Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 7, July 2015, pg.253 – 262

RESEARCH ARTICLE

Extracting Spread Spectrum Data from Image Using MLS Algorithm

Sahej Redekar
Department of Computer Engineering, Padmashree Dr. D.Y. Patil Institute of Engineering & Technology,
Savitribai Phule Pune University, India
Sahej1992@gmail.com

Prof. Dr. Reena Gunjan
Department of Computer Engineering, Padmashree Dr. D.Y. Patil Institute of Engineering & Technology,
Savitribai Phule Pune University, India
reenagunjan@gmail.com

ABSTRACT:

Data embedding and extraction schemes are increasing rapidly to protect against the fast increment of data tampering attacks. So a robust data hiding scheme is required to get security from these attacks. The problem of extracting blind data embedded in a spectrum domain of digital medium (image) is considered. Multicarrier Least Square (MLS) algorithm is developed to extract unknown data hiding in the image via multicarrier SS embedding. This paper proposes a blind extraction technique. In this technique, the original host and embedding carriers are not available. The data is embedded in the image via DCT multicarrier DSSS (Direct Sequence Spread Spectrum). Multicarrier Least Square algorithm is the extraction algorithm used to extract the hidden data from image. MLS is provides high performance and the complexity of algorithm is low. Cross correlation enhanced MLS is based on statistical analysis of repeated independent MLS processing of the host, it offers most effective hidden message recovery.

INTRODUCTION:

Data hiding is used for safe communication which is used in military communication like encrypted message is used by sender and receiver for hiding its existence. Data tracking are growing quickly in everywhere like mobile tracking. Because of these attacks for transmitting data secure communication schemes are used. There are multiple data hiding and extraction schemes used for the secure communication. This paper mainly focuses on blind extraction of hidden data in host image via spread spectrum. The original host and embedding carriers both are not available, so here the fully blind extraction is considered. Watermarked content only (WOA) attack in watermarking security is also referred as blind extraction of data.

The unknown host acts as a source of disturbance in blind extraction of SS embedded data to be recovered and the problem parallels blind signal separation (BSS) applications as they arise in the fields of array processing, biomedical signal Processing. ICA-based BSS algorithms are not effective in the presence of correlated signal interference as is the case in Spread Spectrum multimedia embedding and reduce immediately as the dimension of the carrier decreases relative to the message size. A least squares (LS) procedure was developed to blindly recover unknown messages hidden in image hosts via SS embedding. However, the algorithm was developed for single carrier SS embedding where messages were hidden with one signature only and was not generalizable to the multicarrier case. Realistically, an embedder would favour multicarrier SS transform-domain embedding to increase security and/or payload rate. In this paper, a Multicarrier Iterative Generalized Least Square (MLS) algorithm is used for extraction of hidden data and to improve the performance. The complexity of MLS is low and it provides high performance. The greatest challenge is to extract the small hidden messages. The proposed algorithm is well-known to test security of Spread Spectrum data hiding. Characteristics of data hidings are a) payload b) robustness c) transparency d) security. Active hidden data extraction is relatively new research while passive detection-only of the presence of embedded data is investigated in past years. In blind extraction of SS embedded data the unknown host acts as disturbance to the improvement of data and the blind signal separation appear in the array processing and CDMA communication terms. In this paper, our attention is on the blind recovery of secret data hidden in medium hosts via multicarrier/signature direct sequence spread-spectrum (DS-SS) transform domain embedding. The embedded hidden messages are distributed on identical basis. For hidden Data extraction independent component analysis is done.

Applications of the developed algorithm are not for attacking steganographic covert communication by recovering hidden embedded messages. All the carriers are provided with embedded data. Hence to enhance the recovery performance of MLS algorithm we are using Cross Correlation MLS algorithm (CC-MLS). CC-MLS is nothing but a statistical analysis of independent MLS executions on the host. CC-MLS can achieve the recovery of the hidden data with probability of error very close to be attained with known embedding signatures and the autocorrelation matrix of the known original host.

RELATED WORK:

In the existing system reversible data hiding technique the image is compressed and encrypted by using the encryption key and the data to hide is embedded in to the image by using the same encryption key. The user who knows the secret encryption key used can access the image and decrypt it after extracting or removing the data hidden in the image. After extracting the data hidden in the image then only can be the original image is retrieved.

