ISSN 2321 3361 © 2018 IJESC

Research Article Volume 8 Issue No.4

Extraction of Spread Spectrum Hidden Data from Image using M-IGLS Algorithm

Pankaj P. Pawar[1], Sharayu T. Mane[2]
Assistant Professor[1], ME Scholar[2]
Department of Computer Science and Engineering
MGM College of Engineering, Nanded, Maharashtra, India

Abstract:
Data embedding and extraction schemes are increasing in today's communication world due to rapid increment of data tracking and tampering attacks. Extracting embedded data without understanding the analysis over a wide band in a spectrum domain of digital medium like image, audio and video and it considered a problem of active blind spread spectrum steganalysis and attempted to recover unknown messages hidden in image hosts via multi-signature spread spectrum embedding. So we need an efficient and robust data hiding schemes to protect from these attacks. We develop novel multi-carrier/signature iterative generalized least-squares (M-IGLS) procedure to find or to extract unknown data hidden in hosts via multicarrier spread spectrum embedding. In this we assumed that the original host and the embedding carriers are available. M-IGLS is a low complexity algorithm and it attains the probability of error recovery equals to known host and embedding carriers. Its peak signal to noise ratio value obtained is high. Experimental studies on images show that the developed algorithm can achieve recovery probability of error close to what may be attained with known embedding carriers and autocorrelation matrix.

Keywords: Data hiding, Tracking, Tampering, Blindly extraction, Spread spectrum embedding.

I. INTRODUCTION

In the field of Data Communication, security issues have the major problem. 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. For that, we are having many data hiding schemes and extraction schemes. Hiding the information is a vital issue in the 21st century in the field of Data Communication security. It's an important issue because the virtual and digital information transmission faces critical setbacks due to hacking and hackers threats. The transmission of information via the Internet may expose it to detect and theft. So the data embedding technologies are developed to provide personal privacy, commercial and national security interests. Digital data embedding in digital media is rapidly growing commercial as well as national security interest. Data hiding schemes are initially used in military communication systems like encrypted message, for finding sender and receiver or its very existence. Initially the data hiding schemes are used for 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 also provides a good recovery of hidden data. This is the goal of secured communication. The main applications of data hiding are annotation, copyright-marking and watermarking, single-stream media merging (image, audio, text) and Steganography [1].

The proposed paper having the blindly extraction technique is considered. Blindly extraction means the original host and the embedding carriers are not require to be known. Here, data embedded to the host digital signal, via multicarrier spread spectrum embedding. Hence it has developed rapidly in this area due to the advantages of good robustness and immunity to noise attack. Spread spectrum techniques really of digital communications systems. Two commonly spread spectrum techniques are used: Direct Sequence Spread Spectrum and Frequency Hopped Spread Spectrum. Implementation of hiding data in image data using direct sequence spread spectrum has been presented in this work. Hidden data is extracted from digital media like image, video or audio. The extraction algorithm used to extract the hidden data from digital media is Multicarrier Iterative Generalized Least Squares (M-IGLS). This blind hidden data extraction problem has also been referred to as "Watermarked Content Only Attack" (WOA) in the watermarking security context [2]-[4]. In blind active spread spectrum detection is the unknown host acts as a source of interference/disturbance to the data to be extracted with, in a way; the trouble parallels blind signal separation (BSS) applications as they arise in the fields of biomedical signal processing, array processing and code division multiple access (CDMA) communication systems. In proposed paper, we implement new multicarrier iterative generalized least squares (M-IGLS) SS algorithm for hidden data extraction.

II. LITERATURE SURVEY

The growing use of the Internet allows users to access, share, manipulate, and distribute digital media data easily and it has affected our daily life profoundly. However, it also makes unauthorized proliferation of digital media much easier, which poses key challenges to the copyright industry and raises critical issues for intellectual protection of digital media. To address the above concern, watermarking and data hiding have been applied as promising ways for post delivery protection of digital media data. The basic idea of watermarking and data hiding is to embed some invisible information into the host media signal and the hidden data can later be extracted for desired purpose. The hidden information could be used for digital media authentication, copyright protection, information embedding, database annotation, traitor tracing, and so on.

