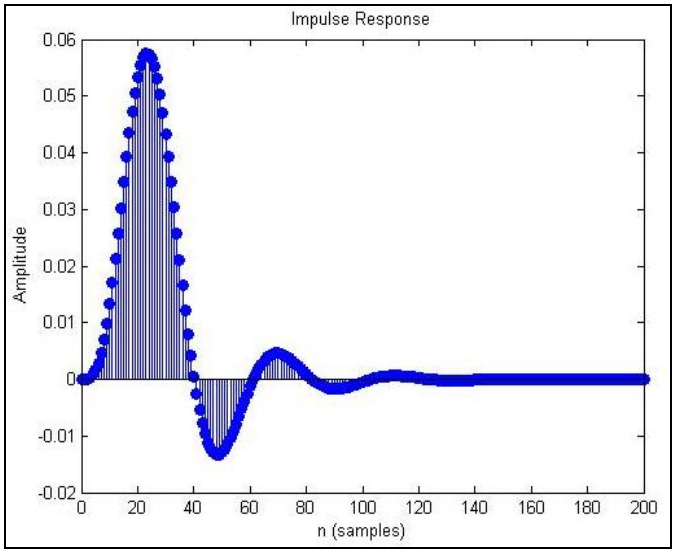

Fig 9: A graph plot of Impulse Response of audio file with hidden massage

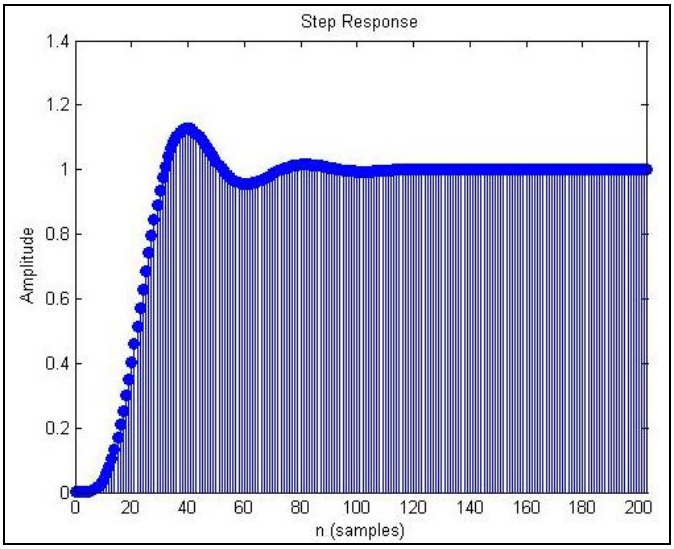

Fig 10: A graph plot of Step Response of audio file with hidden massage

Our Proposed Algorithm is able to detect any type of attack if applied in stegano graphy Audio File. And improves the speed of detection, and also test the robustness of the Audio Files.

Conclusion

We have proposed in this Research paper a combination of Echo Hiding discrete wavelet transform and least significant bit (LSB) coding technique is also a steganography embedded technique on different regions of the host Audio File approach. To increase the reliability of the LSB, DWT and Echo Hiding based stegano detection in WhatsApp. The audio Steganalysis algorithms exploit the variations in the characteristic topographies of the audio signal as a result of message embedding. Audio Steganalysis algorithms that detect the discontinuities in phase (as a result of phase coding), differences in the amplitude (as a result of Echo hiding) and the changes in the perceptual and non-perceptual audio superiority as a result of message embedding have been planned. In summary, each carrier media has its own special attributes and reacts differently when a message is embedded in it. Therefore, the Steganalysis algorithms have also been developed in a manner specific to the target stego file and the algorithms developed for one cover media are generally not effective for a different media. Our amixture of Echo Hiding discrete wavelet transform and least significant bit (LSB) coding method is also a steganography embedded technique on different regions of the host Audio File steganography technique is appreciated by using two steganography Audio Files, each with a strong High Frequency or Low Frequency components. Non overlapping regions of these steganography Audio Files are inserted into the host Audio File using a combination of Audio File segmentation. The experimental results will performed and analyze of different Audio Files file is implemented in Matlab tool.

Reference

1. Voyatzis G, Nikolaidis N, Pitas I. Digital steganography: An overview, EUSIPCO, 1998; 1:9-12.
2. Deborah Radcliff. Computer World URL:

http://www.computerworld.com/securitytopics/security/story/0,10801,71726,00.html.
3. Mazdak Z, Azizah AM, Rabiah BA, Akram MZ, Shahidan A. A Secure audio steganography approach, World Acad. Sci. Eng., Technol., 2009; 52:360-363.
4. Wang HX. Overview of content based adaptive audio steganography. Journal of Southwest Jiao tong University. 2009; 44(3):430-437.
5. Petitcolas FAP, Anderson RJ, Kuhn MG. Information Hiding-A survey, Proceeding of the IEEE, Special Issue on Protection of Multimedia Content, 1999, 1062-1078.
6. Nikolaidis A, Pitas I. Region-based Audio File steganography, Audio File Processing, IEEE Transactions on, 2001; 10(11):1726-1740.
7. Er. Niranjan Singh, Dr. Bhupendra Verma. Steganalysis of Audio Signals, Audio Quality and Distortion Measures ICCET - International Conference on Computer Engineering and Technology CET6011.0.607 ISBN No 978-81-920748-1-8, 2010.
8. Mayank Srivastava, Mohd Qasim Rafiq. A Novel Approach to Secure Communication Using Audio Steganography, Advanced Materials Research, 20011; (403-408):963-969.
9. Dilip Vishwakarma1, Prof. Satyam Maheshwari2, Prof. Sunil Joshi3. Efficient Information Hiding Technique Using Steganography, International Journal of Emerging Technology and Advanced Engineering ISSN 2250-2459, 2012; 2(1).
10. Michael Arnold, Xiao-Ming Chen, Peter G. Baum and Gwenaël Doërr, Improving Tonality Measures for Audio Watermarking, Lecture Notes in Computer Science, 2011; 6958/2011:223-237.
11. Kitawaki N, Nagabuchi H, Itoh K. Objective Quality Evaluation for Low-Bit-Rate Speech Coding Systems, IEEE J Select. Areas Commun, 1988; 6:242-248.
12. Wang S, Sekey A, Gersho A. An Objective Measure for Predicting Subjective Quality of Speech Coders, IEEE J Select. Areas Commun, 1992; 10:819-829.
13. Zwicker E, Fastl H. Psychoacoustics Facts and Models, Springer-Verlag, 1990.
14. Yang W, Dixon M, Yantorno R. A Modified Bark Spectral Distortion Measure Which Uses Noise Masking Threshold, IEEE Speech Coding Workshop, Pocono Manor, 1997, 55-56.