International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Secure Spread Spectrum Data Embedding and Extraction

Vaibhav Dhore[1], Pathan Md. Arfat[2]

[1]Professor, Department of Computer Engineering, RMD Sinhgad School of Engineering, University of Pune, India

[2]Department of Computer Engineering, RMD Sinhgad School of Engineering, University of Pune, India

Abstract: ***Embedding the secrete data into the digital media like image, video audio is rapidly growing technology from the commercial as well as national security perspective. From the past few years the only the presence of the embedded data intensively investigated which is called passive detection .While the active hidden data extraction is relatively new branch in research. This paper gives the survey of the methods for embedding the data using direct sequence spread spectrum technique and gives the method for blindly extracting the hidden data embedded over a wide band of spectrum or transformed domain of digital medium. To extract the hidden data blindly we have used the Multi Carrier Iterative Generalized Least square (M-IGLS) procedure to seek the unknown data hidden in the image host. Neither the original host nor the embedding carriers assumed available. Then to enhance the MIGLS we have used the Cross Correlation MIGLS(CC-MIGLS) procedure, which leads to most effective hidden message recovery by performing the statistical analysis of repeated independent MIGLS processing of the host.***

Keywords: Steganalysis, steganography, data hiding, data embedding, cross correlation, blind detection, spread spectrum embedding, watermarking.

1. Introduction

The purpose of the data hiding technique is to establish the covert communication between two trusting parties. The steganography is one of the technique to embed the secrete data into the cover medium such as image , audio or video and the steganalysis is the counter measure technique to steganography which aims to discover the presence of the embedded data and extract the embedded data from the host. There are two types of staganalysis i.e passive analysis and active analysis. Passive analysis is used to decide the presence of absence of secrete hidden data into the digital medium while the active analysis focuses on extracting the actual hidden data from the past few years the passive analysis is intensively investigate while the active analysis is relatively new branch of research.

In this paper we focus our attention on active detection of data using Spread Spectrum steganalysis. We are using the Direct Sequence Spread Spectrum (DS-SS) method to embed our secrete data into the image. In case of telecommunications, direct-sequence spread spectrum (DS-SS) is a technique,In which the transmitted signal having more bandwidth than the information signal that modulates the carrier or broadcast frequency. The name 'spread spectrum' implies that the carrier signals occur over the full bandwidth or say spectrum of transmitting frequency. As we are using the transform (spectrum) domain of the image to embed the data ,hence we can use the DS-SS technique.
Direct-sequence spread-spectrum transmissions multiply the data being transmitted by a noise signal. This noise signal is nothing but a pseudorandom sequence of 1 and −1 values, at a frequency much higher than that of the original signal. The resulting signal appears like white noise, like an audio recording of "static". However, this noise-like signal is used to exactly reconstruct the original data at the receiving end by multiplying the data by the same pseudorandom sequence (because 1 × 1 = 1, and −1 × −1 = 1). This process, known as "de-spreading", mathematically constitutes a correlation of the transmitted PN sequence with the PN sequence that the receiver already knows the transmitter is using. An iterative generalized least squares (IGLS) procedure was developed to blindly Extract unknown messages hidden in image hosts via SS embedding. The MIGLS algorithm is having low complexity and strong recovery performance of the data. However, the scheme is designed only for single

carrier SS embedding in which messages are hidden with one signature only and cant be used for the multi-carrier case.To increase security Realistically, an embedder would favor multicarrier SS transform-domain embedding and/or payload rate. In this paper, we gave a novel multi-carrier iterative generalized least squares (M-IGLS) algorithm for SS hidden data extraction . For improved recovery performance, for relatively small hidden messages that pose a big challenge, experimental studies indicate, a few independent re-initializations of MIGLS and executions on the host can lead recovery of hidden data with probability of error close to attained by known embedding carriers and known original host autocorrelation matrix.Hence to enhance the recovery performance of MIGLS algorithm we are using Cross Correlation MIGLS algorithm (CC-MIGLS). CCM-IGLS is nothing but a statistical analysis of independent M-IGLS executions on the host. Experimental studies on active steganalysis indicate that CC-MIGLS can achieve the recovery of the hidden data with probability of error very close as what may be attained with known embedding signatures and the autocorrelation matrix of the known original host.