2.1 Related Work in Data Hiding

F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn [1] gave an overview of information hiding and the study of analysis for the steganography techniques. They described the number of attacks on information hiding and also described the number of transform techniques. It is often thought communications may be secured by encrypting the traffic but this has rarely been adequate in practice. For example, an encrypted email message between a known drug dealer and somebody not yet under suspicion, or between an employee of a defence contractor and the embassy of a hostile power, has obvious implications. The study of communications security includes not just encryption but also traffic security, whose essence lies in hiding information. This discipline includes such technologies as spread spectrum radio, which is widely used in tactical military systems to prevent transmitters being located; temporary mobile subscriber identifiers, used in digital phones to provide users with some measure of location privacy; and anonymous remailers, which conceal the identity of the sender of an email message. An important sub discipline of information hiding is steganography. While cryptography is about protecting the content of messages, steganography is about concealing their very existence. Until recently, data hiding techniques received much less attention from the research community. The embedded data is the message that one wishes to send secretly. It is usually hidden in an innocuous message referred to as a cover-text, or cover-image or cover-audio as appropriate, producing the stego-text or other stego-object. A stego-key is used to control the hiding process so as to restrict detection and/or recovery of the embedded data to parties who know it (or who know some derived key value). As the purpose of steganography is having a covert communication between two parties whose existence is unknown to a possible attacker, a successful attack consists in detecting the existence of this communication. Copy-right marking, as opposed to steganography, has the additional requirement of robustness against possible attacks. The authors claimed that the spread spectrum technique is robust information hiding system with low probability errors.

2.2 Related Work in Spread Spectrum Techniques:

R. Chandramouli [5] had proposed a method called mathematical framework for steganalysis with linear steganography in 2004. In this method a mathematically formal definition of steganalysis is given. Steganalysis is relatively a new branch of research, while steganography deals with techniques for hiding information (such as fingerprinting), the goal of steganalysis is to detect and/or estimate potentially hidden information from observed data with little or no knowledge about the steganography algorithm and/or its parameters. In traditional steganography set-up formulated as a prisoner's problem: Alice wishes to send a secrete message to Bob by hiding information in a cover message. The stego message (cover + secrete message) passes through Wendy (a Warden) who inspects it to determine if there is anything suspicious about it. Wendy could perform one or several tests to decide if the message from Alice to Bob contains any secrete information – Wendy acts as a Passive Warden. If her decision is negative then the Wendy forwards the message to Bob. On the other hand, Wendy can take a conservative approach and modify all messages from Alice to Bob irrespective of whether any information is hidden by Alice or not. In this case Wendy is called Active Warden. In this paper the steganalysis techniques used by the authors focused on detecting the presence/absence of secrete message in an observed message. In general, extraction of secrete message is harder problem than mere detection simply because the former outputs multiple bits of information while the latter results in a two bit (secret message present or absent) information.

Therefore, based on the ultimate outcome of the effort authors classified steganalysis into two categories:

(a)Passive Steganalysis: Detect the presence or absence of hidden message in a stego signal, identify the stego embedding algorithm

(b)Active Steganalysis: Estimate the embedded message length, estimates location(s) of hidden message, estimate the secrete key used in embedding, estimates some parameters of stego embedding algorithm, extract the hidden message.
In this paper, authors derived a mathematical framework for active steganalysis when a class of linear steganography algorithms was employed. In this paper the steganography key is same for at least two stego messages. Based on the discussion of information collection for steganalysis authors again classified steganalysis methods into two general categories:

(a)Spatial diversity information based steganalysis:
In this method, Steganalysis methods can look for information in the spatial domain that repeats itself in various forms in different spatial locations (e.g. different blocks within an image or, in different images).

(b)Temporal diversity information based steganalysis: Steganography information that appears repeatedly over time can also aid steganalysis. Such techniques are called temporal diversity information based steganalysis. In this paper authors also described a generic linear additive steganography algorithm and then mathematically setup the corresponding steganalysis problem. An additive steganography models seems to fit a wide range of popular steganography techniques. In this paper, data embedding was based on employing two different quantizers to represent the message bits 0 and 1 then the quantization error can be modeled as additive noise interfering with the cover message. LSB embedding for image steganography changes the pixel values by ±1. Many steganography methods first use the message bits to modulate a carrier signal which is then added to cover message. The authors worked on active steganalysis which is a new branch of research, in this paper only one copy of stego message is available.

III. SYSTEM METHODOLOGY

Now in today's communication world data embedding and extraction schemes are increasing 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. Blindly extraction of hidden data from digital media is considered in this dissertation work. In this paper, 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 digital media like image. The extraction algorithm which is used to extract the hidden data from digital media. Multi-carrier iterative generalized least square (M-IGLS) method is used to extract the hidden data from digital media. M-IGLS is a low complexity algorithm and it attains the probability of error recovery equals to known host and

embedding carriers. Its peak signal to noise ratio value obtained is high.

