An Analytical Approach to an Implementation of Reversible Image Transformation with Steganography for an Online Document Leakage Detection & Prevention

Mrunalinee S. Patil
PG Student
Department of Computer Science & Engineering
Dr. R.G.I.T. & R, Amravati

Prof. Z. I. Khan
Assistant Professor
Department of Computer Science & Engineering
Dr. R.G.I.T. & R, Amravati

Abstract

Steganography are the most popular or widely used information security scheme or techniques. It is the act of covert communications. The algorithms for hiding data or message in a cover image, and extracting a message from a stego-image were implemented using C# programming language. In this paper, LSB Replacement technique was adopted as the embedding method. In this paper we propose a reversible high capacity data hiding method applying on binary images. We also proposed leakage detection and prevention with the help of unique hash values of the document/images. The paper proposes a novel framework for reversible data hiding in encrypted image (RDH-EI) based on reversible image transformation (RIT). Different from previous frameworks which encrypt a plaintext image into a cipher text form; RIT-based RDH-EI shifts the semantic of original image to the semantic of another image and thus protects the privacy of the original image. In our proposed system, we proposed a new mosaic image technique in which we divide the image/pdf document into parts, and the parts will then transformed, encrypted and encoded into target images. Steganography meaning covered writing. It includes the concealment of information within computer files. It is one of the techniques used for securing data or information over the internet.

Keywords: Steganography, Least Significant Bit (LSB), Reversible image data hiding (RIDH), Stego Image, BMP Image

---

I. INTRODUCTION

A. *Steganography: -*

Steganography and Cryptography, both are used for security purposes but with different implementation and approaches. Steganography is defined as a technique to hide data into images in such a manner, which is unperceivable. In cryptography, the text file get converted to other form which provide confidentiality to sensitive data but in steganography we hide the actual data file in image form so that if leakage get occurred the third party fails to recognize the actual data. This provides confidentiality as well as security to the sensitive data. The idea is to hide text in image with the conditions that the image quality is retained along with the size of the image instead we can encrypt the data. While in steganography hiding message in an image, along with the conditions, it make seem of just an exchange of picture between two user ends.

The steps being followed in steganography are as under:-

1) Firstly the text message is being written, then encryption of the message is done.
2) Later, text is hidden in the selected media like image file and transmitted at the receiver side.
3) At receiver end, reverse method is done to implement and recover the original text message.

1) *Classification of Steganography:*

For decades people try to develop innovative methods for secret communication. Classification of information hiding can be depicted as follows:

Steganography is mainly of two types, linguistic steganography and technical steganography as shown in figure no.1. In linguistic steganography, machine readable data is encoded to innocuous natural language text, thereby providing security against any arbitrator tolerating natural language as a communication medium.[13] In this approach, linguistic properties of a text are customized to hide information. Language has the property that a small local change to a text, e.g. replacing a word by a word with same context, may result in text which is anomalous at the document level, or with respect to the state of the world. Hence finding linguistic transformations which can be applied reliably and often is a challenging problem for Linguistic steganography.[14]

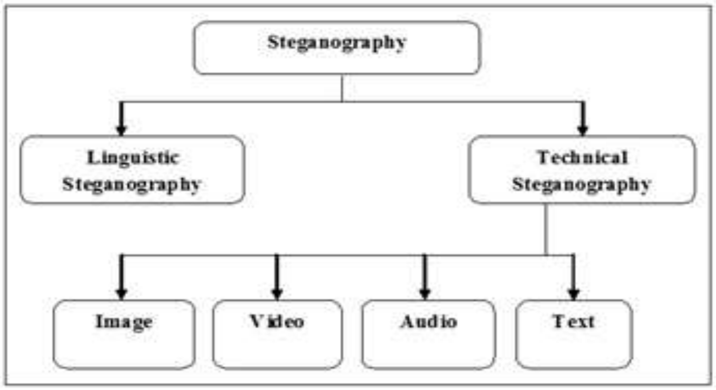


Fig. 1: Classification of Steganography

Technical Steganography: It is a carrier rather than a text which can be presented, as any other substantial medium such as microdots and invisible inks. In this context, the cover_media is the file in which we will hide the secret_data, which may also be encrypted using the stegokey. The resultant file is the stego_media. There are four ways to implement steganography:

1) Using text.
2) Using images.
3) Using audio files.
4) Using video files

2) Process and Techniques of Steganography:-

Various techniques are used in the field of steganography by arranging the different bits of the character of the text message in the image file and other media. In order to encrypt the data two files are needed: (i) image file and (ii) the text file containing the data. Our algorithm is simple and flexible using LSB (Least Significant Bit) technique. Least significant bits (LSB) insertion is a simple approach to embedding information in image file. The technique we are using is LSB i.e. storing in LSB of a byte (pixel).

