Hide Text in Images Using Steganography and a Review of Methods and Approach for Secure Stegnography

Navdeep[1],
PG Student Department of Computer Science,
Shri Ram college of Engineering & Management,
NH-2, Delhi-Mathura Road, Palwal, Haryana, India

Ms Neha Goyal[2]
Assistant Professor,
Department of Computer Science,
Shri Ram college of Engineering & Management,
NH-2, Delhi-Mathura Road, Palwal, Haryana, India

Abstract – ***In this paper, we have proposed a new steganographic algorithm that is used to hide text file inside an image. In order to increase/ maximize the storage capacity we have used a compression algorithm that compresses the data to be embedded. The compression algorithm we have used works in a range of 1bit to 8 bits per pixel ratio. By applying this algorithm we have developed an application that would help users to efficiently hide the data. In this paper we also survey different steganography techniques for encrypting the data. Steganography is a technique that allows the one to hide the data within an image while adding few noticeable changes. This paper discusses the concept behind the steganography by exploring firstly what is the steganography and the terms that are related to steganography. This paper explores the steganography methods –image steganography, audio steganography, video steganography, text steganography that are used to embed the information in digital carriers. The two most important aspects of image based stegnography system are the quality of stego image and the capacity of the cover image. By reviewing this paper, researchers can develop a better steganography technique to increase the MHC and PSNR value by analyzing the existing steganalysis techniques.***

1. Introduction

Steganography is derived from Greek words Steganous meaning "covered" and graphy meaning "writing".So it is known as "covered writing".Steganography is a technique which is used to hide the message and prevent the detection of hidden message. Image stegnography [1] is a modern way of hiding information in a way that the unwanted people may not access the information. Data used to hide data in stegnography can be text or image. In modern times, image stegnography can be helpful in a number of ways such as hiding the secret data [2], data authentication, ensuring authenticated data availability for academic usage, monitoring of data piracy, labelling electronic data/contents, copyright protection, ownership identification, providing confidentiality and integrity enhancement control of electronic data piracy etc.[3].

The primary objective of steganography is to avoid drawing attention to the transmission of hidden information. If suspicion is raised, then this objective that has been planned to achieve the security of the secret message because if the hackers noted any change in the sent message then this observer will try to know the hidden information inside the message. **[1][2].**

The basic terminologies used in the steganography systems are: the cover message, secret message, the secret key and embedding algorithm **[5].** The cover message is the carrier of the message such as image, video, audio, text or some other digital media. The secret message is the information which is needed to be hidden in the suitable digital media. The secret key is usually used to embed the message depending on the hiding algorithms. The embedding algorithm is the way or the idea that usually used to embed the secret information in the cover message **[8][9].**

In steganography, before the hiding process, the sender must select an appropriate message carrier, an effective message to be hidden as well as a secret key used as a password. A robust steganographic algorithm must be selected that should be able to encrypt the message more effectively. The sender then may send the hidden message to the receiver by using any of the modern communication techniques. The receiver after receiving the message decrypts the hidden message using the extraction algorithm and a secret key **[8][9].**

This paper proposes a new algorithm to hide data inside an image using steganographic technique. The algorithm that we have proposed is an enhanced version of LSB technique, that is not very much robust. Also we have implemented a compression technique to increase the hiding capacity. This all is demonstrated using an application we have build in java.

The rest of the paper is organized as follows: Section 2 would be presenting the proposed algorithm. Section 3 reviews the Steganographic Terms. Section 4 provides a state-of-art review and analysis of different existing methods of Steganography drawn from literature survey. Steganography Applications are presented in Section 5. Section 6 review the different types of Steganography. Finally, the Proposed method and conclusion is presented in Section 7.