In the traditional architecture there existed only the server and the client. In most cases the server was only a data base server that can only offer data. Therefore majority of the business logic i.e., validations etc. had to be placed on the clients system. This makes maintenance expensive. Such clients are called as “fat clients”. This also means that every client has to be trained as to how to use the application and even the security in the communication is also the factor to be considered. Since the actual processing of the data takes place on the remote client the data has to be transported over the network, which requires a secured format of the transfer method. How to conduct transactions is to be controlled by the client and advanced techniques implementing the cryptographic standards in the executing the data transfer transactions. Present day transactions are considered to be "un-trusted" in terms of security, i.e. they are relatively easy to be hacked. And also we have to consider the transfer the large amount of data through the network will give errors while transferring. Nevertheless, sensitive data transfer is to be carried out even if there is lack of an alternative. Network security in the existing system is the motivation factor for a new system with higher-level security standards for the information exchange.

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. However 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 encodes the information by changing the several statistical properties of a cover and uses the hypothesis testing in the extraction.

5) **Distortion technique**: This method stores 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 secret communication is created.

Disadvantage of Existing system:

1) It needs lot of effort for hiding a few bits of information.
2) The user can access the image when he knows the secret encryption key and decrypt it after extracting or removing the data hidden in the image.
3) Frequency planning is needed.

PROPOSED SYSTEM:

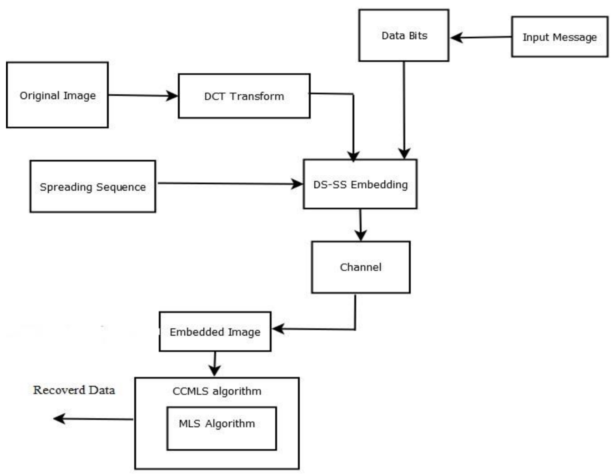

Figure 1. Workflow of system

In this paper, the problem of extracting blind data embedded over a wide band in a spectrum (transform) domain of a digital medium (image) is considered. A original multicarrier iterative generalized least-squares (MLS) core procedure to find unknown data hidden in image via multicarrier spread spectrum embedding and improved spread spectrum embedding is developed. The original host and the embedding carriers are not available. The proposed scheme is also used as a performance analysis tool for the data hiding schemes. And also to enhance the performance of MLS algorithm, we are using Cross Correlation MLS procedure (CC-MLS). CC-MLS is nothing but a statistical analysis of independent MLS executions on the host. CC-MLS can achieve the recovery of the hidden data with probability of error very close to be attained with known embedding signatures and the autocorrelation matrix of the known original host.

Advantages of Proposed system:

1) Resistance to noise attack.
2) The data is embedded in the host signal itself so no separate broadcast channel is required which is one of the main advantages of the technique.

3) No need for frequency planning.
4) High resilience and variable data rate transmission.

Our objective is to blindly extract the unknown Hidden data B from the Observation matrix Y without prior knowledge of the embedding carriers S_k and amplitude A_k.

MODULE DESCRIPTION:

1) Partitioning of Host Image: The message is hidden in digital media like image. Image is taken as host for hiding the data. Image can be either gray scale image or colour image. Image is not divided into overlapping blocks. Each block carries hidden message bits. The image is divided into 6 non overlapping blocks. These blocks are independently processed for embedding. Then, the embedded blocks are synchronised.

2) Application of DCT: The DCT transform for image transformation is taken. It is important that DCT transformation provides outstanding energy compaction in low spectral coefficients for highly correlated data. Any interruption directly or indirectly added in the frequency domain may result in a change of statistical properties. DCT is applied in blocks of matrix.

3) Multicarrier SS embedding: The main feature of spread spectrum is its ability to stand with strong conflict, sometimes generated by a third party to block the communication. This is one reason of using spectrum spreading in military communications. The embedding method is designed to satisfy the perceptual constraints and improve the delectability as well as the embedding rate. In this embedding technique, the hidden data is spread across many samples of host signal or image by adding the DCT coefficient as the carrier. SS embedding algorithms for blind image (that is, hidden message recovery without knowledge of the novel image) have been based on the understanding that the host signal acts as a source of interference to the secret message of interest. It should also be understood that this interference is known to the message embedded. Such knowledge can be exploited appropriately to facilitate the task of the blind receiver at the other end and decrease the recovery error rate for a given host misinterpretation level. Advanced embedding methods can facilitate host interference suppression at the receiver side when knowledge of the host signal is adequately exploited in system design.