2. Literature Survey

In [4] “Information hiding a survey” by fobien A.P petitcolas, Rosc J. Anderson and markus G. Kuhan gave an overview of information hiding and study of analysis for the steganography techniques. They described the number of

Licensed Under Creative Commons Attribution CC BY

attacks on information hiding and described the number of transformed domain techniques. The authors claimed that the spread spectrum technique is robust information handing system with low probability of errors.

In [1] "A survey of security mechanism with the direct sequence spread spectrum signals" published by Teaho Kang, Xian G. li, Chansu Yu and Jong Kim gave the basic idea of DS-SS spreading process to embed the watermarked information, which can be considered as a counter measure for the attacker who try to obtain the hidden data and sometimes may intended to replace it with the fake message.

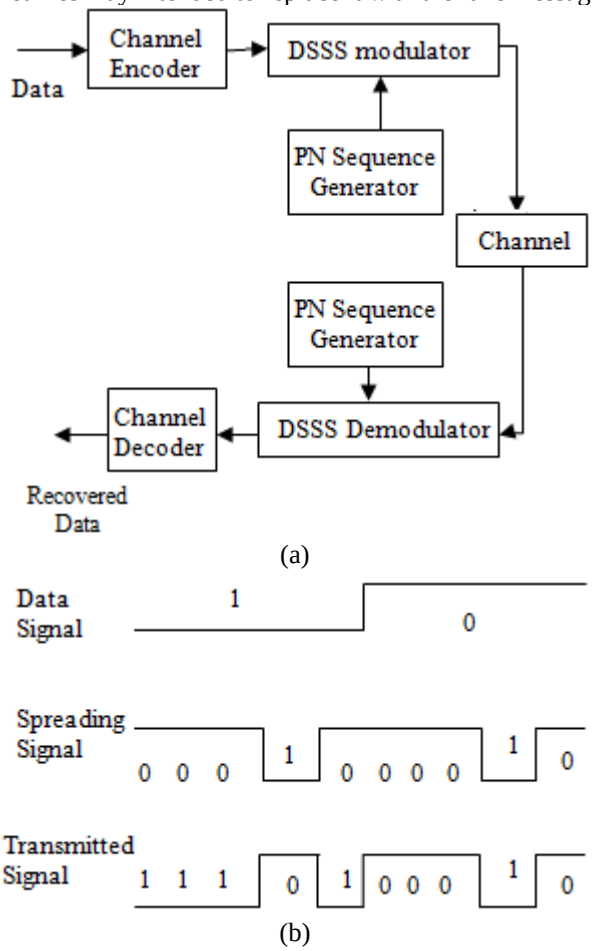

Figure 1: (a) Direct sequence spread spectrum (DSSS) system model and (b) DSSS spreading operation; the pseudonoise (PN) sequence

Figure (a) shows the generic DS-SS system model. In typical DS-SS system transmitter first modulate the signal with a carrier signal by applying modulator addition to it with a spreading signal . The Pseudo Noise(PN) sequence generates the spreading signal periodically at a higher rate than that of original signal.

Figure(b) shows that each individual bit in the PN sequence is used to spread one bit data as the PN sequence resembles the white noise. The spread spectrum occupies lager bandwidth than the necessary bandwidth and posses the lower spectral density as compare to the original signal.

In the base paper published by J.R. Henandez, M. Amando and P. Perez-Gonzalez "DCT-Domain watermarking technique for still images" proposes the DCT-Domain watermarking technique. For copy-right protection of still images is analyzed. This technique use the DCT 8x8 blocks of an image as used in JPEG algorithms.

To ensure that the whether a given image contained a watermark generated with a particular key, two steps are involved, (1) Watermarked decoding, (2) Watermarked extraction. The generalized Gaussian distribution is applied to the DCT coefficients of the image. The author shown in this paper how the resulting detector structure improves the performance with respect to the correlation receiver which uses the Gaussian noise assumptions. As the results of authors work the analytical expressions for performance measures can be derived and contrasted with the results of experiments. The performance errors in this case are probability of watermarked decoding and the probability of false alarms in watermarked detection.

