ISSN:0975-9646

A Novel Approach for Data Leakage Detection Using Steganography

P.P.Dandavate[#1], Prof.S.S.Dhotre[#2]

[#]*Computer Engineering Department*

Bharati Vidyapeeth Deemed University,Pune

***Abstract*—Now a days data is transmitted over internet. During data transmission it becomes necessary to provide security to message.Steganography algorithms are used to provide security to data .To conceal the existence of hidden secrete data inside a cover object steganography is one of the powerful technique. In this paper comparison of three different steganography algorithm is done using clustering based approach to provide security to data and to detect leakage if data gets leaked .Clustering is used to group pixels for embedding process.**

***Keywords*— data leakage,steganography,k-means clustering algorithm.**

I. Introduction

Unintentional distribution of sensitive data to unauthorized person is the data leakage. In business process sometimes sensitive data is given to supposed third party. Owner of data is called distributor and supposed third party is called agent. During business process if there is data leakage then we have to detect agent who leaked data.[1]The technique to hide the information in some cover media so that attacker cannot identify that information is hidden in cover media is called steganography.In information security steganography plays important role. Now a days in steganography image has been used as a carrier to transmit or send the secrete message from sender to a receiver. An image can be defined as array of numbers in a computer that represent light intensities at various points. (pixels.). Image's raster data is made by these pixels.24-bit (true color image) or 8-bit per pixel files are used to store digital images.640*480 pixels and 256 colors(or 8 bits per pixel)are common size image. Such an image contains 300kb of data. When sending over the network or internet such large size images are avoided.[2]

As compared to other steganography image steganography has attracted researchers. This is because in image a huge amount of information can be hidden without noticeable impact to the image which is used as carrier. Second reason is that use of image in information hiding provides security because to human visual system digital image is insensitive. [3]

II Literature Survey

Simple LSB substation technique is used and implementation of OPAP has increased security and complexity is reduced [5].Embedding capacity and image quality is achieved by using new steganography approach with revised LSB substitution and pixel differencing. Private k-bit modified substitution techniques is implemented for embedding.[4].To enhance security for color image steganography is implemented with combination of OPAP and PI technique and pixel value for content hiding. By using this

MSE is reduced and efficiency is increased.[5] An efficient hiding approach is proposed for content hiding. In this approach 8*8 image is taken and DCT is applied on it and in diagonal pixels secrete message is embedded and in place of text random bits are replaced.[6].Scan path based approach are discussed in [7].

III System Architecture

Fig.1 shows the system architecture.

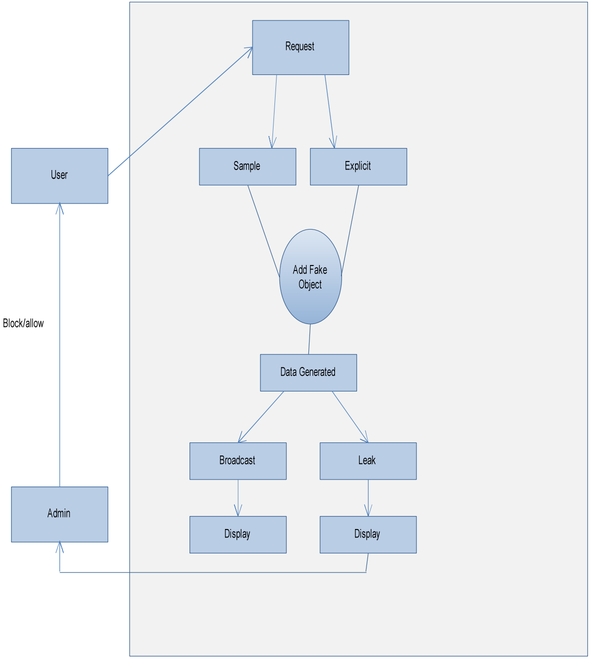

Fig.1 system architecture

In this leakage detection system agent will first do registration. After registration, admin will either give approve or decline to agent. If approval is given by administration then agent will request for data i.e. image by either sample request or explicit request. Data (image) will be given to agent by adding unique used ID using different steganography LSB algorithm. After adding secret data, image will be stegano image. After receiving data agent will broadcast data to authorized channel or leak data to unauthorized channel. If data is leaked system will find which agent has leaked data. That agent is called guilty agent.Admin will not allow guilty to enter in the system. Guilty agent will be placed in black list by administrator.