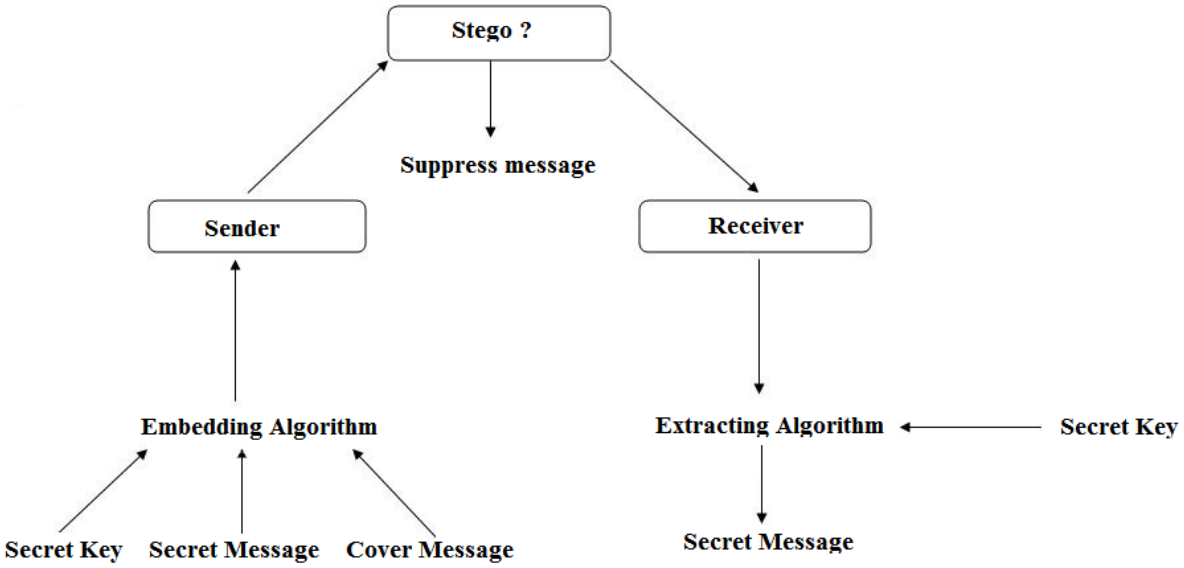
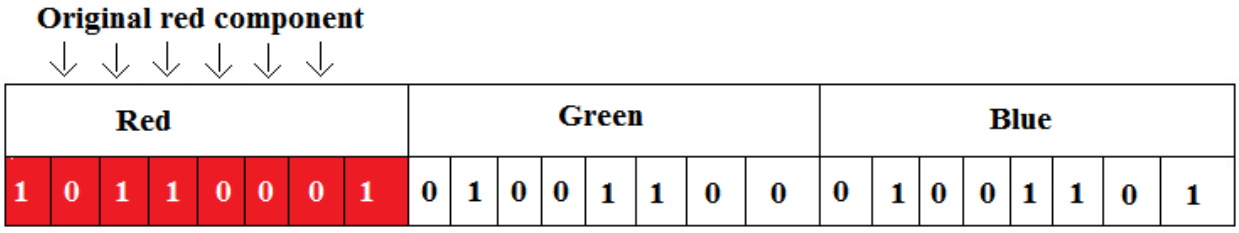


Figure 1: General Steganographic Approach.