Flow diagram of method is given in figure 1.

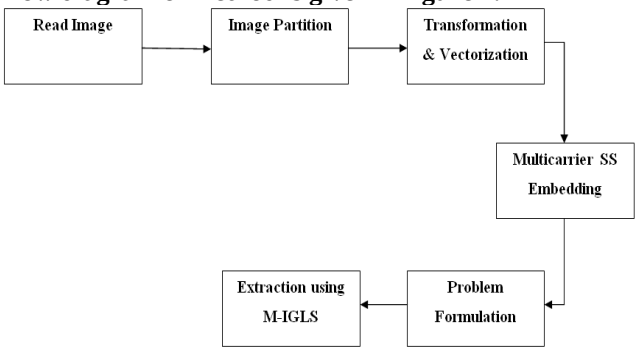

Figure .1. Modules for Data Hiding and Data Extraction

The above fig. gives the detail view of the modules used for data hiding and data extraction. The modules that have been included for embedding and extraction of hidden data are discussed below.

1. Preprocessing and Image Partition:

The proposed spread spectrum technique uses blind recovery of hidden data. DCT transform is used as a carrier for embedding the data in digital media. Multicarrier spread spectrum technique is used for embedding process. In the first step, consider the host image. The hidden message has to hide in digital media like image. Consider a host image $H \in M^{N1\times N2}$ where M is the finite image alphabet and $N1 \times N2$ is the image size in pixels. Image is partitioned into non-overlapping blocks. Each block should carry hidden information bits. For that, block division features to be known. The image is divided into blocks on the basis of 8*8 matrix. This 8 * 8 matrix blocks are independently processed for embedding in different domain. By this the embed blocks are synchronized. Without the loss of generality, the image H is partitioned into M local non-overlapping blocks of size $\frac{N1N2}{M}$. Each block, H1, H2, ……., HM is to carry K hidden information bits (KM bits total image payload).

2. Transformation and Vectorization:

For image transformation, we will take 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. It is common and appropriate to avoid the dc coefficient (if applicable) due to high perceptual sensitivity in changes of the dc value.

The autocorrelation matrix of the host data x is an important statistical quantity for the developments and is defined as:

$$R_x \triangleq \mathbb{E}\{xx^T\} = \frac{1}{M}\sum_{m=1}^{M} x(m)x(m)^T$$

It is easy to verify that in general $R_x \neq \alpha I_L, \alpha > 0$.

R_x is not a constant value diagonal or white in field language. For example, 8 * 8 DCT with 63-bin host data formation (excluding only the dc coefficient) for the gray-scale image.

3. 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.

In this method, K distinct message bit sequences, $\{b_k(1), b_k(2), \ldots\ldots\ldots\ldots, b_k(M)\}$, k = 1, 2, …, K, $b_k(m) \in \{\pm 1\}$, m = 1, 2, …, M, each of length M bits.

In particular, the m^{th} bit from each of the K sequences, is simultaneously hidden in the m^{th} transform-domain host vector x(m) via additive SS embedding by means of K spreading sequences (carriers) $s_k \in R^L$, $\| s_k \| = 1$, k = 1,2,….,K.

$$y(m) = \sum_{k=1}^{K} A_k b_k(m) s_k + x(m) + n(m), m = 1,\ldots, M$$

4. Extraction using MIGLS

In this method, to blindly extract spread-spectrum embedded data from a given host image, we needs first to convert the host image to observation vectors of the form of y (m), m=1, 2… M. This requires the knowledge of: i) the partition, ii) transform domain, iii) subset of coefficients, and iv) number of carriers used by embedder. We denote the combined "disturbance" to the hidden data (host plus noise) by:

$$y(m) = \sum_{k=1}^{K} A_k b_k(m) s_k + z(m), m = 1,\ldots, M$$

Table.1.Multicarrier Iterative Generalized Least Squares Data Extraction Procedure

1	$d := 0; Initialize\ \hat{B}^{(0)} \in \{\pm 1\}^{K\times M} arbitarily.$
2	$d := d + 1;$ $\hat{V}^{(d)} := Y(\hat{B}^{(d-1)})^T[(\widehat{B}^{(d-1)})(\hat{B}^{(d-1)})^T]^{-1};$ $\hat{B}^{(d)} := sgn\{((\hat{V}^{(d)})^T R_y^{-1}(\hat{V}^{(d)}))^{-1}(\hat{V}^{(d)})^T R_y^{-1} Y\}$
3	Repeat step 2 until $\hat{B}^{(d)} = \hat{B}^{(d-1)}$.

