IMPROVED SPREAD SPECTRUM: A NEW METHOD FOR INFORMATION HIDING AND OPTIMUM DETECTION STRUCTURE

#1Duba.Sriveni- M.Tech Pursuing,
#2B.Sevanaik -Associate Professor,
Department of Computer Science & Engineering,
MAHAVEER INSTITUTE OF SCIENCE & TECHNOLOGY, Hyderabad.

Abstract: Data hiding and extraction schemes are increasing in today‟s communication world due to rapid increment of data tracking and tampering attacks.So we need an efficient and robust data hiding schemes to protect from these attacks. In this project the blindly extraction technique is considered. Blindly extraction means the original host and the embedding carriers are not need to be known. Here,the hidden data embedded to the host signal,via multicarrier SS embedding. The hidden data is extracted from the digital media like audio, video or image. The extraction algorithm used to extract the hidden data from digital media is Multicarrier Iterative Generalized Least Squares (M-IGLS).It is a low complexity algorithm and it attains the probability of error recovery equals to known host and embedding carriers. It‟s peak signal to noise ratio value obtained is high.

Keywords: *Data hiding, Tracking, Tampering, Blindly extraction, Steganography and Watermarking Spread spectrum embedding.*

I. INTRODUCTION

Data tracking and tampering are rapidly increasing in everywhere like online tracking, mobile tracking etc. So we need a secured communication scheme for transmitting the data. Forthat, we are having many data hiding schemes and extraction schemes. Data hiding schemes are initially used in military communication systems like encrypted message, for finding the sender and receiver or it‟s very existence. Initially the data hiding schemes are used for the copy write purpose. In [1] Fragile watermarks are used for the authentication purpose ,i.e to find whether the data has been altered or not. Likewise the data extraction schemes also provides a good recovery of hidden data .This is the goal of the secured communication.

Before the invention of steganography and cryptography, it was challenging to transfer secure information and, thus, to achieve secure communication environment [1]

Normally an application is developed by a person or a small group of people and used by many. Hackers are the people who tend to change the original application by modifying it or use the same application to make profits without giving credit to the owner. It is obvious that hackers are more in number compared to those who create. Hence, protecting an application should have the significant priority. Protection techniques have to be efficient, robust and unique to restrict malicious users. The development of technology has increased the scope of steganography and at the same time decreased its efficiency since the medium is relatively insecure. This lead to the development of the new but related technology called "Watermarking‟. Some of the applications of watermarking include ownership protection, proof for authentication, air traffic monitoring, medical applications etc. [1] [5] [8].

A. *Steganography and Watermarking*

Steganography

Steganography is evolved from the ancient technique known as the "Cryptography". Cryptography protects the contents of the message [13]. On the other hand, steganography is a technique to send information by writing on the cover object invisibly. Steganography comes from the Greek word that means covered writing (stego = covered and graphy = writing) [3]. Here the authorized party is only aware of the existence of the hidden message. An ideal steganographic technique conceals large amount of information ensuring that the modified object is not visually or audibly distinguishable from the original object.

The steganography technique needs a cover object and message that is to be transported. It also requires a stego key to recover the embedded message. Users having the stego key can only access the secret message. Another important requirement for efficient steganographic techniques is that, the cover object is modified in a way that the quality is not lost after embedding the message.

Watermarking

Watermarking is a technique through which the secure information is carried without degrading the quality of the original signal. The technique consists of two blocks:

(i) Embedding block
(ii) Extraction block

The system has an embedded key as in case of a steganography. The key is used to increase security, which does not allow any unauthorized users to manipulate or extract data. The embedded object is known as watermark, the watermark embedding medium is termed as the original signal or cover object and the modified object is termed as embedded signal or watermarked data [13].

B. Image and Audio Watermarking

Watermarking technique has evolved considerably from its origin [8]. Due to evolution of technology the medium of transmission has been changed. Watermarking is employed in digital media such as image and audio. The watermarking technique, in which the cover objects as discussed in Section 1.1.2, is image (audio) then the process is termed as Image (Audio) Watermarking. Audio watermarking is quite challenging than image watermarking due to the dynamic supremacy of human auditory system (HAS) over human visual system (HVS) [12].