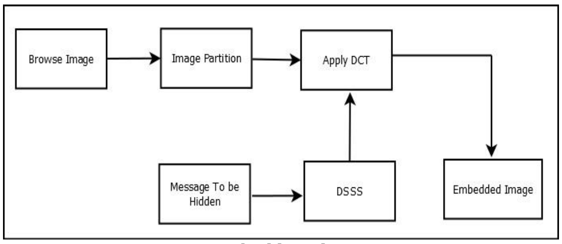

Figure 2: Embedding data in image

4) Data extraction using MLS: After detecting the image, if the image contains data, then it is to be extracted. So an algorithm known as Multicarrier Least Squares is used. This contains the following steps. Each time a least squares is computed. The unknown matrices are updated and modified conditioned on a previously obtained estimate for the other matrix. This is repeated until convergence of the least squares cost function is reached. Convergence is the property that a different transformation of the same state has a transformation to the same end state. Convergence of the least square cost is guaranteed since each update may either improve or maintained. The final result is generally dependent on the initialization. Least Squares (LS) is a technique for estimating the unknown parameters in linear regression model. LS is applied when variances of observation are unequal or there is certain degree of correlation between the observations. Linear regression is approach for modelling the relationship between scalar dependent variable X and one or more explanatory variables. Dependent variable is measured variable Y.
Least Square is standard approach to the approximate solution of over determined system. Over determined system, in maths system of linear equations is considered over determined if there are more equations than unknown.

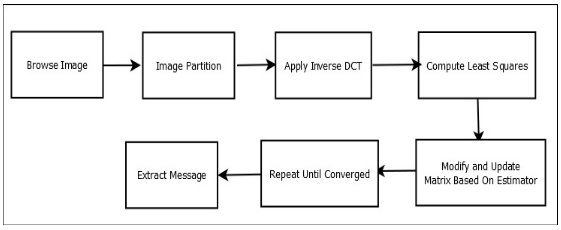

Figure 3: Extracting data from image using MLS Algorithm

MLS Algorithm is having following steps.

1. Each time compute a least squares.
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.

MLS ALGORITHM:

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

2. d:=d+1;

$\hat{V}^{(d)}$:= $Y(\hat{B}^{(D-1)})^T\ [(\hat{B}^{(D-1)})\ (\hat{B}^{(D-1)})^T]^{-1}$
$\hat{B}^{(d)}$:=sgn $\hat{R}_y^{-1}\left(\hat{V}^{(d)}\right)^{-1}\left(\hat{V}^{(d)}\right)^T\ \hat{R}_y^{-1}\ \ Y\}$

3. Repeat step 2 until $\hat{B}^{(d)} = \hat{B}^{(D-1)}$.

5) CC-MLS: For detection and Extraction process we are using The CC-MLs. As the CC-MIGLS is nothing but a statistical analysis of independent MLS executions. CC-MLS can achieve the recovery of the hidden data with probability of error very close to be attained with known embedding signatures and the autocorrelation matrix of the known original host. The term cross-correlations is used for referring to the correlations between the entries of two random vectors *X* and *Y*, while the autocorrelations of a random vector *X* are considered to be the correlations between the entries of *X* itself, those forming the correlation matrix (matrix of correlations) of *X*.

The sample cross-correlation between any two bit streams is:

Criterion 1 : $n_{ij} = \hat{b}_i^T \hat{b}_j | M$, i≠j, i, j = 1,2,..........k

The basic idea behind the final refinement of MLS blind 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 MLS procedure.

Criterion 2 : for signatures k and runs p can be defined as
$$\rho_{k,p} \triangleq \sum_{j=1, j\neq p}^{p} \frac{|\hat{V}^{T}_{k,p}\hat{V}_{K,j}|}{\|\hat{V}_{K,P}\|\|\hat{V}_{K,P}\|}$$

Where,
is p-th run estimate of Vk.
Now set threshold,

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

Then $\hat{V}_{k,p}$ consider "reliable" estimate of V_k otherwise "unreliable".

CCMLS ALGORITHM:

For j := 1 to P

1) Execute MLS with arbitrary initialization and obtain estimates $\widehat{V_k}$, k= 1.....K.