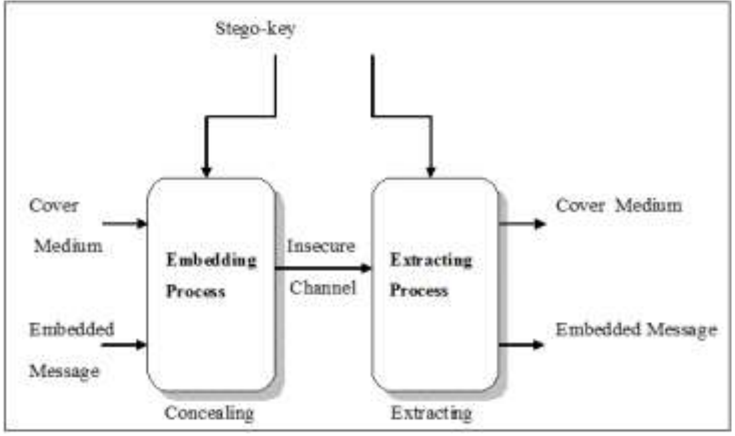


Fig. 2: Steganographic Process

Some of the techniques used in steganography are domain tools or simple system such as least significant bit (LSB) insertion. In our security model, we divide the document / scanned images of document into parts. The parts will be reversibly transformed, encrypted and then encoded into target images which enhances the security of document and reduces possible attacks. At the time of file transfer we will check the access permission for the user as well as hash value of that document. If the document is an official document and particular employee is trying to leak that particular document, system will automatically prevent it and send notifications to higher authority for further actions to be taken. In order to protect the data files being open source the data leakage must be detected in the early stage.

Embedding data, which is to be hidden into an image, requires two files. The first is the image that will hold the hidden information, called the cover image. The second file is the message- the information to be hidden. When combined the cover image and the embedded message make a stegoimage or stego-file as shown in figure 2.

Steganography system is designed for encoding and decoding a secret file embedded into an image file using random LSB insertion method in which the secret data is spread out among the image data in a seemingly random manner. This could be achieved using a secret key.

3) Applications of Steganography

1) Confidential communication and secret data storing
2) Access control system for digital content distribution

3) Digital watermarks
4) Modern printers

There are many applications for digital steganography of images, including copyright protection, feature tagging, and secret communications.

B. Data Leakage Detection (DLD) and prevention:-

Data leakage prevention is a technique used to hide the confidentiality of data being accessed by unauthorized user. Most DLP solutions include a suite of technologies that facilitates three key objectives:

- Locate and catalog sensitive information stored throughout the enterprise
- Monitor and control the movement of sensitive information across enterprise networks
- Monitor and control the movement of sensitive information on end-user systems [19].

C. Reversible Image Transformation (RIT) with steganography

1) At the time of document uploading, user have to choose security level of that document.
2) Security level can be 1 to 4.
3) System will divide the uploaded document/image into specified no of parts (1 to 4).
4) The parts then transformed in reverse way i.e. the text/bytes will be arrange in reverse fashion.
5) The transformed parts will be encrypted using any advanced encryption algorithm.
6) The encrypted part of document/image will be then encoded into target image using LSB algorithm according to number of parts.
7) Required encryption keys (as per the number of parts) will be stored into the database in encrypted format.

D. Reversible Image Data Hiding (RIDH)

It is a special category of data hiding technique, which ensures perfect reconstruction of the cover image upon the extraction of the embedded message. The reversibility makes such an image data hiding approach particularly attractive in the critical scenarios. The majority of the existing RIDH algorithms are designed over the plaintext domain, namely, the message bits are embedded into the original unencrypted images. The early works mainly utilized the lossless compression algorithm to compress certain image features, to vacate room for message embedding.

In this paper, we propose an encrypted-domain RIDH scheme by specifically taking the above-mentioned design preferences into consideration. The proposed technique embeds message through a public key modulation mechanism and performs data extraction by exploiting the statistical distinguish ability of encrypted and no encrypted image blocks.

II. PROPOSED WORK

The proposed method uses XOR Ciphering technique to find the maximum PSNR value and hiding capacity with less MSE shown in Fig 3. The proposed block diagram is described as follows

A. Original Image

The original image is any input image selected to hide the secret data.

B. XOR Ciphering

XOR Cipher also called as an exclusive OR and a type of additive cipher, an encryption algorithm Operates according to principles

$$A \oplus 0 = A$$
$$A \oplus A = 0$$
$$0 \oplus 0 = 0$$
$$0 \oplus A = A$$

C. Secret data

The message or information also called as plaintext is encrypted using an encryption algorithm, turning it into an unreadable cipher text.