In [2] 2013 paper published by Ming Li, Michael Kulhadjian, Dimitris A. Pados "Extracting spread spectrum hidden data from digital media" proposed a method for blindly extracting the hidden data from digital medium like image, by using the multi-carrier/signature spread spectrum embedding. The data bits are embedded into the frequency domain by using the embedding carriers sequences into a wide range of transformed domain image, while extracting the data neither the original host nor the embedding carriers assumed available. Authors developed a low complexity algorithm called multicarrier iterative generalized least square algorithm (M-IGLS). This algorithm extracts the unknown hidden data with the very low probability of error.

In [3] Active spread spectrum steganalysis for hidden data extraction" by Ming Li, Stella N. Batalama, Dimitri A. Pados and Michael J. Medley proposed the cross correlation enhanced MIGLS(CC-MIGLS) procedure, which enhance the recovery performance of the MIGLS algorithm. Actually the CC-MIGLS based on statistical analysis of repeated MIGLS runs on the stego image and hence it gives most effective method for hidden message recovery.

3. Existing System

The most common technique for the data hiding is the steganography. Steganography is basically classified into three categories. (1) Spatial Domain Techniques ,(2) Transformed Domain techniques (3) Other than spatial or transformed domain. How ever we can further classify these techniques into six categories as follows.

1) ***Substitution(Spatial Domain) :*** This method substitute redundant parts of a cover image with a secrete key.
2) ***Transformed domain techniques***: This steganographic method embed the secret information in a transformed space of a signal.
3) ***Spread Spectrum Technique :*** This technique also uses the transformed space of a signal and also adopt the from the spread spectrum communication technique.
4) ***Statistical method*** : This method encode the information by changing the several statistical properties of a cover and uses the hypothesis testing in the extraction.

Licensed Under Creative Commons Attribution CC BY

5) ***Distortion technique*** : This method store the information by signal distortion and in the decoding step it measures the deviation of the original cover.
6) ***Cover generation technique*** : This method encode the information in such a way that a cover for secrete communication is created.

Table 1: Drawbacks of current steganography methods

Methods	Descriptions
Spatial domain techniques	• Having Large payload but it is often offset the statistical properties of the image. • It is not robust against lossy compression and the image filters. • Not robust against rotation, and translation. • Not robust against noise. • Many work only on the BMP format.
Transformed domain Techniques	• Less prone to attacks than the former methods at the expense of capacity • There is a breach of second order statistics. • There is a breach of DCT coefficients distribution • Work only on the JPEG format • Double compression of the file • It's not robust against rotation, translation and cropping.

The Table 1 gives the drawbacks of the Spatial domain and the Transformed domain techniques in contrast to the spread spectrum technique. In 2013 paper published by Ming Li, Michael Kulhadjian, Dimitris A. Pados "Extracting spread spectrum hidden data from digital media" proposed a method for embedding the secrete data using the multi-carrier spread spectrum embedding in the transformed domain (DCT/ Wavelet transform) and gave the solution for the extraction of the embedded data by using M-IGLS seek procedure.

Drawback : In this paper author does not gave the threshold for the number of runs of the M-IGLS seek procedure and the criterion for the reinitializaion and re-execution of the M-IGLS seek procedure, as it is stated in the paper that certain number of runs of the MIGLS algorithm recover the data which having probability of errors close to what may attend with the known embedding carrier. Also this paper does not provide the signal model representation for embedding the Psedonoise Sequences with the data bits into the image.