2) If estimate are criterion-1- compliant.

$\hat{V}_k^{(j)} := \widehat{V_k}$, k=1,.......K;

Else go to 1

End

For k: 1 to K

3) Identify reliable estimates for V_k according to criterion 2.

4) Calculate the average over all reliable estimates $\widehat{V_k}$

End

5) Set $\overline{\hat{\hat{V}}} \triangleq [\, \overline{\hat{\hat{V}}_1} \ldots\ldots\ldots \overline{\hat{\hat{V}}_k} \,]$.
6) Execute MLS with initialization.

$\hat{B}(0) = \text{sgn}\{ (\hat{\bar{V}}^T \hat{R}_y^{-1} \hat{\bar{V}}^T)^{-1} \hat{\bar{V}}^T \hat{R}_y^{-1} Y \}$

Steps:

1. Run Criterion 1 equipped M-IGLS P times .
2. Identify "reliable" Estimate by Criterion 2.
3. Execute MLS Initialized with average "reliable" signature estimate.

MATHEMATICAL MODEL:

Input Data: Host Image (H)

Step 1: Let H be the host image.

H= {H1,H2,....H_M}

$H \epsilon M^{N1*N2}$

Where,

N1, N2 = Image size in pixel

M= alphabets

Step 2: Apply DCT.

T(H_m)∈ $\mathbb{R}^{\frac{N1N2}{M}}, m = 1,2, \dots \dots . . M.$

Step 3: By choosing fixed subset of L ≤ $\frac{N1N2}{M}$

Final host vector will be X(m) ∈ $\mathbb{R}^L$, m=1,2,.......M.

Step 4: Let M be the message to be embed with K distinct message bit sequence.

M= $\{b_k(1), b_k(2)$,......... $b_k(M)\}$

Where,

k= 1, 2........K, $b_k(m) \in \{\pm 1\}$, $m = 1 \dots \dots \dots, M$

Step 5: Apply SS embedding with following equation.

$$y(m) = \sum_{k=1}^{k} A_k B_k(m) S_k + X(m) + n(m), \qquad m = 1,2,3 \dots \dots, M$$

Where,

S_k = sequence carrier

A_k = Amplitude

B_k= bit

X = Original host data

M= Message bit Sequence

n (m)= Potential external while Gaussian noise.

Step 6: z(m) combined disturbance to hidden data

$$y(m) = \sum_{k=1}^{k} A_k B_k(m) S_k + z(m), \qquad m = 1,2,3 \dots \dots, M$$

Where ,

Z(m)= sequence of zero mean

Step 7: Let $v_k \triangleq A_k\ s_k \in \mathbb{R}^L$,

k=1,.....,K be the amplitude including embedding carriers.

Reformulate SS embedding as

$y(m) = \sum_{k=1}^{k} b_k(m)v_k + z(m)$

=Vb (m) +z (m), m=1... M,

Where,

V ≜ [v1,.....,v_K]∈ $\mathbb{R}^{L\times K}$ is amplitude including carrier matrix.

b(m) ∈ $\{\pm 1\}^{K\times 1}$ bits embedded in block.

Step 8: Extract Hidden data B from the Observation matrix Y without prior knowledge of the embedding carriers Sk and amplitude Ak.

Y=VB+Z

Where,

Y = y(m)
B= unknown hidden data
V=Vector matrix
V=$A_1\ S_1$,.... ,$A_k S_k$
Z = [x(1) + n(1), . . . , x(M) + n(M)].

Results:

Fig. 4 shows the embedding of data in the image using spread spectrum embedding. Embedded data is converted into bits and then embedded in each partition of image. Fig. 5 shows same data is extracted from image using multicarrier least square algorithm and number of iterations required to extract embedded data.