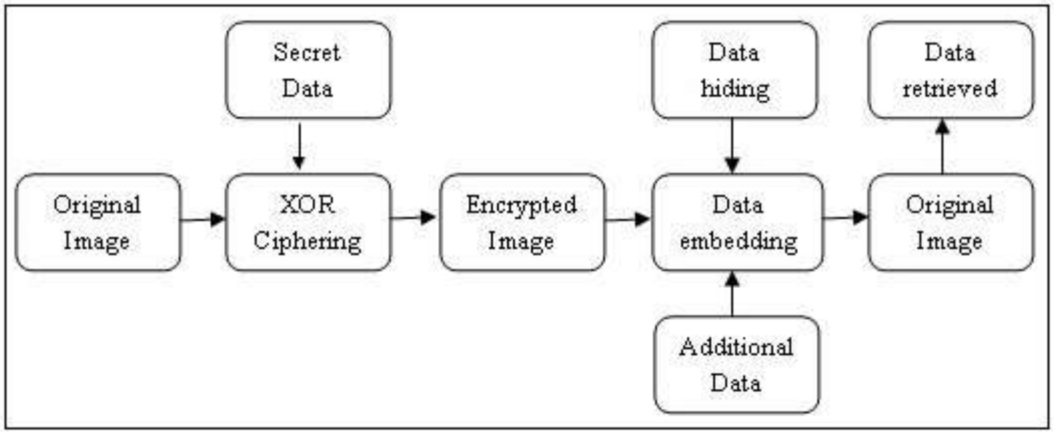


Fig. 3: Proposed Block Diagram

D. Encrypted image

After rearranged self-embedded image, denoted by X, is generated, we can encrypt X to construct the encrypted image, denoted by E. With a stream cipher, the encryption version of X is easily obtained.

E. Data hiding key

Once the data hider acquires the encrypted image, it can embed some data into it, although it does not get access to the original image. The embedding process starts with locating the encrypted version.

F. Data embedding

Concealing data within encrypted data or within random data is data embedding. Here once the data embedding process is done the PSNR value and MSE is obtained. Peak signal-to-noise ratio is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. PSNR is usually expressed in terms of the logarithmic decibel scale. PSNR is most easily defined through the mean squared error (MSE). Given a noise-free m×n monochrome image I and its noisy approximation K, MSE is defined as:

$$MSE = \frac{1}{m\,n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2$$

The PSNR is defined as:

$$PSNR = 10\,.\log_{10}\left(\frac{MAX_I^2}{MSE}\right)$$

$$= 20\,.\log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right)$$

$$= 20\,.\log_{10}(MAX_I) - 10\,.\log_{10}(MSE)$$

Here, MAX_I is the maximum possible pixel value of the image. When the pixels are represented using 8 bits per sample, this is 255. More generally, when samples are represented with B bits per sample, MAX_I is 2^B-1. For color images with three RGB values per pixel, definition of PSNR is the same except the MSE is the sum over all squared value differences divided by image size and by three. Alternately, for color images the image is converted to a different color space and PSNR is reported against each channel of that color space.

III. METHODOLOGY

In our proposed system, we proposed a new mosaic image technique in which we divide the image/pdf document into parts, and the parts will then transformed, encrypted and encoded into target images. We proposed enterprise social network application in which organizational employees will communicate and share documents with each other using our security model. In our security model, we divide the document / scanned images of document into parts. The parts will be reversibly transformed, encrypted and then encoded into target images which enhances the security of document and reduces possible attacks.

At the time of document uploading as shown in figure 4, users have to choose security level of that document. Security level can be 1 to 4. System will divide the uploaded document/image into specified no of parts (1 to 4). The parts then transformed in reverse way i.e. the text/bytes will be arrange in reverse fashion. The transformed parts will be encrypted using any advanced encryption algorithm. The encrypted part of document/image will be then encoded into target image using LSB algorithm according to number of parts. Required encryption keys (as per the number of parts) will be stored into the database in encrypted format.

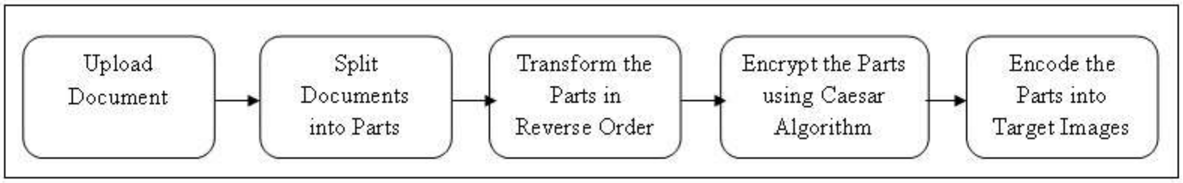

Upload Document
Split Documents into Parts
Transform the Parts in Reverse Order
Encrypt the Parts using Caesar Algorithm
Encode the Parts into Target Images

Fig. 4: Working of Encryption and Encoding

We also proposed leakage detection and prevention with the help of unique hash values of the document/images. At the time of file transfer we will check the access permission for the user as well as hash value of that document. If the document is an official document and particular employee is trying to leak that particular document, system will automatically prevent it and send notifications to higher authority for further actions to be taken.