C. Requirements of efficient watermarking techniques

According to IFPI (International Federation of the Phonographic Industry) [4], audio watermarking algorithms should meet certain requirements. The most significant requirements are perceptibility, reliability, capacity, and speed performance [9].

Perceptibility: One of the important features of the watermarking technique is that the watermarked signal should not lose the quality of the original signal. The signal to noise ratio (SNR) of the watermarked signal to the original signal should be maintained greater than 20dB [4]. In addition, the technique should make the modified signal not perceivable by human ear.

Reliability: Reliability covers the features like the robustness of the signal against the malicious attacks and signal processing techniques. The watermark should be made in a way that they provide high robustness against attacks. In addition, the watermark detection rate should be high under any types of attacks in the situations of proving ownership. Some of the other attacks summarized by Secure Digital Music Initiative (SDMI), an online forum for digital music copyright protection, are digital-to-analog and analog-to-digital conversions, noise addition, band-pass filtering, time-scale modification, echo addition, and sample rate conversion [10].

Capacity: The efficient watermarking technique should be able to carry more information but should not degrade the quality of the audio signal. It is also important to know if the watermark is completely distributed over the host signal because, it is possible that near the extraction process a part of the signal is only available. Hence, capacity is also a primary concern in the real time situations [4].

Speed: Speed of embedding is one of the criteria for efficient watermarking technique. The speed of embedding of watermark is important in real time applications where the embedding is done on continuous signals such as, speech of an official or conversation between airplane pilot and ground control staff. Some of the possible applications where speed is a constraint are audio streaming and airline traffic monitoring. Both embedding and extraction process need to be made as fast as possible with greater efficiency [4].

Asymmetry: If for the entire set of cover objects the watermark remains same; then, extracting for one file will cause damage watermark of all the files. Thus, asymmetry is also a noticeable concern. It is recommended to have unique watermarks to different files to help make the technique more useful [4].

D. Applications of Watermarking

Ownership protection and proof of ownership: In ownership protection application, the watermark embedded contains a unique proof of ownership. The embedded information is robust and secure against attacks and can be demonstrated in a case of dispute of ownership. There can be the situations where some other person modifies the embedded watermark and claims that it is his own. In such cases the actual owner can use the watermark to show the actual proof of ownership [5] [11] [4].

Authentication and tampering detection: In this application additional secondary information is embedded in the host signal and can be used to check if the host signal is tampered. This situation is important because it is necessary to know about the tampering caused to the media signal. The tampering is sometime a cause of forging of the watermark which has to be avoided [5] [11] [4].

Finger printing: Additional data embedded by a watermark in the fingerprinting applications are used to trace the originator or recipients of a particular copy of a multimedia file. The usage of an audio file can be recorded by a fingerprinting system. When a file is accessed by a user, a watermark, or called fingerprint in this case, is embedded into the file thus creating a mark on the audio. The usage history can be traced by extracting all the watermarks that were embedded into the file [7].

Broadcast monitoring: Watermarking is used in code identification information for an active broadcast monitoring. No separate broadcast channel is required as the data is embedded in the host signal itself which is one of the main advantages of the technique [4].

Copy control and access control: A watermark detector is usually integrated in a recording or playback system, like in the DVD copy control algorithm [8] or during the development of Secure Digital Music Initiative (SDMI) [7]. The copy control and access control policy detects the watermark and it enforces the operation of particular hardware or software in the recording set [11].