1) *Assumption1:* .Each agent should have unique id
2) *Assumption 2:* Data should be leaked by registered agent only.
3)

IV .STEGANOGRAPHY

When communication between sender and receiver takes place a method to hide secrete message in a cover media is called steganography.Steganography is Greek word meaning concealed writing. Meaning of steganos means "covered" and meaning of "graphical" means "writing".Steganography is not only used for hiding data but also hiding the secrete data of transmission. When secret data is added in a file its existence is known only to receipent.Steganography has three elements: the secret message, stegano image (message is embedded in cover object) and cover image (it contains secrete message). In ancient time protection of data is done by hiding data on of wax, on scalp of the slaves, writing tables. But in today's world data is transmitted in the form of text, images, video, and audio. Over the medium. Audio, video images are used as cover sources to hide data for safely transmission of confidential data.[8] In this paper image steganography is implemented.

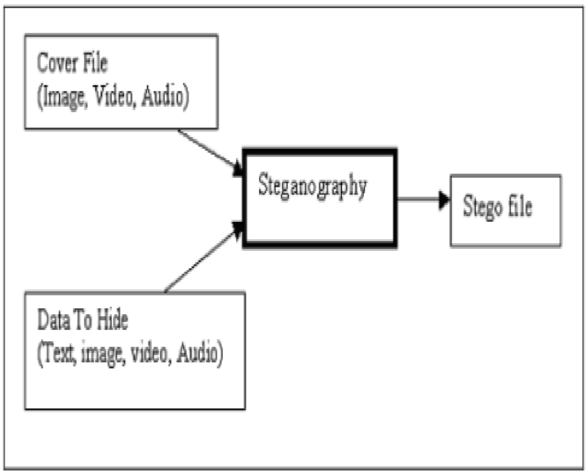

Fig.2 Process of hiding data

A. *Types of steganography*

1) Text *Steganography:* In text steganography information is hidden in text. Every nth letter of every words of text message is taken to hide data inside text file. Various methods are available for hiding data in text file.i) Random and statistical method ii) Linguistics Method iii) Format Based Method.

2) *Image Steganohraphy*: In image steganography for hiding data image object is taken. To hide data pixel insenties are taken.Mostely in digital steganography, images are widely used cover media.

3) *Audio Steganography:* In audio file data is hidden in audio file. Data is hidden in WAV, AU and MP3.There are various methods of audio steganography i) Low bit Encoding ii) Phase Coding iii) Spread Spectrum.

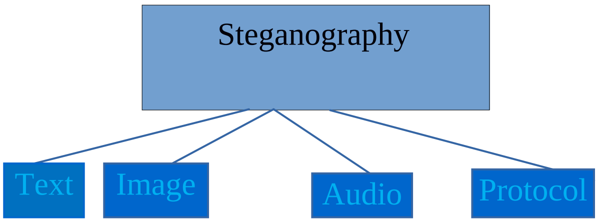

Fig.3 Types of steganography

V STEGANOGRAPHY ALGORITHMS

Following algorithm are implemented and analysed for detection of guilty agent. In this paper one new LSB algorithm is proposed Negate+nth bit. Proposed algorithm provides better security as compared to other two steganography LSB algorithm.

A. *Least significant Bit (LSB-1)*

This is simplest steganography method based in the use of LSB.While embedding LSB of the image pixel is replaced by bit message. It replace the LSB of cover image with MSB of secrete text. E.g. 240 can be hidden in the first eight bytes of three pixels in a 24 bit image.[2]

(00100111 11101001 11001000)

(00100111 11001000 11101001)

(11001000 00100111 11101001)

240 : 011110000

Result: (0010011**0** 11101001 1100101

(0010011**1** 11001001 1110100**0**)

(1100100**0** 00100110 11101000)

B. Least Significant Bit(LSB-2)

In Least significant bit-2 while embedding LSB-2 of the image pixel is replaced by bit message. It replace the LSB-2 of cover image with MSB of secrete text. E.g. 240 can be hidden in the first eight bytes of three pixels in a 24 bit image

(00100111 11101001 11001000)
(00100111 11001000 11101001)
(11001000 00100111 11101001)

240 : 011110000

Result: (001001**0**1 11101011 11001010)

(001001**1**1 11001010 11101001)

(110010**0**0 00100101 111010**0**1)

C .Negate +nth bit

In this paper new steganography algorithm is proposed named Negate + nth bit steganography algorithm, in this algorithm complement of original image pixel is taken and bit message is embedded after every in nth bit in each image using LSB algorithm. And again complement of stegano image is taken after that we get the original image. This proposed method increase the security of transmission message as compared to previous two algorithms.

Above three algorithms are implemented and analyzed for

1) Whole Image

2) Clustered image using K-Means clustering algorithm.