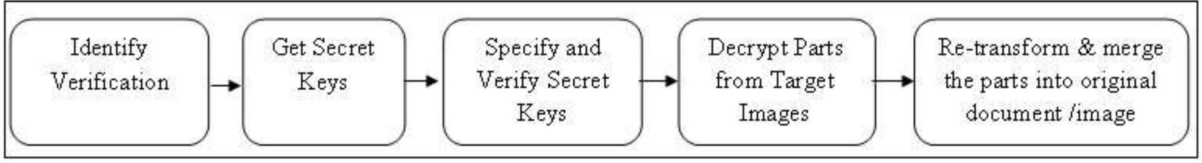
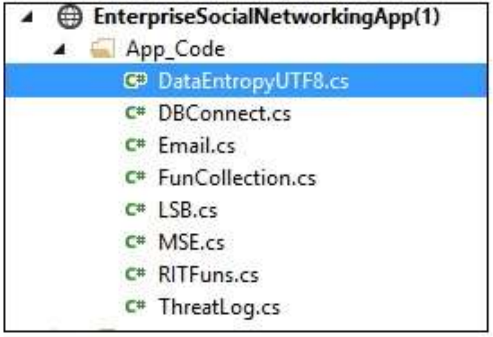

Identify Verification
Get Secret Keys
Specify and Verify Secret Keys
Decrypt Parts from Target Images
Re-transform & merge the parts into original document /image

Fig. 5: Working of Decryption and Decoding

This paper proposes a novel framework for reversible data hiding in encrypted image (RDH-EI) based on reversible image transformation (RIT). Different from previous frameworks which encrypt a plaintext image into a cipher text form; RIT-based RDH-EI shifts the semantic of original image to the semantic of another image and thus protects the privacy of the original image.

IV. Implementation

The given idea is implemented as a desktop application by using asp.net and C# module to check the nature and time complexity of the application. The application is made with the help of some important files as shown in figure 6 like,



EnterpriseSocialNetworkingApp(1)
App_Code
DataEntropyUTF8.cs
DBConnect.cs
Email.cs
FunCollection.cs
LSB.cs
MSE.cs
RITFuns.cs
ThreatLog.cs

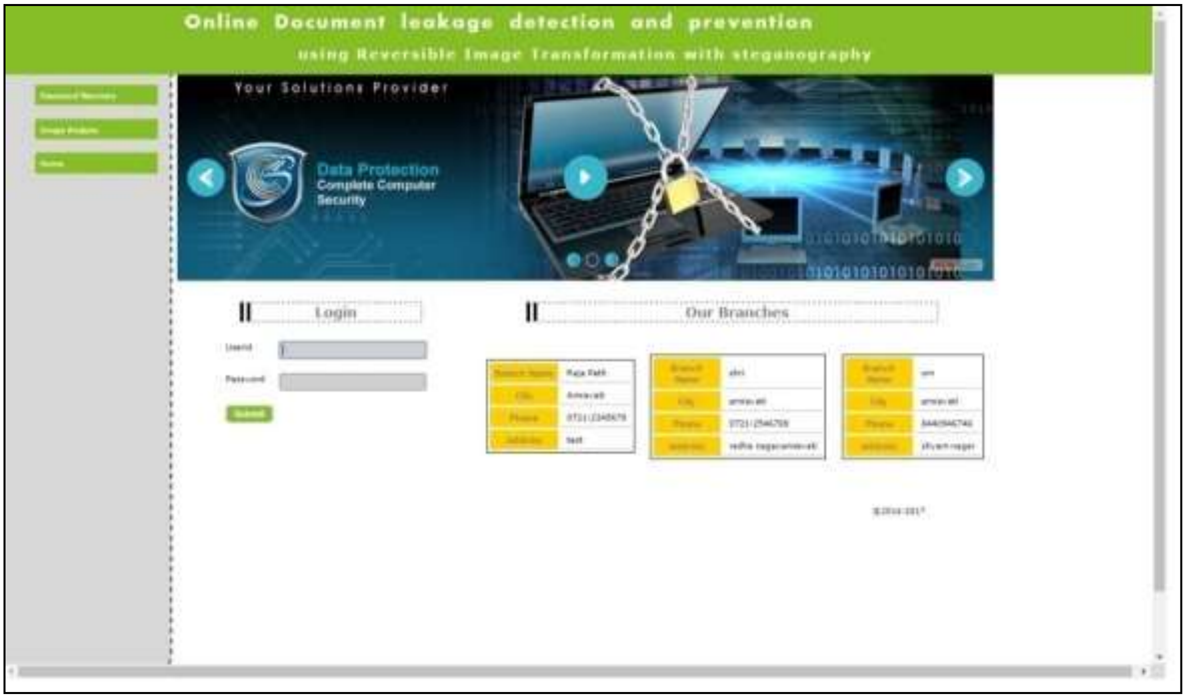
Fig. 6: Classification of different codes