II. RELATEDWORK

There are many data hiding and data extraction schemes are comes into existence. The important data hiding technique is steganography.It is differ from cryptography in the way of data hiding. The goal of steganography is to hide the data from a third party whereas the goal of cryptography is to make data unreadable by a third party. In [2] The steganalysis method is used. The goal of steganalysis is to determine if an image or other carrier contains an embedmessage. In my project the concept of „Watermarked Content only attack" in the watermarking security context is taken.i.e the blindly recovery of data is considered. In [3],in steganalysis concept it is said to be Universal Steganalysis means instead of using any priori information ,they take into account all available steganography methods to devise a single steganalysis framework. This approach can detect any steganography if sufficient numbers of cover and stego images have been taken into account during the design process.In [4] spread spectrum embedding algorithm for blind steganography have based on the understanding that the host signal acts as a source of interference to the secret message of interest.Such knowledge can be useful for the blind receiver at the recovery side to minimize the recovery error rate for a given host signal.To increase the security and payload rate the embedder will take multicarrier embedding concept. In [5] the spread spectrum communication is explained. Here a narrow band signal is transmitted over a much larger bandwidth such that the signal energy present in any single frequency is imperceptible. Similarly in SS embedding scheme, the hidden data is spread over many samples of host signal by adding a low energy gaussian noise sequence. The DCT transformation is taken for embedding purpose as a carrier since it is a fast algorithm and for it‟s efficient implementation. In [6] the Generalized Gaussian Distribution (GGD) has been used to model the statistical behavior of the DCT coefficients. In [7] there are many extraction procedures to seek the hidden data. Butit is having some disadvantages. Iterative Least Square Estimation (ILSE) is prohibitively complex even for moderate values. Pseudo-ILS (ILSP) algorithm is not guaranteed to converge in general and also it provides measurably worse results.So,these two algorithms coupled and so called Decoupled weighted ILSP(DW-ILSP).But here also have an disadvantage like ,it may not be valid for large N..

III. PROPOSED SYSTEM

The proposed system uses blind recovery of data and it uses the DCT transform as a carrier for embedding the data in digital media.Embedding is performed by using multicarrier SS embedding technique.It uses M-IGLS algorithm for the extraction of the hidden data.It is a low complexity algorithm and provides strong recovery performance. It attains equal probability of error recovery to known host and embedding carriers.It is used as a performance analysis tool for the data hiding schemes.

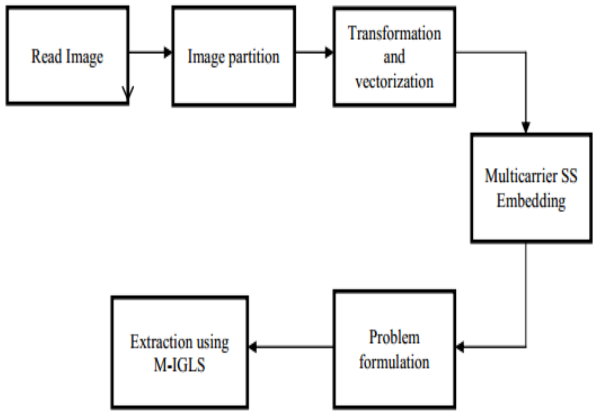


Fig. 1 Modules for data hiding and extraction

IV. SPREAD SPECTRUM METHOD

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. In frequency-hopping SS, the audio file's frequency spectrum is altered so that it hops rapidly between frequencies. The math theory behind SS is quite complicated and goes beyond the scope of this project. However, Katzenbeisser and Petitcolas write about a generic steganography system that uses direct-sequence SS in Information Hiding Techniques for Steganography and Digital Watermarking. The following procedural diagram illustrates the design of that system when applied to our specific topic of audio steganography.