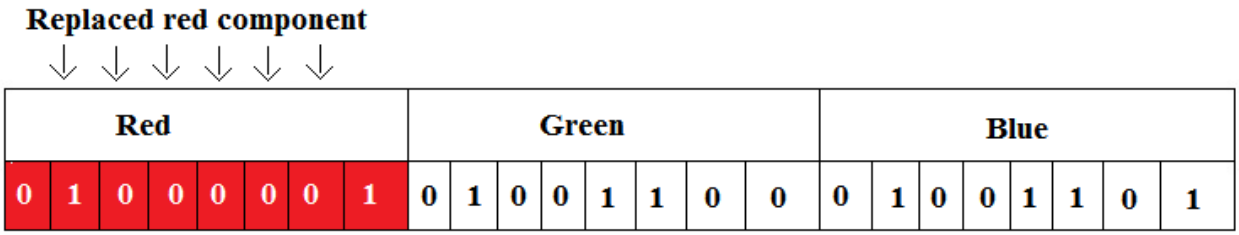
II. Proposed Algorithm

The algorithm that we have proposed in this system is basically an extension of the original LSB which is quite vulnerable. Instead of hiding the data in least significant bits of the RGB components of a pixel, we in this algorithm, would be hiding data as shown below: - Let the data to be hidden is word "ABC" ASCII code of A= 65 and corresponding binary is 01000001. ASCII code of B= 66 and corresponding binary is 01000010. ASCII code of C= 67 and corresponding binary is 01000011.

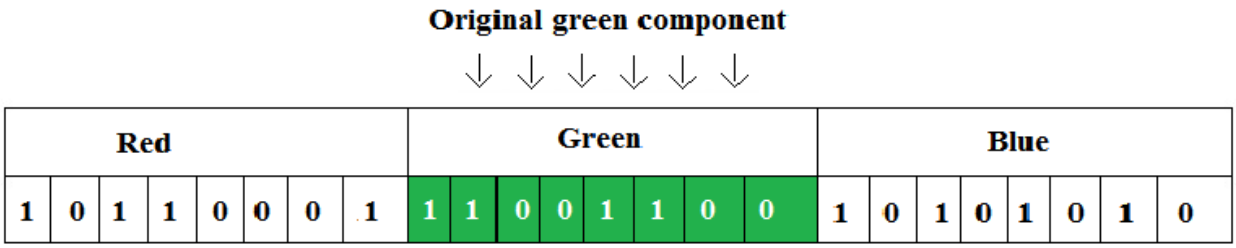
Let the first pixel‟s RGB component be: -



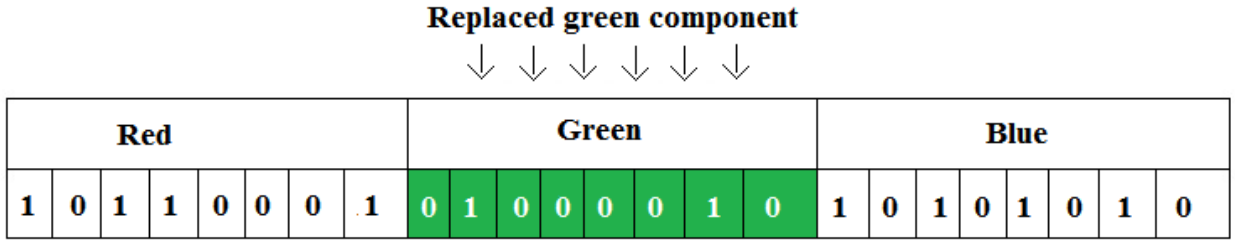
Red component is replaced with binary of 65 i.e. A.



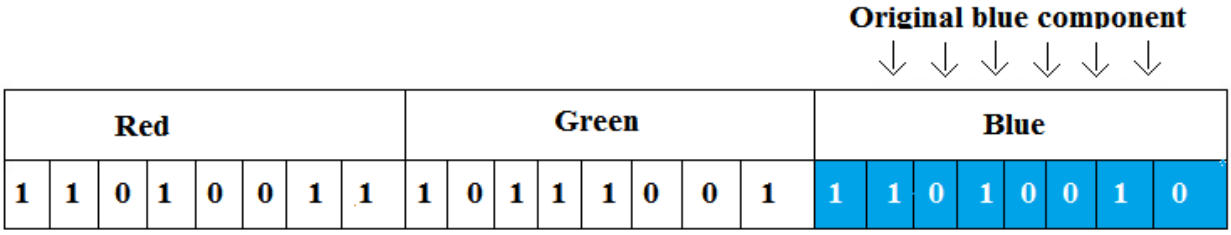
Let the second pixel‟s RGB component be: -