1) Data EntropyUTF8.cs- The tool can calculate entropy for bits, bytes, or Unicode code points.
2) DBConnect.cs - The first step in a C# application is to create an instance of the Server object and to establish its connection to an instance of Microsoft SQL Server.
3) Email.cs- Represents an e-mail message that can be sent using the SmtpClient class.
4) FunCollection.cs- Collection classes are specialized classes for data storage and retrieval. These classes provide support for stacks, queues, lists, and hash tables
5) LSB.cs- The least significant bit (LSB) is the bit that when flipped from 0 to 1 or from 1 to 0, then no significant change will occur on the total value.
6) MSE.cs- calculate the "Mean Square Error" of the RGB values of corresponding pixels.
7) RITFuns.cs- Represents Reversible Image Transformation functions in a different class.
8) ThreatLog.cs- To detect the threats in a class as security point of view.

V. Flow of Application and Result

This section presents the steps of the application runs through which is as shown in below. The flows of application in panel show how the application allows secret communication using image steganography.

A. Home Page having Login Option, Login process collects user id and password


Online Document leakage detection and prevention
using Reversible Image Transformation with steganography
Password Recovery
Image Products
Home
Your Solutions Provider
Data Protection
Complete Computer Security
Login
Our Branches
Submit

Userid

Password

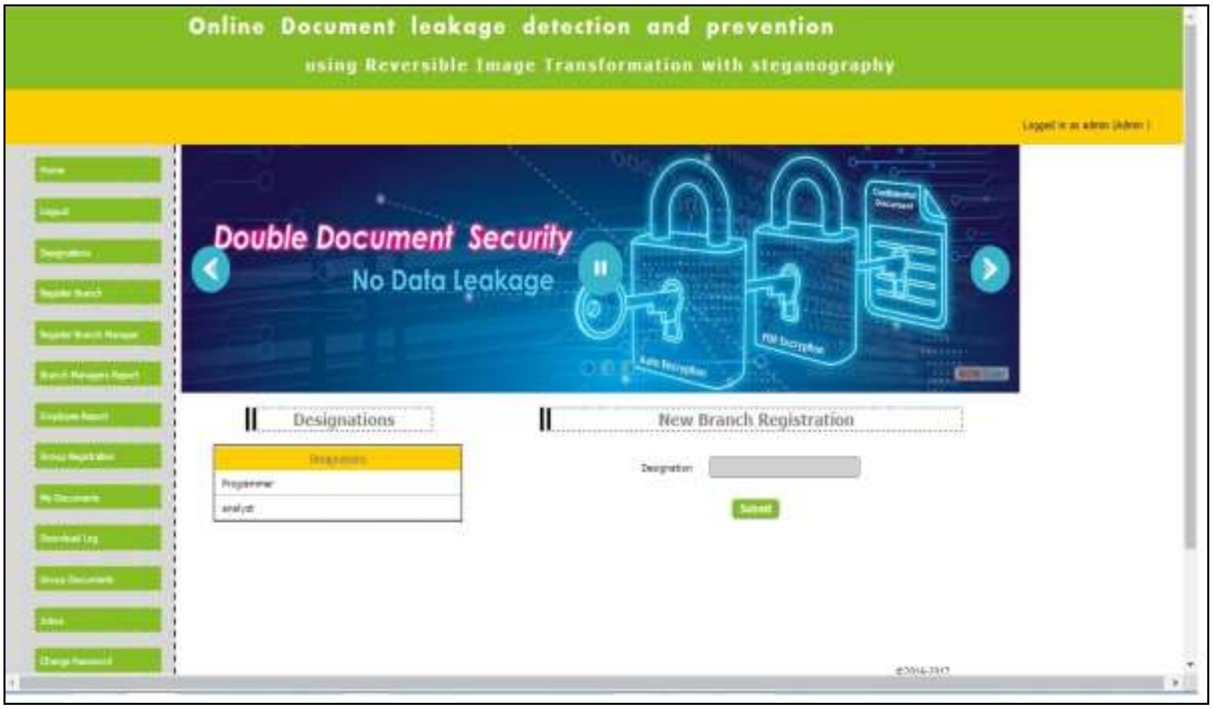
Branch Name	Raja Peth
City	Amravati
Phone	07212345678
Address	test

Branch Name	abc
City	amravati
Phone	0721-2546759
Address	radha nagaramravati

Branch Name	om
City	amravati
Phone	8446946740
Address	shyam nagar

©2014-2017

B. 1st & 2nd panel containing New Branch registration which collects information of Designation, Branch name, Address and phone number of New Branch.