V. EMBEDDING AND EXTRACTION ALGORITHM

A. *Embedding of Watermark*

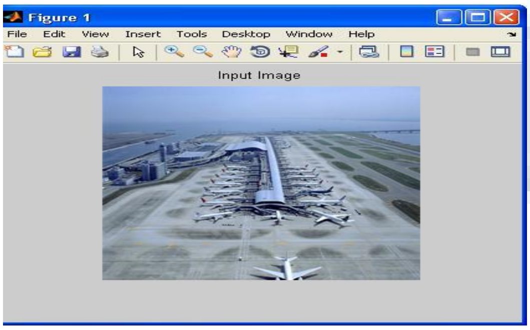
1. Reading of cover audio signal and getting of equivalent 2D matrix and calculate size.
2. Reading of watermark image and getting of equivalent 2D matrix and Calculate size of matrix.
3. Conversion of watermark matrix into binary matrix.
4. Getting of spreading size by multiplying spreading factor with total number of elements of binary watermark matrix.
5. Generation of a random binary key sequence according to spreading size, so as to provide security.
6. Encoding of watermark matrix by Binary XOR-ing of row vector watermark matrix with key sequence.
7. Now, encoded watermark matrix has a double size as compared to that of original.
8. If cover image is too big than division of cover image matrix into two parts.
9. Selection of a block size, which must be suitable to the size of first part of cover image matrix.
10. Division of cover image matrix into first and second part.
11. Segmentation of first part matrix into an array of sub-matrix.
12. Each sub-matrix has a specific number of elements which depends upon block size.
13. Application of Discrete Cosine Transform (DCT) on each element of all the sub-matrices.
14. Embedding of watermark by multiplication of encoded watermark matrix with cosine transform matrix.
15. Reconstruction of matrix by application of inverse discrete cosine transform on resultant matrix.
16. Joining of reconstructed matrix with second part of cover image matrix and Resizing of embedded image according to original audio cover signal
17. Plotting of frequency coefficients of both audio cover signals, so as to make comparison.

B. Extraction of Watermark

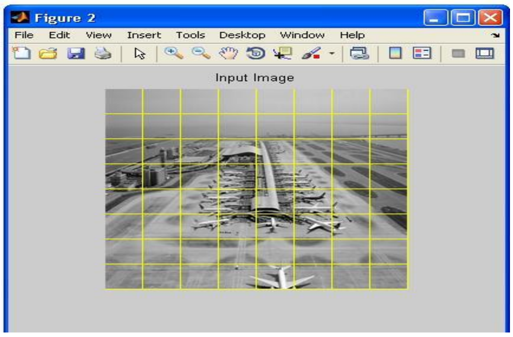
1. Reading of audio cover signal.
2. Reading of audio watermarked signal and Calculation of size of cover audio signal.
3. Reading of watermark image and Calculate size of watermark image.
4. Selection of block size, so as to increase the spreading.
5. Division of both images i.e. cover and marked audio into two parts.
6. Declaration of empty cell having array of empty matrices so as to fill these with first part of both matrices.
7. Declaration of threshold value so as to fill the empty cell up to a certain limit.
8. Application of discrete cosine transform on both cell.
9. Division of 3rd element of each matrix of watermarked signal by that of original audio cover signal.
10. Decoding of watermark components or removal of key sequence.
11. Reconstruction of extracted watermark according to size of original watermark image
12. Plotting of both watermark images i.e. original and extracted.

VI. MODULEDESCRIPTION

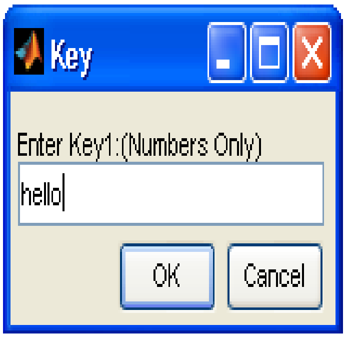
Preprocessing and image partition:The hidden message has to hide in digital media like audio,video or image. Herefor hiding the data image is taken as host.Image can either as RGB or gray scale image.Image is partitioned into non-overlapping blocks. Each block should carry hidden information bits. Forthat, block division features are to be known. The image is divided into blocks on the basis of 8*8 matrix. This 8*8 blocks are independently processed for embedding in different domain. Bythis, the embed blocks are synchronised.



2.a



2.b



Enter Key1:(Numbers Only)

hello

2.c

Fig. 2(a,b,c) Host image and image partition and the data to be hidden.

Transformation and vectorization: For image transformation, we will take the DCT transform. It is well known that DCT transformation provides excellent energy compaction in low spectral coefficients for highly correlated data. Any disturbance directly or indirectly added in the frequency domain may result in a change of statistical properties.DCT is applied in blocks of 8*8 matrix. The gaussian distribution is used to model the statistical properties of the DCT coefficients. Then the vectorization process will undertaken.Vectorization is the process of converting raster graphics to vector graphics. Multicarrier SS embedding: The embedding method is designed to satisfy the perceptual constraints and improve the detectability as well as the embedding rate. Instead of the pixel value, the histogram can be modified to embed the data. If we examine typical histograms of DCT coefficients we will find some samples have high amplitudes that the generalized Gaussian model cannot adequately found. We will consider the DCT coefficients whose amplitude is below a certain threshold value. In this embedding scheme, the hidden data is spread over many samples of host signal or image by adding the DCT coefficient as the carrier.

