IJDCST @June-July-2015, Issue- V-3, I-5, SW-07
ISSN-2320-7884 (Online)
ISSN-2321-0257 (Print)

An Efficient and Scalable Multi-signature Iterative Generalized Least-Squares for Hidden Data Extraction

[#1]Panchumarthy Bhaskar, [#2]S.Suresh Babu
[#1]M.Tech Student, Sri Mittapalli college of Engg, #2 Assistant Professor, Dept of CSE
#1bskrmail13@gmail.com, #2cse.sureshbabu@gmail.com

Abstract: Digital data from digital media is the major source contains text, images and video etc. Data Extraction (Knowledge Extraction) from digital media corpus becomes commercial interest for business entrepreneurs to mine customer interest and for national security purpose also. As on many former researchers were implemented various knowledge extraction techniques are suffering from recovery probability. In this paper we introduced an Efficient and Scalable Multi-signature Iterative Generalized Least-Squares (ESM-IGLS) algorithm to improvise the recover probability to mine expected information. We designed the embedding signatures and host auto-correlation matrix to achieve scalability and relevance in terms of knowledge engineering from wide band digital media spectrum. Experiments on real digital media corpus data results are showing that our ESM-IGLS algorithm is having the high-level efficiency and scalability over other existing techniques.

Keywords: Annotation, Data Hiding, Information Hiding, Spread-Spectrum Embedding, Authentication, Blind Detection

1. Introduction

With the advent of digitalization, embedding the personal and public data in digital media (data hiding [1]) become an interest. In order to provide the security for this digital data sake they are using watermarking [2], copyrights, read-only and iron-binding etc. Along with security these technologies also embeds the secondary information for the sake of providing side information like Meta data. A famous Greek steganography technique is called "covered writing" means covering the actual data by using the digital objects like watermarks, images[10], metadata, video and audio etc is used to share the data among the authorized people. Although many data hiding methods [6 and 7] were designed to implement the information hiding, each approach should be compared against the below properties of information hiding are: PayLoad, Robustness, Transperency and Security. These four are the main attributes of information hiding to determine the efficiency and scalability.

Majority of the past information hiding and evaluating technologies were suffering from the threats on privacy, disclosures, complexity and security. More complex design of information hiding technique may help to design more secured encryption but leads to slow decryption and accessibility due to complexity. Similarly the design of hiding information with less complexity causes to less secure, but fast accessibility and decryption. Coordinating these both become a challenging issue in, the design of information hiding algorithm.

In this paper we are concentrating on implementing the algorithm for recovering the hidden data from the given medium in a scalable and secured way with the help of blind recovery procedure. To achieve this we proposed an Efficient and Scalable Multi-signature Iterative Generalized Least-Squares [8 and 9] (ESM-IGLS) algorithm to improvise the recover probability to mine expected information. We designed the embedding signatures and host auto-correlation matrix to achieve scalability and relevance in terms of knowledge engineering from wide band digital media spectrum.

The main contributions of this paper are:

- Along with passive extracting the active hidden data efficiently with ESM-IGLS algorithm
- Extending the capability of M-IGLS algorithm with more efficiency and scalability
- Design of new ESM-IGLS algorithm to extract even identically distributed independent data.
- SS embedding used to introduce the generalized multi-career case.

The main aim of ESM-IGLS algorithm is used to recover the data from SS hidden data, which was a challenging task since a decade. ESM-IGLS uses the re-initializations to deal with the small messages with huge complexity of challenges efficiently with scalability also. This process is enough robust to resist from tampering attacks and stenographic conversions.

Related work

In this section we discuss about the basic information, terminology and concepts which are required and used through out of this research paper.

A) Information Hiding

In this case we are assuming that the information hiding with the media object is image means image level or image oriented information hiding. Initially an image K is a collection of pixels scattered through a canvas for representation. Each image is having the size as N1 and N2 and the host image is divided to the non-isolate block are having the size (N1 * N2) / M. The block representation for the given image is represented as K1, K2,K3 …. Kn which are belongs to an image K. After this the process of embedding secret information will starts and the bits of data will be embedded as chunks with the real encryption techniques are 2-D transform domain T.

The below diagram shows the examples for the sample host images [10] with hidden information.