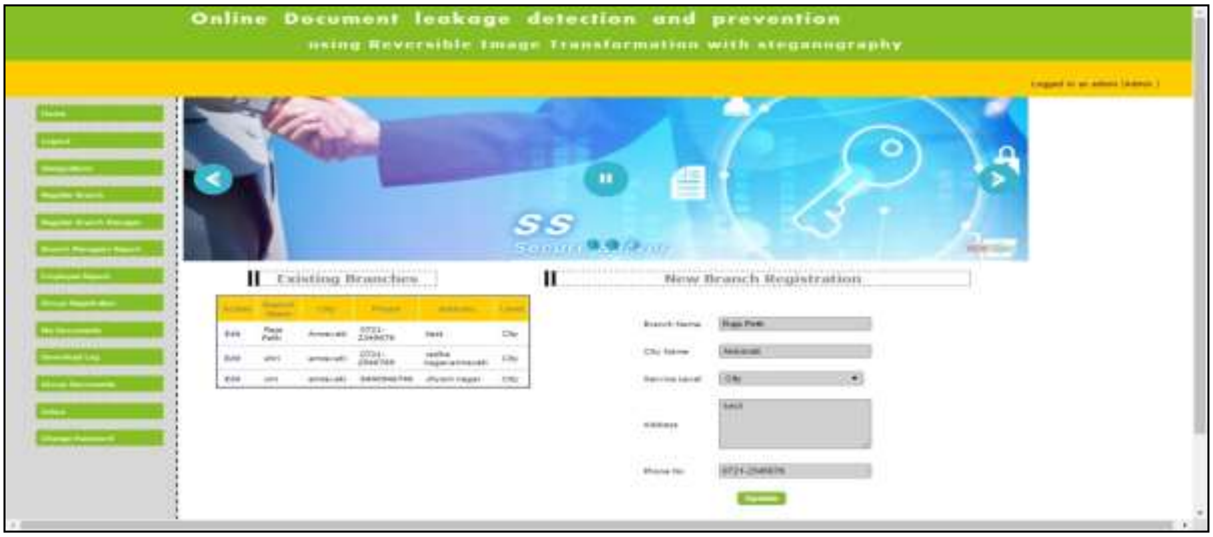


Logged in as admin (Admin)

Designation
Programmer
analyst

Designation

©2014-2017



Action	Branch Name	City	Phone	Address	Level
Edit	Raja Peth	Amravati	0721-2349676	test	City
Edit	shri	amravati	0721-2566769	radha nagar,amravati	City
Edit	shri	amravati	[illegible]	shyam nagar	City

Branch Name: Raja Peth

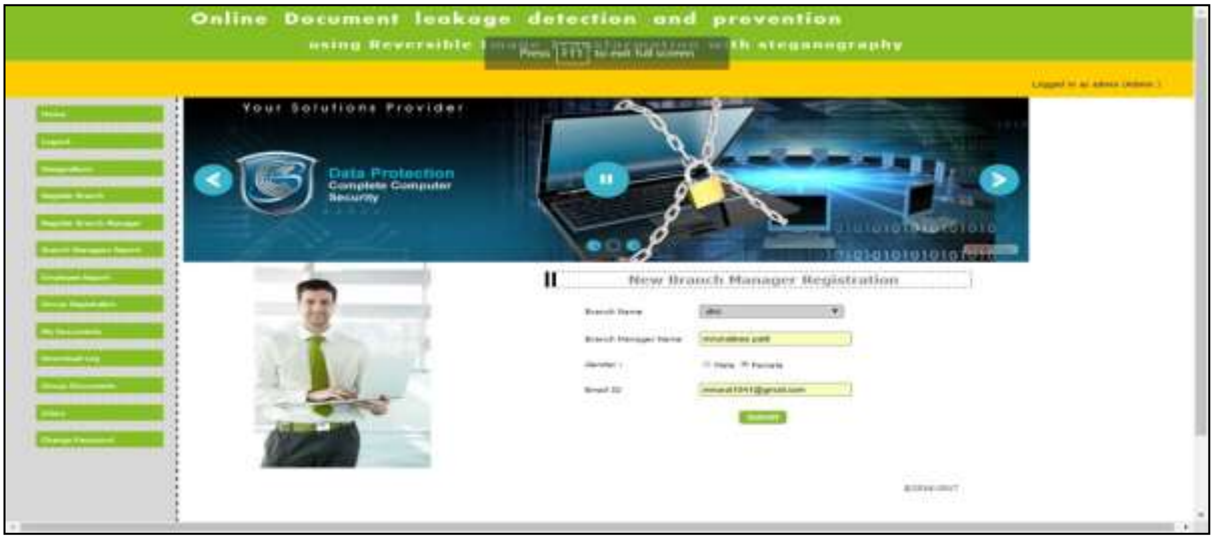
City Name: Amravati

Service Level: City

Address: test

Phone No: 0721-2549676

C. 3rd panel containing New Branch Manager registration collects the data of Branch Name, Branch Manager Name and email id


Online Document leakage detection and prevention
using Reversible Image Transformation with steganography
Press [F11] to exit full screen
Logged in as admin (Admin)
Home
Logout
Designation
Register Branch
Register Branch Manager
Branch Manager Report
Employee Report
Group Registration
My Documents
Download Log
Group Documents
Inbox
Change Password
Your Solutions Provider
Data Protection
Complete Computer Security
New Branch Manager Registration
Submit