The above table gives the multicarrier iterative generalized least square algorithm for blind extraction of hidden data.

IV. RESULTS

The proposed method is used to extract the hidden data from digital media like image. Here the blindly recovery of hidden data is considered. Blindly extraction means the original host and the embedding carrier is not need to be known. The M-IGLS algorithm is used for extraction process. Quality of hidden message extraction solution is the difference in bit-error-rate (BER) experienced by the intended recipient. The intended recipient using the: (i) ideal MMSE filtering with known carriers and known true host autocorrelation matrix. In terms of blind extraction, we examine: (ii) M-IGLS. We consider a gray scale host image of "Boat" size 512*512 as shown in fig. 2. Then we perform 8*8 blocks DCT transform embedding and consider all bins except dc coefficient. Each message is having a length of 4K bits. The embedding carriers of length id L=63. The per message block mean square distortion due to each embedded message is set to be the same for all messages. The graph between average BER versus per message block distortion is shown in fig.3.

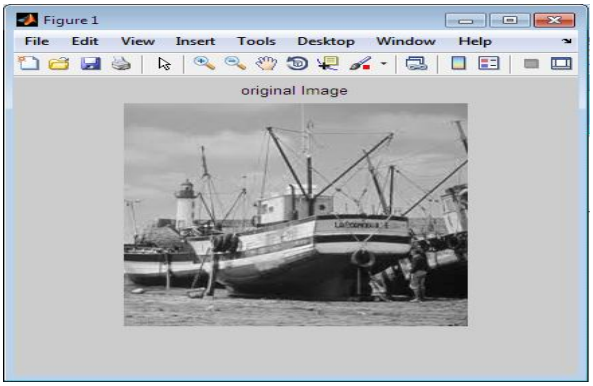

Figure.2. Host Image of Boat

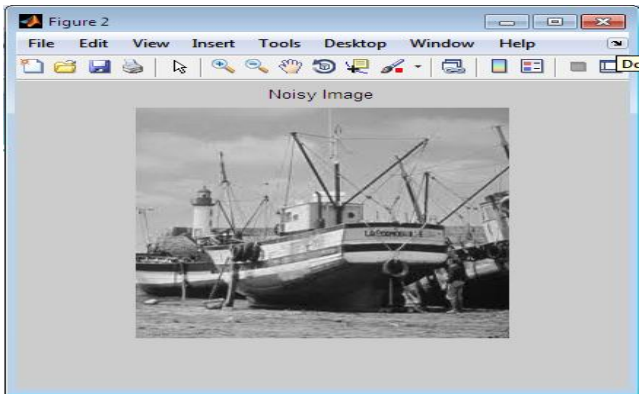

Figure.3. Noisy Image

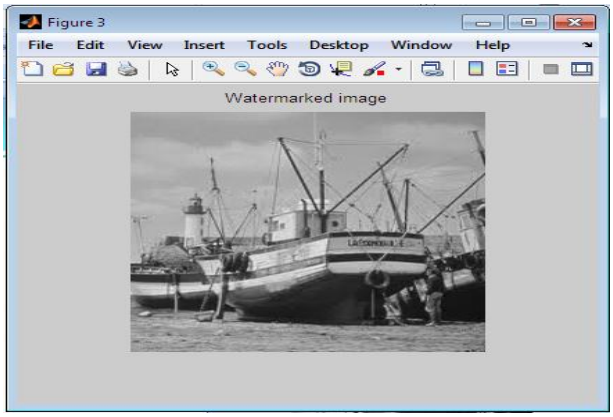

Figure .4. Watermarked Image

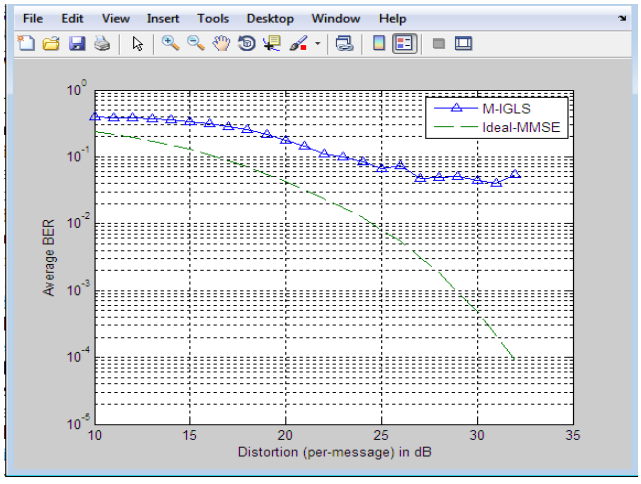

Figure.5. Average BER versus Distortion (per-message) in db for Boat Image

In figure 3, the graph shows, the gap between M-IGLS and Ideal MMSE increases as the hidden message size decreases. It shows the extraction performance of MIGLS seems to be satisfactory than ideal MMSE. The following table shows the performance measurement of M-IGLS and Ideal MMSE method by calculating distortion of per block message and bit error rate.