VI CLUSTERING

A process of grouping or partitioning a given set of patterns into disjoint clusters is called clustering. In clustring patterns in same cluster are same and pattern belonging to two different clusters are different. Clustering is example of unsupervised algorithm. Classification is a procedure that assigns data objects to set of classes. In unsupervised algorithm while classifying the data objects clustering process does not depend on predefined classes. In clustering partition of given data set into groups is done based on specified feature so that all points within groups are more similar to each other than points in different groups. Therefore in cluster collection of objects is similar to each other and belonging to different cluster they are dissimilar. In data mining cluster analysis is one of the primary tools. Clustering is used in variety of application domains such as neural network, statics, bioinformatics ,pattern recognition image processing data mining economics.[9]

Several algorithms are proposed for clustering .One of clustering algorithm is k-means clusting algorithm. This algorithm is good in producing results for many practical applications.

VII K-MEANS CLUSTERING ALGORITHM

K-means clustering is a well known partitioning method. In this algorithm belonging to one of K-groups objects are classified. After applying algorithm dataset is divided into K clusters, each object dataset belonging to one cluster. There may be a centroied or cluster representative in each cluster .K-means algorithm uses iterative approach.

A. *Steps in k-means clustering algorithm*

In K-means clustering algorithm given data set is classified into k clusters values of k is defined by user which is fixed. For calculating the distance of data point for m particular centroied euclidean distance is used. Algorithm consists of four steps.

1. Initialization: In this step data set, number of clusters and centroid are defined for each cluster.
2. Classification: For each data point distance is calculated for each data point from the centroid and from centroid data point having minimum distance from centroid of cluster is assigned to that particular cluster.
3 Centroid Recalculation: For previously generated cluster cenroid is again calculated means centroid recalculation.
4. Convergence condition: Some convergence conditions are given as below.

1. Stopping when reaching a given or defined number of iterations.
2 Stopping when there is no exchange of data points between the clusters.
3. Stopping when a threshold value is achieved.
4.If all of the above conditions are not satisfied, then go to step 2 and the whole process repeat again, until the given conditions are not satisfied [10]

In this paper for embedding ,data image is clustered into one or two clusters using k-means clustering algorithm and secrete message is embedded in clusters.

On stegano image we can perform operations like smoothening, filtering, After performing operations there may be possibility that secret data may be removed from stegano image. Suppose that data is given to agent by embedding secret data using above steganography algorithm. After giving data to agent, agent may leak data. Suppose agent has leaked the data then there will be two possibilities image may be tempered or may not temper by third party. In case of tempered image it is difficult to detect guilty agent after performing above mentioned operations. Solution for this is that we can find minimum euclidean distance for each image that is given by administrator.

VIII RESULTS

Experimental results are shown in this section, after applying above mentioned algorithms images are compared using MSE and PSNR values.

1) *MSE: It* computed by byte by byte comparison of cover image and stegano image.

$$MSE=\frac{1}{MxN}\sum_{i=1}^{M}\sum_{j=1}^{N}(p_{ij}-q_{i_j})^2$$

2) *PSNR:* Quality of stegano image is compared with CVR.Higher PSNR quality will be better.

$$PSNR = 10\log_{10}\frac{L^2}{MSE}$$

Following graph shows the comparison of different steganography algorithm on image with respect to MSE and PSNR

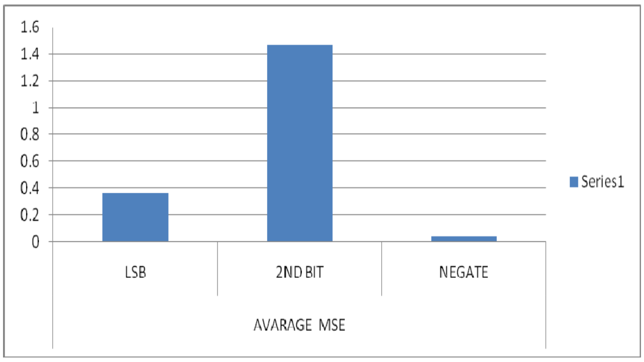

Fig.4 .Bar graph showing comparison of algorithm with respect MSE for whole image

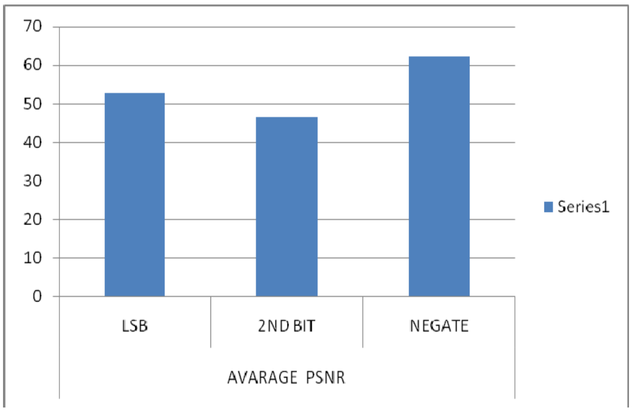

Fig 5.Bar graph showing comparison of algorithm with respect PSNR for whole image