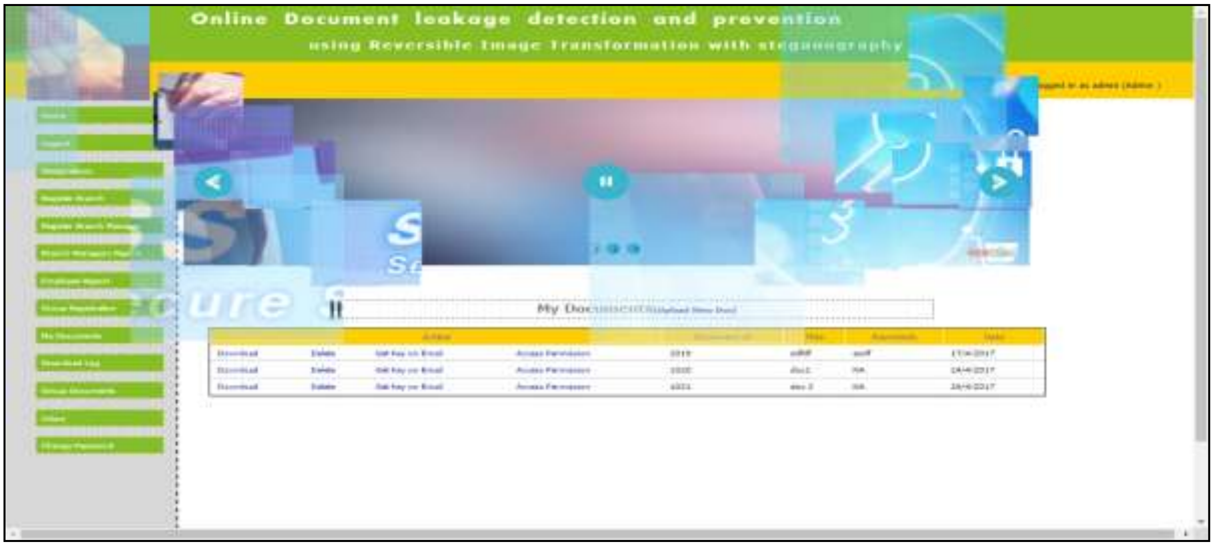
Branch Name: shri

Branch Manager Name: [illegible]

Gender : ○ Male ● Female

Email ID: [illegible]@gmail.com

D. 4th panel containing My Documents in which we have to upload a new document, we can download it (need secret key), Get secret key on registered email.



Action				Document ID	Title	Keywords	Date
Download	Delete	Get key on Email	Access Permission	1019	[illegible]	[illegible]	17/4/2017
Download	Delete	Get key on Email	Access Permission	1020	doc1	NA	24/4/2017
Download	Delete	Get key on Email	Access Permission	1021	doc 2	NA	28/4/2017

E. 5th panel containing Image Analysis which collects the data of Original Image and Stego Image/ Encoded Image which gives the information of values of PSNR and MSE.

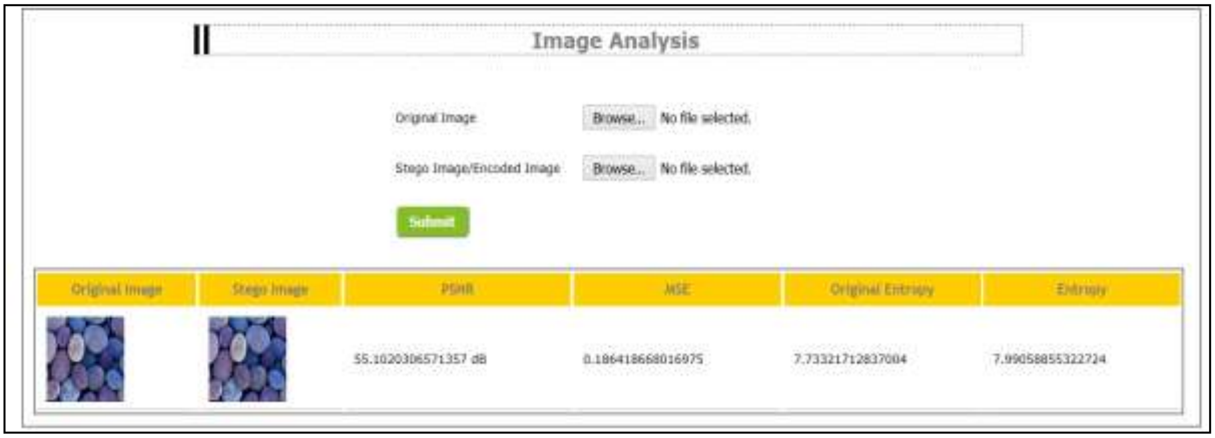


Image Analysis

Original Image Browse... No file selected.