Table.2. Distortion and BER values for Ideal MMSE and MIGLS

Distortion in dB	Pe_Ideal_MMSE	Pe_MIGLS
10	0.2412	0.4033
11	0.2207	0.3922
12	0.1959	0.3813
13	0.1682	0.3662
14	0.1518	0.3499
15	0.1314	0.3296
16	0.1056	0.3085
17	0.0907	0.2873
18	0.0713	0.2411
19	0.0546	0.2231
20	0.0425	0.1710
21	0.0313	0.1475
22	0.0243	0.1279
23	0.0170	0.1000
24	0.0121	0.0834
25	0.0083	0.0649
26	0.0047	0.0568
27	0.0033	0.0629
28	0.0017	0.0488
29	0.0009	0.0587
30	0.0005	0.0388
31	0.0002	0.0351
32	0.0001	0.0371

V. CONCLUSION

The M-IGLS method has been proposed to provide a good recovery of extraction technique which considered the blindly recovery of hidden data. The data is embedded via DCT transform by using multicarrier SS embedding method. This technique of extraction will provides a high peak signal to noise ratio and it will attains the probability of error recovery equals to known host and embedding carriers. Experimental studies shows that the M-IGLS can achieve probability of error rather close to what may be attained with known embedding signatures and known host autocorrelation matrix.

VI. REFERENCES

[1]. F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Information hiding: A survey," Proc. IEEE, Special Issue on Identification and Protection of Multimedia Information, vol. 87, no. 7, pp. 1062–1078, Jul. 1999.

[2]. G. C. Langelaar, I. Setyawan, and R. L. Lagendijk, "Watermarking digital image and video data: A state-of-the-art overview," *IEEE Signal Process. Mag*., vol. 17, no. 5, pp. 20–46, Sep. 2000.

[3]. N. F. Johnson and S. Katzenbeisser, S. Katzenbeisser and F. Petitcolas, Eds., "A survey of steganographic techniques," in *Information Hiding*. Norwood, MA, USA: Artech House, 2000, pp.43–78.

[4]. Y. Wang and P. Moulin, "Perfectly secure steganography: Capacity, error exponents and code constructions," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2706–2722, Jun. 2008.

[5]. R. Chandramouli, "A mathematical framework for active steganalysis," *ACM Multimedia Syst., Special Issue on Multimedia Watermarking*, vol. 9, pp. 303–311, Sep. 2003.

[6]. H. S. Malvar and D. A. Florencio, "Improved spread spectrum: A new modulation technique for robust watermarking," *IEEE Trans. Signal Proc*., vol. 51, no. 4, pp. 898–905, Apr. 2003.

[7]. J. Hernandez, M. Amado, and F. Perez-Gonzalez, "DCT-domain watermarking techniques for still images: Detector performance analysis and a new structure," *IEEE Trans. Image Process*., vol. 9, no. 1, pp. 55–68, Jan. 2000.

[8]. M. Gkizeli, D. A. Pados, S. N. Batalama, and M. J. Medley, "Blind iterative recovery of spread-spectrum steganographic messages," in *Proc. IEEE Int. Conf. Image Proc. (ICIP)*, Genova, Italy, Sep. 2005, vol. 2, pp. 11–14.

[9]. M. Gkizeli, D. A. Pados, and M. J. Medley, "Optimal signature design for spread-spectrum steganography," *IEEE Trans. Image Process*., vol. 16, no. 2, pp. 391–405, Feb. 2007.

[10]. L. Pérez-Freire and F. Pérez-González, "Spread-spectrum watermarking security," *IEEE Trans. Inf. Forensics Security*, vol. 4, no. 1, pp. 2–24, Mar. 2009.

[11]. A. Valizadeh and Z. J. Wang, "Correlation-and-bit-aware spread spectrum embedding for data hiding," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 2, pp. 267–282, Jun. 2011.