Figure1. 256 X 256 Gray Scale Host images with Hidden Data

B) SS Embedding

We consider K distinct message bit sequences, {bk(1), bk(2), . . . , bk(M)}, k = 1, 2, . . . ,K, bk(m) ∈ {±1}, m = 1, . . . ,M, each of length M bits. The K message sequences may be to be delivered to K distinct corresponding recipients or they are just K portions of one large message sequence to be transmitted to one recipient. In particular, the mth bit from each of the K sequences, b1(m), . . . , bK(m), is simultaneously hidden in the mth transform-domain host vector x(m) via additive SS embedding by means of K spreading sequences (carriers) sk ∈ RL, kskk = 1, k = 1, 2, . . . ,K.

Gaussian noise1 of mean 0 and auto correlation matrix σ2 (nIL), σ2 (n > 0). It is assumed that bk(m) behave as equi-probable binary random variables that are independent in m (message bit sequence) and k (across messages). The contribution of each individual embedded message bit bk to the composite signal is Ak(bk(sk)) and the block mean-squared distortion to the original host data x due to the embedded k message. Under statistical independence of messages, the block mean-squared distortion of the original image due to the total, multi-message, insertion of data

C) What is Hidden Extraction?

Hidden extraction stands for extracting the actual secret information from the host data which covers the information. In order to implement this process we introduced a blind hidden extraction method which divides the host data to a set of interrelated and singular vectors. These vectors contain the meta data of host image as partitions boundary values, co-related matrices, intellectual data transforms and the collection of neighbor information.

We used the SS attribute population algorithms to carry out the estimated attribute relations in a jumbling manner. Finally, determination of the transform domain used in embedding seems to be a hurdle not yet tackled by current research. The natural approach would be to consider individually and exhaustively one transform at a time starting from the most common (for example, 2D-DCT [5], common wavelet transforms [6], and so on).

2. Efficient and Scalable Multi-signature Iterative Generalized Least-Squares (ESM-IGLS) algorithm

In this section we discuss about our implementation algorithm ESM-IGLS in detail to describe the working efficiency and scalability. We use the Guassian model to identify the disjoint Maximum Likelihood (ML) for the estimation of text T and decode of host image Hn. The whole algorithm we described here as like this:

Algorithm: ESM-IGLS

Input: Host Data Image M with hidden information

Output: Hidden Information K

1. Begin
2. Split the image M to vector set Vn
3. Estimate Niegbours information and embed metadata
4. Apply SS signal transform on each Mi
5. For i=0 to n {
6. Apply 2D-DCT for Mi

7. }
8. If (M > 0)
9. Construct auto-covariance matrix with all vector sets for Vi to Vn
10. Now transform the subset to actual info by using blind extraction
11. End

This algorithm ESM-IGLS takes the host data image with hidden information as input value and returs the hidden information which is extracted using ESM-IGLS co-relation pattern mining [4]. To find the co-relations first the input data will be divided to a set vector elements. Each element consists of the equal partitioned image with the specified attributes as follows: Here R is an interact co-relation designed to process the text I from the given host image subset V. The given below model is showing how the extraction is happening while implementing the expression for hidden data.

$$R_y^{-1} = R_z^{-1} - R_z^{-1}V(I + V^T R_z^{-1} V)^{-1} V^T R_z^{-1}.$$

$$\begin{aligned} V^T R_y^{-1} V &= V^T R_z^{-1} V - \\ &\quad V^T R_z^{-1} V(I + V^T R_z^{-1} V)^{-1} V^T R_z^{-1} V \\ &= V^T R_z^{-1} V[I - (I + V^T R_z^{-1} V)^{-1} V^T R_z^{-1} V] \\ &= V^T R_z^{-1} V(I + V^T R_z^{-1} V)^{-1} \\ &\quad [(I + V^T R_z^{-1} V) - V^T R_z^{-1} V] \\ &= V^T R_z^{-1} V(I + V^T R_z^{-1} V)^{-1}. \end{aligned}$$

The above expression set is representing the methodologies and processing operations in a detailed manner for the given hidden image partition V with the relation R and mapped text I in an elegant manner. This same process we implemented to perform the experiments on it.