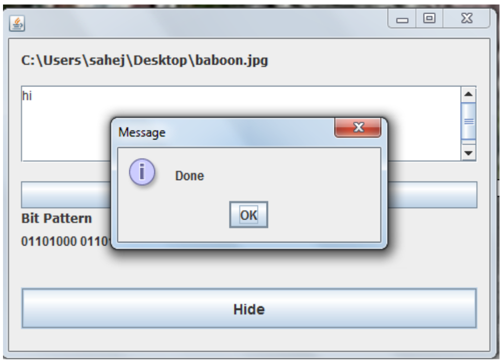

Figure 4. Data embedded in image

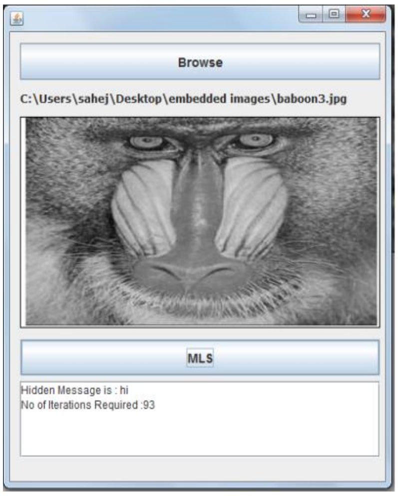

Figure 5. Extracted data from image using MLS

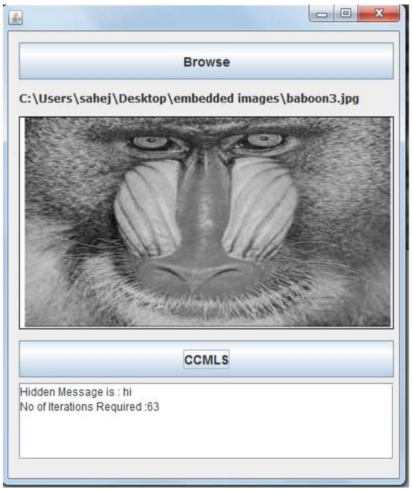

Figure 6. Extracted data from image using CCMLS

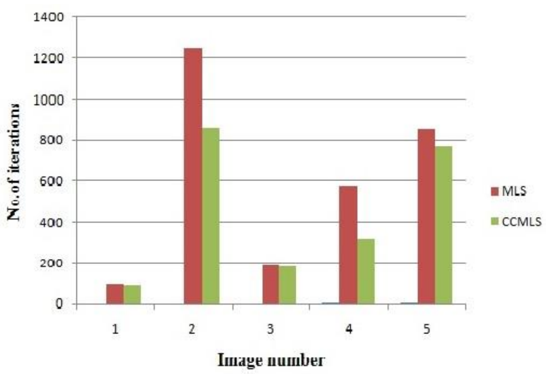

Figure 7. Graph of image number verses no. Of iterations

CONCLUSION:

The proposed system provides a good extraction technique which considered the blind recovery of data. Neither the original host nor the embedding signatures are assumed available. To extract the spread spectrum embedded data from a given host image, the host is converted to observation vectors. Then MLS algorithm is used for the extraction. The MLS extraction provides high peak signal to noise ratio and it attains the probability of error recovery.

We first developed a low complexity multi signature least-squares (MLS) core algorithm. Cross correlation enhanced MLS (CC-MLS), a procedure based on statistical analysis of repeated independent MLS processing of the host, is seen to offer most effective blind hidden message recovery and presents itself as an effective countermeasure to conventional SS data hiding.

REFERENCES:

[1] F. Cayre, C. Fontaine, and T. Furon, "Extracting Spread-Spectrum Hidden Data From Digital Media," IEEE Transactions On Information Forensics And Security, Vol. 8, No. 7, July 2013.

[2] F. Cayre, C. Fontaine, and T. Furon, "Watermarking security: Theory and practice," IEEE Trans. Signal Process., vol. 53, no. 10, pt. 2, pp. 3976-3987, Oct. 2005.

[3] T. Li and N.D. Sidiropoulos, "Blind digital signal separation using successive interference cancellation iterative least squares," IEEE Trans. Signal Process., vol. 48, no. 11, pp. 3146-3152, Nov. 2000.

[4] 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.

[5] C. Qiang and T. S. Huang, "An additive approach to transform-domain information hiding and optimum detection structure," IEEE Trans. Multimedia, vol. 3, no. 3, pp. 273-284, Sep. 2001.

[6] 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.

[7] 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.

[8] Y. Wang and P. Moulin, "Optimized feature extraction for learning based image steganalysis," IEEE Trans. Inf. Forensics Security, vol. 2, no. 1, pp. 31-45, Mar. 2007.

[9] 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.

[10] W. Lie and G. Lin, "A feature-based classification technique for blind image steganalysis," IEEE Trans. Multimedia, vol. 7, no. 6, pp.1007-1020, Dec. 2005.

[11] L. Prez-Freire and F. Prez-Gonzlez, "Spread-spectrum watermarking security," IEEE Trans. Inf. Forensics Security, vol. 4, no. 1, pp. 2-24, Mar. 2009.

[12] 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.