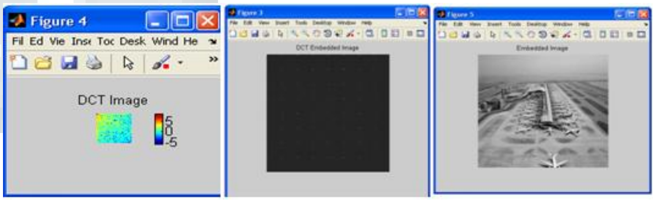


Fig. 3 Transformed and embedded image

Problem formulation: With high resolution digital images as carrier, detecting hidden nessages is also considerably difficult. The hiddenimage often a binary sequence in embedded by substituting a host signal component with a quantized value. Thequantization error will produce due to data embedding. As per the embedding rule, the quantization error α>0.5 is added back to quantization value in order to compensate for quantization distortion. During the detection, the original image is not available we must treat it as additive noise. The distortion can be occurring due to the original host image and due to the embedding and compression process. Featurevector extraction is used to achieve the detection.

Data extraction using M-IGLS: After the detection process, if the data is present in the image, then we have to extract it.

So we are using an algorithm known as Multicarrier Iterative Generalized Least Squares. This comprises the following steps. Each time compute a least squares. Update for one of the unknown matrices conditioned on a previously obtained estimate for the other matrix. Proceed to update the other matrix and repeat until convergence of the least squares cost function is reached. Convergence is the property that different transformation of the same state have a transformation to the same end state. Convergence of the least square cost is guaranteed since each update may either improved or maintained. The final output is generally dependent on the initialization.

VII. RESULT

The proposed method is to extract the hidden data from the digital media. Here blindly recovery of data is considered. That is the original host end embedding carrier is not need to be known. This method uses multicarrier embedding and DCT transformation for the embedding the data into the host image. The M-IGLS algorithm is used for the extraction purpose. This algorithm is a low complexity algorithm and it attains the probability of error recovery equals to known host and embedding carriers. It is used as a tool to analyse the performance of the data hiding schemes.

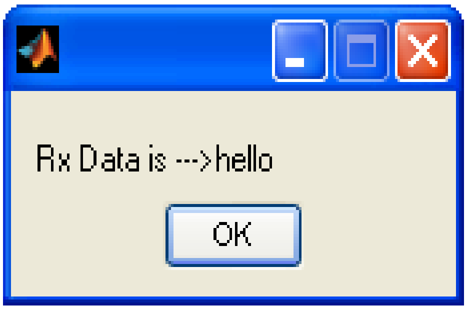

Rx Data is --->hello

Fig. 4 Extracted Data

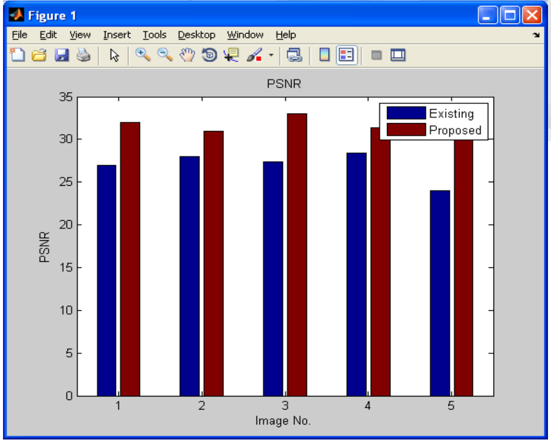


Fig. 5 Graph for PSNR verses image number

In this performance graph peak signal-to-noise ratio,extract the hidden data from the digital media.PSNR is most commonly used to measure the quality of reconstruction of image compression. PSNR is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. For color images the image is converted to a different color space and PSNR is reported against each channel of that color space. Typical values for the PSNR in lossy image and video compression are between 30 and 50dB.This was obtained in our proposed system. For higher value the bit rate is to kept better.