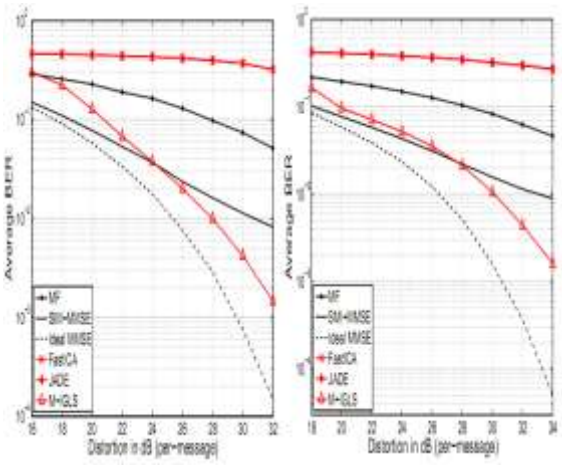
3. Experimental Results

In this section we discuss about the implementation of experiments on hidden data from host images. We implemented this algorithm from MATLAB and JAVA combination to get the information from experiments. We collected almost 170 host images which contain the hidden data information with complex challenge based encryption. Once set up the environment we started the data extraction from media as host images by using our algorithm.

A technically firm and keen measure of quality of a hidden-message extraction solution is the difference in bit-error-rate (BER) experienced by the intended recipient and the analyst. The intended recipient in our studies may be using any of the following three message recovery methods: (i) Standard carrier matched-filtering (MF) with the known carriers sk, k = 1, ...,K; (ii) sample-matrix-inversion MMSE (SMI-MMSE) filtering with known carriers sk and estimated host autocorrelation matrix.

Soon after processing is completed for the given set of images we analyzed the dissertation value for each image, data and image partition in detail and the results information we shown as below graph 1. WE compared the result against the other popular extraction methods like MF[8], SMI-MMSE[7], IdealMMSE [3 and 7], FastiCA [9] and JADE [5] as shown below:

The respective Distortion information is shown with respective to the above figure1 with below graph model is:



Experiments on real digital media corpus data results are showing that our ESM-IGLS algorithm is having the high-level efficiency and scalability over other existing techniques.

4. Conclusion

In this paper we are concentrating on implementing the algorithm for recovering the hidden data from the given medium in a scalable and secured way with the help of blind recovery procedure. To achieve this we proposed an Efficient and Scalable Multi-signature Iterative Generalized Least-Squares (ESM-IGLS) algorithm to improvise the recover probability to mine expected information. We designed the embedding signatures and host auto-correlation matrix to achieve scalability and relevance in terms of knowledge engineering from wide band digital media spectrum.

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, pp. 1062-1078, July 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 Processing Magazine, vol. 17, pp. 20-46, Sept. 2000.
3) J. Fridrich, Steganography in Digital Media, Principles, Algorithms, and Applications. Combridge, UK: Combridge Univeristy Press, 2010.
4) D. G. Manolakis, V. K. Ingle, and S. M. Kogon. Statistical and adaptive signal processing: Spectral estimation, signal modeling, adaptive filtering and array processing. Boston, MA: McGraw-Hill, 2000.
5) T. Li and N. D. Sidiropoulos, “Blind digital signal separation using successive interference cancellation iterative least squares,” IEEE Trans. Signal Proc., vol. 48, pp. 3146-3152, Nov. 2000.
6) J. Hernandez, M. Amado, and F. Perez-Gonzalez, “DCT-domain water-marking techniques for still images: Detector performance analysis and a new structure,” IEEE Trans. Image Proc., vol. 9, pp. 55-68, Jan. 2000.
7) M. Li, S. N. Batalama, and D. A. Pados, “Population size identifica-tion for CDMA eavesdropping,” in Proc. IEEE Military Comm. Conf. (MILCOM), Orlando, FL, Oct. 2007, pp. 1-6.
8) C. Fei, D. Kundur, and R. H. Kwong, “Analysis and design of watermarking algorithms for improved resistance to compression,” IEEE

Trans. Image Proc., vol. 13, pp. 126-144, Feb. 2004.

9) A. Valizadeh and Z. J. Wang, "Correlation-and-bit-aware spread spectrum embedding for data hiding," IEEE Trans. Inform. Forensics and Security, vol. 6 , pp. 267-282, June 2011.

10) G. Schaefer and M. Stich, "UCID–An uncompressed colour image database," in Proc. SPIE, Storage and Retrieval Methods and Applications for Multimedia, San Jose, CA, Jan. 2004, pp. 472-480.