Stego Image/Encoded Image Browse... No file selected.

Submit

Original Image	Stego Image	PSNR	MSE	Original Entropy	Entropy
		55.1020306571357 dB	0.186418668016975	7.73321712837004	7.99058855322724

VI. CONCLUSION

After performing experiments we can conclude that the value of PSNR is maximum and obtain the less MSE.

We are getting value of PSNR is 55.10 and value of MSE is 0.18. Also we came to conclude that an embedding capacity of proposed scheme is much higher. In this paper an experimental implementation of image steganography in securing data or information over a communication channel was carried out by using various types of input images as cover images. At the time of file transfer we will check the access permission for the user as well as hash value of that document. Cloud computing can be implemented. Ceaser Algorithm can be used for data hiding along with reversible data hiding algorithm. Image data hiding can be applied for Videos, audio type of data Protect the data files being open source the data leakage must be detected in the early stage.

REFERENCES

[1] M. U. Celik, G. Sharma, A. M. Tekalp, and E. Saber, “Lossless generalized-LSB data embedding,” IEEE Trans. Image Process., vol. 14, no. 2, pp. 253–266, Feb. 2005.

[2] M. U. Celik, G. Sharma, and A. M. Tekalp, “Lossless watermarking for image authentication: A new framework and an implementation,” IEEE Trans. Image Process., vol. 15, no. 4, pp. 1042–1049, Apr. 2006.

[3] X. Zhang, “Separable reversible data hiding in encrypted image,” IEEE Trans. Inf Forensics Security, vol. 7, no. 2, pp. 826–832, Apr. 2012.

[4] W. Puech, M. Chaumont, and O. Strauss, “A reversible data hiding method for encrypted images,” Proc. SPIE, vol. 6819, pp. 1–9, Feb. 2008.

[5] X. Zhang, “Reversible data hiding in encrypted image,” IEEE Signal Process. Lett., vol. 18, no. 4, pp. 255–258, Apr. 2011.

[6] W. Hong, T.-S. Chen, and H.-Y. Wu, “An improved reversible data hiding in encrypted images using side match,” IEEE Signal Process. Lett., vol. 19, no. 4, pp. 199–202, Apr. 2012.

[7] Provos, N., Honeyman, P, Hide and seek: An introduction to steganography, IEEE Security & Privacy Magazine 1 (2003) pp. 32-44.

[8] Silman, J., Steganography and Steganalysis: An Overview, SANS Institute 2001.

[9] N.F. Johnson, S. Jajodia, "Exploring Steganography: Seeing the Unseen", IEEE computer, Vol. 31, No. 2, pages 26-34, February 1998.

[10] P. Hayati, V. Potdar, E. Chang, A survey of steganographic and steganalytic tools for the Digital forensic investigator,available from: http://debii.curtin.edu.au/~pedram/images/docs/survey_of_steganograph y_and_steganalytic_tools.pdf

[11] A.Joseph Raphael, Dr.V Sundaram, “Cryptography and Steganography – A Survey”, Int. J. Comp. Tech. Appl., Vol 2 (3), 626-630 ISSN:2229-6093.

[12] Gandharba Swain, Saroj Kumar Lenka,“A Hybrid Approach to Steganography Embedding at Darkest and Brightest Pixels”, Proceedings of the International Conference on Communication and Computational Intelligence – 2010, Kongu Engineering College, Perundurai, Erode, T.N.,India.27– 29 December,2010.pp.529-534.

[13] Richard Bergmair ,”Towards Linguistic Steganography: A Systematic Investigation of Approaches, Systems, and Issues Oct-03 – Apr-04.

[14] Richard Bergmair. 2007 A comprehensive bibliography of linguistic steganography. In Proceedings of the SPIE Conference of security. Steganography and watermarking of multimedia contents, 6505.

[15] Chamkour Singh, Gauravdeep, “Cluster based Image Steganography using Pattern Matching”, IJAIR, vol. 2, issue 5, 2013.

[16] Neil, F. and Jajodia, J. S. Exploring Steganography: Seeing the Unseen. George Mason University USA. 0018-9162/98/$10.00 © 1998 IEEE

[17] Neeta, D., Snehal, K. and Jacobs, D. Implementation of LSB Steganography and Its Evaluation forVarious Bits. I EEE Digital Information Management, 2006 1st International Conference on, Bangalore, Pp173 –178, 6-6 Dec. 2006 .

[18] Mamta Jain and Saroj Kumar Lenka “A Review on Data Leakage Prevention using Image Steganography”, International Journal of Computer Science Engineering (IJCSE), ISSN : 2319-7323 Vol. 5 No.02 Mar 2016 (56-59)

[19] Raman, Preeti, H. G. Kayacık and A. Somayaji. "Understanding Data Leak Prevention."6th Annual Symposium on Information Assurance (ASIA’11).2011.