4. Proposed System

The proposed spread spectrum technique uses Multi-carrier Direct Sequence Spread Spectrum (DS-SS) Technique which uses the DCT transform as a carrier for embedding the data in an image. To extract the hidden data embedded by multicarrier DS-SS technique we are using M-IGLS algorithm. The proposed algorithm is having a low complexity and it provides strong recovery performance. It provides probability of error recovery equal to the known host and embedding carriers recovery. The proposed scheme is also used as a performance analysis tool for the data hiding schemes. And also to enhance the performance of M-IGLS algorithm,we are using Cross Correlation MIGLS procedure (CC-MIGLS). CC-MIGLS is nothing but a statistical analysis of independent M-IGLS executions on the host. Experimental studies on active steganalysis indicate that CC-MIGLS can achieve the recovery of the hidden data with probability of error very close as what may be attained with known embedding signatures and the autocorrelation matrix of the known original host.

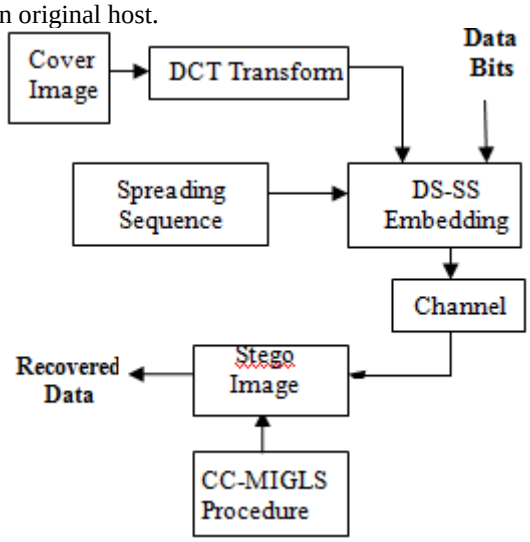

Figure 2: Direct Sequence Spread Spectrum Embedding and Extraction of data from an image

4.1 Problem Formulation

The hidden message is often a binary sequence which is embedded by substituting a host signal component with a quantized value. Data embedding produces the quantization error. while in the processes of detection, the original image is not available,i.e we have stego image, we must treat it as additive noise. The process of distortion can be occurring due to the original host image and due to the embedding and compression process. Spread Spectrum method is known to be very robust, except the c the cost, is very high, the implementation is relatively complex, less secure and the information capacity is very limited. Current spread spectrum steganographic applications with audio media are primarily limited to providing proof of copyright and assurance of content integrity. There is the potential to expand the applications to include the embedding of covert communications. By considering the mentioned problems related to spread spectrum can be overcome by using proposed methodology.

Our objective is to blindly extract the unknown Hidden data B from the the Observation matrix Y without prior knowledge of the embedding carriers S_k and amplitude A_k, $_{k=1,\ldots K}$ in $V=\{A_1S_1,\ldots A_K.S_K\}$

$$\mathbf{Y = VB + Z}$$

Where Y = y(m) , V = $A_k.S_k$ and the host itself x(1), . . . , x(M) in **Z** = [x(1) + n(1), . . . , x(M) + n(M)].

4.2 Data extraction using CC-MIGLS algorithm

For detection and Extraction process we are using The CC-MIGL.As the CC-MIGLS is nothing but a statistical analysis of independent M-IGLS executions. M-IGLS Algorithm is having following steps.
1. Each time compute a least squares.

Licensed Under Creative Commons Attribution CC BY

2. Modify or Update for one of the unknown matrices conditioned on a previously obtained estimate for the other matrix.
3. Process to Proceeding updates the other matrix and repeat until convergence of the least squares cost function is reached.

Convergence is the property that each transformation of the same state has a transformation to the same end state. Convergence of the least square cost is always guaranteed since each update may either improve or maintained. The final output is usually dependent on the initialization.