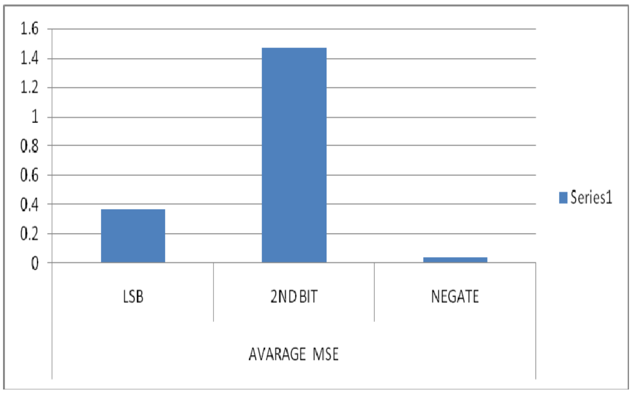

Fig.6 Bar graph showing comparison of algorithm with respect MSE for clustered image

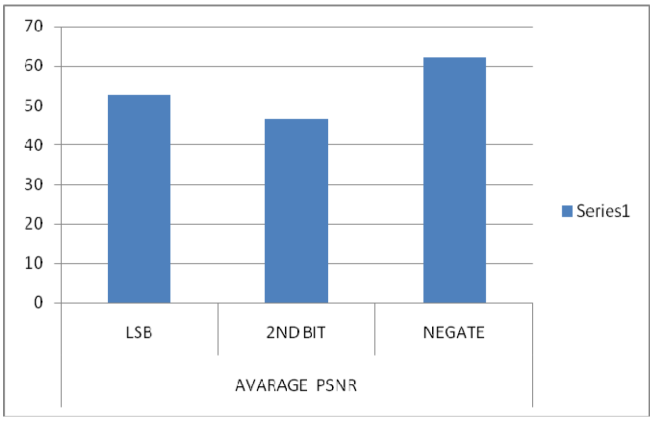

Fig.7 Bar graph showing comparision of algorithm with respectPSNR for clusterd image

CONCLUSION

Steganography is the art of hiding secret message in such a way that no one, apart from the sender and receiver suspect the existence of the message For data leakage detection and to find guilty agent steganography LSB algorithms are used. In this paper one new steganography algorithm is proposed named negate + nth bit. Also Comparison of Least significant (LSB-I), Least significant (LSB-2) and negate+nth bit steganography algorithm is done. Comparison is done with respect to MSE and PSNR values. The PSNR shows the quality of image after hiding the Data. Result of comparison shows that negate+nth bit gives better result as compared to other two algorithms.

REFERENCES

[1] Panagiotis Papadimitriou, Hector Garcia Molina (2010), *Data Leakage Detection* IEEE Transactions on Knowledge and Data Engineering, Vol 22,No 3

[2] A. E.Mustafa A.M.F.ElGamal M.E.ElAlmi Ahmed.BD" *A Proposed Algorithm For Steganography*

[3] Siti Dhalila Mohd Satar, Nazirah Abd Hamid, Fatimah Ghazali, Roslinda Muda and Mustafa Mamat" *A New Model for Hiding Text in an Image Using Logical Connective"* International Journal of Multimedia and Ubiquitous Engineering Vol.10, No.6 (2015), pp.195-202

[4] Xin Liao, Qiao-yan Wen, Jie Zhang, "*A steganographic method for digital images with four-pixel differencing and modified LS substitution*", *Journal of Visual Communication and Image Representation*, Vol. 22, 2011, pp. 1-8

[5] C.K. Chan, L.M. Chen, "*Hiding data in images by simple LSB substitution*", *Pattern Recognition*, Vol. 37, No. 3, 2004, pp. 469–474.

[6] Stuti Goel, Arun Rana, Manpreet Kaur, "*ADCT-based robust methodology for image steganography*", International Journal of Image, Graphics and Signal Processing (IJIGSP).

[7] Karthikeyan, B., Ramakrishnan, S., Vaithiyanathan, V., Sruti, S., Gomathymeenakshi, M.,"*An improved steganographic technique using LSB replacement on a scanned path image*", *International Journal of Network Security*,

[8] Jasleen kour, Deepankar Varma"Steganography Paper –Review Paper" International Journal of Engineering Research in Management and Technology ISSN:2278-9359(Volume 3,Issue-5)

[9] Madhu Yedla, Srinivasa Rao Pathakota, T M Srinivasa" *Enhancing K- means Clustering Algorithm with Improved Initial Center"* Madhu Yedla et al. / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 1 (2) , 2010, 121-125

[10] Manpreet Kaur,Usvir Kaur"Comparision between K-mean and Hirachical Algorithm Using Query Redirection"iNternational Journal of Advanced Research in Computer Science and Software Engineering Volume 3,Issue 7, July 2013,ISSN 2277 128X