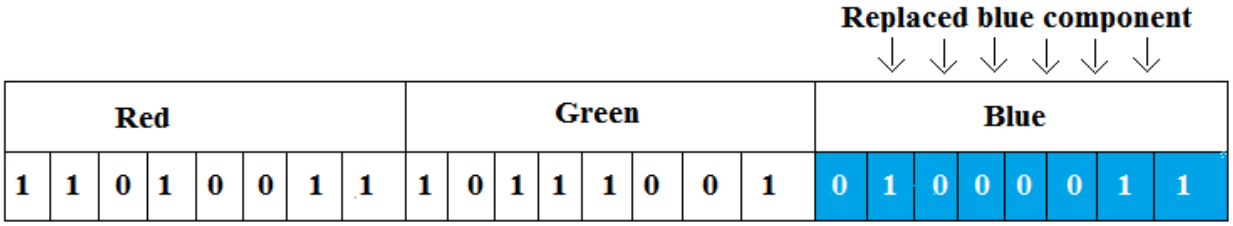
Green component of second pixel is replaced with binary of 66 i.e. B.



Let the third pixel‟s RGB component be: -



Blue component of third pixel is replaced with binary of 67 i.e. C.



And the process continues. The resulting stego image that we are obtaining after the algorithm completes its execution, is distorted and is easy to detect, that some kind of alteration has been done to the image. So, to enhance the security of the secret message we would be covering the resulting stego image with a new cover image, this is the first level of security. By just looking at the resulting image no one would be able to predict that something is hidden inside it. The new cover image can be the same or different than the original. In order to increase the storage capacity of the image, a compression algorithm has been used, we know that each component of an RGB pixel is represented with 8 bits. So, the maximum compression would be 8 bits per pixel and minimum would be 1 bit per pixel. The proposed steganographic algorithm comprises of two embedding techniques they are data hiding technique and data retrieving technique. Data hiding technique as the name suggests is used to hide secret message and key in the cover image, while data retrieving technique is used to retrieve the key and the hidden secret message from the stego image. Therefore data is protected in image without revealing to unauthorized party.

A. ***Proposed embedding technique.*** Inputs: - Text file, cover image 1, cover image 2 and secret key. Output: - Stego image.

Begin

1. Select a text file, convert it into binary form and calculate the number of bits in it.
2. Select a carrier image (cover image 1) for hiding purpose, find the number of pixels, convert it into RGB image and calls the compression function.
3. If bits calculated are compatible with the image resolution, then

Start sub iteration

1 Replace red component of the first pixel with first character. Replace green component of the second pixel with second character. Replace blue component of the third pixel with third character. And repeat iterations until pixels get exhaust. **Stop sub iteration 1** Else **Repeat sub iteration 1** Finds necessary compression ratio and perform sub iteration

2. **Sub iteration 2** Replace necessary bits as defined by the compression ratio in immediate component of each pixel. Store the information about bits embedded in a binary address file.

Stop sub iteration2
4. Provide a security key as encryption completes.
5. Select 2nd cover image to hide the distorted stego image.

End

B. Proposed extraction technique. Input**: -** Stego image and secret key. Output: - Secret text file.

Begin

1. Browse the stego image.

2. Choose the folder in which you want to extract the hidden text file.

3. Provide necessary security key.

4. Convert the binary file into human readable form.

End

The main focuses of this proposed steganographic technique is to hide text files in images, compresses the text files so as to increase the overall storage capacity, applying a secret key on the resulting stego image and transferring the secret message without any vulnerability and threat.