Table 2: Multi-carrier generalized least square data extraction algorithm

1) $d := 0$; initialize $\widehat{\mathbf{B}}^{(0)} \in \{\pm 1\}^{K\times M}$ arbitrarily.
2) $d := d + 1$;

$$\widehat{\mathbf{V}}^{(d)} := \mathbf{Y}(\widehat{\mathbf{B}}^{(d-1)})^T \left[(\widehat{\mathbf{B}}^{(d-1)})(\widehat{\mathbf{B}}^{(d-1)})^T\right]^{-1};$$

$$\widehat{\mathbf{B}}^{(d)} := \operatorname{sgn}\left\{\left((\widehat{\mathbf{V}}^{(d)})^T\widehat{\mathbf{R}}_{\mathrm{y}}^{-1}(\widehat{\mathbf{V}}^{(d)})\right)^{-1}(\widehat{\mathbf{V}}^{(d)})^T\widehat{\mathbf{R}}_{\mathrm{y}}^{-1}\mathbf{Y}\right\}.$$

3) Repeat Step 2 until $\widehat{\mathbf{B}}^{(d)} = \widehat{\mathbf{B}}^{(d-1)}$.

Since B and V are jointly detected and estimated, correspondingly,if one is not reliable neither is the other in general. We first examine the reliability of the bit matrix decision B = [b1, . . . , bK]^T returned by the M-IGLS procedure of Table 2. The sample cross-correlation between any two bit streams is

$$\eta_{i,j} \triangleq \widehat{\mathrm{b}}_i^T\widehat{\mathrm{b}}_j/M,\ i \neq j,\ i,j = 1,\ldots,K.$$

$$\textit{Criterion 1:}\ \text{If } |\eta_{i,j}| \leq \tfrac{3}{\sqrt{M}} \text{ for all } i \neq j \in \{1,2,\ldots,K\}$$

The basic idea behind the final refinement of M-IGLS hlind signal extraction procedure is to identify average reliable number of clustered estimation; it means we have to find out the reliable number of runs of the M-IGLS procedure.

Criterion 2 : *for signatures k and runs p can be defined as*

$$\rho_{k,p} \triangleq \sum_{j=1,j\neq p}^{P} \frac{|\widehat{\mathbf{v}}_{k,p}^T\widehat{\mathbf{v}}_{k,j}|}{\|\widehat{\mathbf{v}}_{k,p}\|\|\widehat{\mathbf{v}}_{k,j}\|}$$

Where,
is p-th run estimate of Vk.

Now set threshold,

$$\overline{\rho}_k = \frac{1}{P}\sum_{p=1}^{P}\rho_{k,p}.\ \text{If } \rho_{k,p} \geq \overline{\rho}_k$$

Then $\widehat{\mathbf{v}}_{k,p}$ consider “reliable” estimate of Vk otherwise “unreliable”.

Table 2: Cross Correlation Enhanced M-IGLS

For $j := 1$ to P
1) Execute M-IGLS of Table I with arbitrary initialization and obtain estimates $\widehat{\mathbf{v}}_k$, $k = 1,\ldots,K$.
2) **If** estimates are Criterion-1-compliant, $\widehat{\mathbf{v}}_k^{(j)} := \widehat{\mathbf{v}}_k$, $k = 1,\ldots,K$; **else** go to 1).
End
For $k := 1$ to K
3) Identify reliable estimates for $\mathbf{v}_k$ according to *Criterion 2*.
4) Calculate the average over all reliable estimates $\overline{\widehat{\mathbf{v}}}_k$ by (18).
End
5) Set $\overline{\widehat{\mathbf{V}}} \triangleq [\overline{\widehat{\mathbf{v}}}_1,\ldots,\overline{\widehat{\mathbf{v}}}_K]$.
6) Execute M-IGLS of Table I with initialization

$$\widehat{\mathbf{B}}^{(0)} = \operatorname{sgn}\left\{\left(\overline{\widehat{\mathbf{V}}}^T\widehat{\mathbf{R}}_{\mathrm{y}}^{-1}\overline{\widehat{\mathbf{V}}}\right)^{-1}\overline{\widehat{\mathbf{V}}}^T\widehat{\mathbf{R}}_{\mathrm{y}}^{-1}\mathbf{Y}\right\}.$$

Steps:
1. Run Criterion 1 equipped M-IGLS P times .
2. Identify “reliable” Estimate by Criterion 2.
3. Execute M-IGLS Initialized with average “reliable” signature estimate.