VIII. CONCLUSION AND FUTURE WORK

Data tracking and tampering is rapidly increasing in communication. So we have to secure the data from the trackers. Hence we need a robust and secured data hiding and extraction schemes. The main aim of the proposed system is to provide a good extraction technique which considered the blindly recovery of data. This method uses the M-IGLS algorithm for the extraction. The data is embedded via DCT transform by multicarrier SS embedding. This extraction technique will provides high peak signal to noise ratio and it will attains the probability of error recovery equals to known host and embedding carriers. This technique is enhanced by using harmony search algorithm where it provides low time consumption and high attack resistance.

REFERENCES

[1] Youail, R.S., Samawi, V.W. and Kadhim, A-R. A- K. (2008) "Combining a Spread Spectrum Technique with Error-Correction Code to Design an Immune Stegosystem", Anti counterfeiting, Security and Identification (ASID 2008), IEEE, pp. 245-248.

[2] RU, X.M., ZHANG, H.J. and HUANG, X (2005), "steganalysis of audio: attacking the steghide", Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou (2005), IEEE, pp. 3937-3942.

[3] Kexin, Z. (2010), "Audio Steganalysis of Spread Spectrum Hiding Based on Statistical Moment", 2nd International Conference on Signal Processing Systems (ICSPS-2010), IEEE, vol. 3, pp. 381-384.

[4] Gupta, A., Barr, D.K. and Sharma, D. (2009), *"Mitigating the Degenerations in Microsoft Word Documents: An Improved Steganographic Method"*, 2nd International Conference on Computer, Control and Communication (IC4-2009), IEEE, pp.1-6.

[5] Nutzinger, M., Fabian, C. and Marschalek, M. (2010), *"Secure Hybrid Spread Spectrum System for Steganography in Auditive Media"*, Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (2010), IEEE, pp. 78-81.

[6] Gao, S.; Hu, R.M.; Zeng, W.; Ai, H.J. and Li, C.R. (2008), *"A Detection Algorithm of Audio Spread Spectrum Data Hiding"*, International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM-2008), IEEE, pp. 1-4.

[7] Garay,S.H.; Medina, R.V.; Rivera, L. N. and Ponomaryov, V. (2008),*"steganographic communication channel using audio signals"*, International Conference on Mathematical Methods in Electromagnetic Theory (2008), IEEE, Odesa, Ukraine,

[8] Shah, P.; Choudhari, P. and Sivaraman, S. (2008), *"Adaptive Wavelet Packet Based Audio Steganography using Data History"*, Region 10 Colloquium and the Third ICIIS, Kharagpur, (2008), IEEE.

[9] Li, M., Kulhandjian, M., Pados, D.A., Batalama, S.N., Medley, M.J. and Matyjas, J.D. (2012), *"On the Extraction of Spread-Spectrum Hidden Data in Digital Media"*, Communication and Information Systems Security Symposium, IEEE (ICC- 2012), pp. 1031-1035.

[10] Ghosh, S., De, D. and Kandar, D. (2012), *"A Double Layered Additive Space Sequenced Audio Steganography Technique for Mobile Network"*, International Conference on Radar, Communication and Computing (ICRCC-2012), IEEE, SKP Engineering College, Tiruvannamalai, pp. 29-33.

[11] Skopin, D.E. ; El-Emary, I.M.M. ; Rasras R.J. and Diab R.S.(2010), *"Advanced Algorithms in Audio Steganography for Hiding Human Speech Signal"*, International Conference on Advanced Computer Control (ICACC- 2010) , IEEE, vol. 5, pp. 29-32.

[12] Kumar, H. and Anuradha (2012), *"Enhanced LSB technique for Audio Steganography"*, International Conference on Computing, Communication & Networking Technology (ICCCNT-2012), IEEE-20180, Coimbatore.

[13] Liu, B., Xu, E., Wang, J., Wei, Z., Xu, L., Zhao, B. and Su, J (2011), "Thwarting Audio Steganography Attacks in Cloud Storage Systems*"*, International Conference on Cloud and Service Computing (2011), IEEE, pp. 279-284.