5. Conclusion

This paper provides a review on secure spread spectrum embedding of data into a digital medium and extracting the data using M-IGLS and CC-MIGLS algorithms .Based on literature survey provided in this paper, this paper gives the method for embedding the data using direct sequence spread spectrum technique to embed the data and gives the method for blindly extracting the hidden data embedded over a wide band of spectrum or transformed domain of digital medium. To extract the hidden data blindly we have used the Multi Carrier Iterative Generalized Least square (M-IGLS) procedure to seek the unknown data hidden in the image host. Neither the original host nor the embedding carriers assumed available. Then to enhance the MIGLS I have used the Cross Correlation MIGLS(CC-MIGLS) procedure, which leads to most effective hidden message recovery by performing the statistical analysis of repeated independent MIGLS processing of the host.

References

[1] **A Survey of Security Mechanisms with Direct Sequence Spread Spectrum Signals**, Journal of Computing Science and Engineering, Vol. 7, No. 3, September 2013, pp. 187-197
[2] By : Taeho Kang Department of Computer Science and Engineering, Pohang University of Science and Technology (POSTECH), Pohang, Korea
[3] **Extracting Spread-Spectrum Hidden Data from Digital Media** Ming Li, *Member, IEEE,* Michel Kulhandjian, Dimitris A. Pados, *Member, IEEE,*Stella N. Batalama, *Senior Member, IEEE,* and Michael J. Medley, *Senior Member, IEEE*
[4] **Active Spread-Spectrum Steganalysis for Hidden Data Extraction** Ming Li, Michel Kulhandjian, Dimitris A. Pados†, Stella N. Batalama Department of Electrical Engineering State University of New York at Buffalo

Licensed Under Creative Commons Attribution CC BY

Buffalo, NY 14260 E-mail: {mingli,mkk6,batalama,pados}@buffalo.edu

[5] **Information hiding a survey**, by fobien A.P petitcolas, Rosc J. Anderson and markus G. Kuhan

[6] G. C. Langelaar, I. Setyawan, and R. L. Lagendijk, "**Watermarking digital image and video data: A state-of-the-art overview**," *IEEE Signal Processing Magazine*, vol. 17, pp. 20-46, Sept. 2000.

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

[8] S. Wang and H. Wang, "**Cyber warfare: Steganography vs. steganalysis**," *Communications of the ACM*, vol. 47, pp. 76-82, Oct. 2004.

[9] C. Cachin, "**An information-theoretic model for steganography**," in *Proc.2nd Intern. Workshop on Information Hiding*, Portland, OR, Apr. 1998, pp. 306-318.

[10] G. J. Simmons, "**The prisoner's problem and the subliminal channel**," in *Advances in Cryptology: Proc. CRYPTO'83*. New York, NY: Plenum, 1984, pp. 51-67.

[11] J. Fridrich, ***Steganography in Digital Media, Principles, Algorithms, and Applications***. Combridge, UK: Combridge Univeristy Press, 2010.

[12] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, "**Information hiding:A survey**," *Proc. IEEE (**Special Issue on Identification and Protectionof Multimedia Information**)*, vol. 87, pp. 1062-1078, July 1999.

[13] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*. San Francisco, CA: Morgan-Kaufmann, 2002.

[14] F. Hartung and M. Kutter, "**Multimedia watermarking techniques**," *Proc.IEEE (Special Issue on Identification and Protection of Multimedia Information)*, vol. 87, pp. 1079-1107, July 1999.

Author Profile

Prof. Vaibhav Dhore received the B.E. degree from the Government College of engineering Amravati and M.Tech Degree in Computer Engineering from NIT Jaipur. He is working as Assistant Professor in Department of Computer Engineering, RMD Sinhgad School of Engineering Pune, India. He is having more than four year experience. His research interest is distributed computing and big data.

Pathan Md.Arfat is a Research Scholar of RMD Sinhgad school of Engineering, University of Pune. He has received B.E. in Computer Engineering from BAMU University Aurangabad. Currently he is pursuing M.E. in Computer Engineering from RMD Sinhgad School of Engineering, Pune, University of Pune, Pune , Maharashtra, India

Licensed Under Creative Commons Attribution CC BY