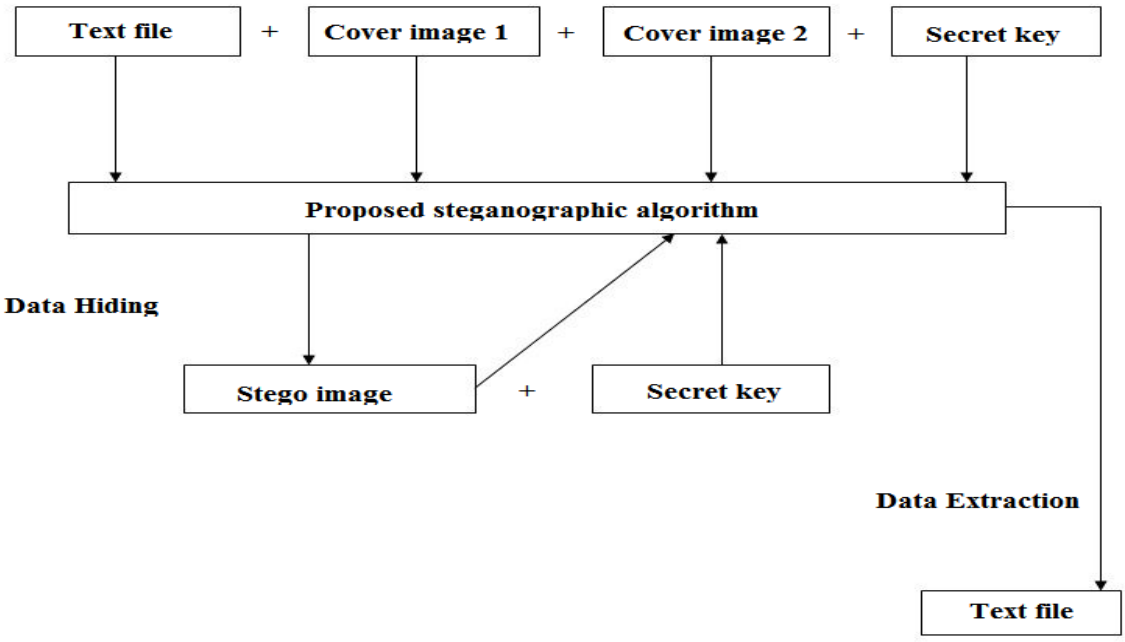


Figure 2: General Layout of Proposed System.

III. Steganographic Terms

- Cover File:It is a file in which hidden information will be stored.
- Stego Medium: Medium through which the information is hidden.
- Message: The data to be hidden or extracted.
- Steganalysis: Identify the existence of message.

IV. Related work:

In the related work, the most common method which is used to hide the message involve the usage of LSB developed by Chandramouli et al.[1],by applying the filtering, masking and transformation on the cover media. Weiqi Luo el al.,[2] proposed LSB matching revisited image steganography and edge adaptive scheme which can select the embedding regions according to the size of secret message For large embedding rates, smooth edge regions are used while for lower embedding rate, sharper regions are used.Ahn et al.,[3] propose an image steganographic method based on chaos and euler Theorem in which hidden message can be recovered using orbits which is different from the embedding orbits, and the original image is not required to extract the hidden message. Hassan Mathkour et al., [4]use a new Image steganography scheme based on LSB replacement technique and pixel value differencing. This scheme involve replacement of least significant bits in order to hide the colored message image with the advanced LSB methodology wherein the bit replacement takes place in accordance to range specified for the color images. Dobisicek et al., [5] proposed an authentication model of steganography to detect any attack on the stego image by modifies two coefficients of the Discrete Wavelet Transform in each row of cover image based on a verification code. Neil Provo et al.,[6]has proposed another method to counter the statistical attack is known as Out Guess .In this method corrections are made to the coefficients to make the stego-image histogram match the cover image histogram.Pavan et al.,[7] used entropy based technique for finding the coefficients in the image where message can be embedded with minimum distortion. Mohammad Shirali-Shahreza [8] proposed a synonym text steganographic technique in which the words in American English are substituted by the words having different terms in British English and vice-versa. Chen Ming et al., [9] discussed different steganography tool algorithms and classified the tools into spatial domain, transform domain, document based, file structure based and other categories such as spread spectrum technique and video compressing encoding.

- Mankun Xu et al., [10] proposed a Model Based steganography technique which is based on least square method to estimate the embedding rates of secret information .

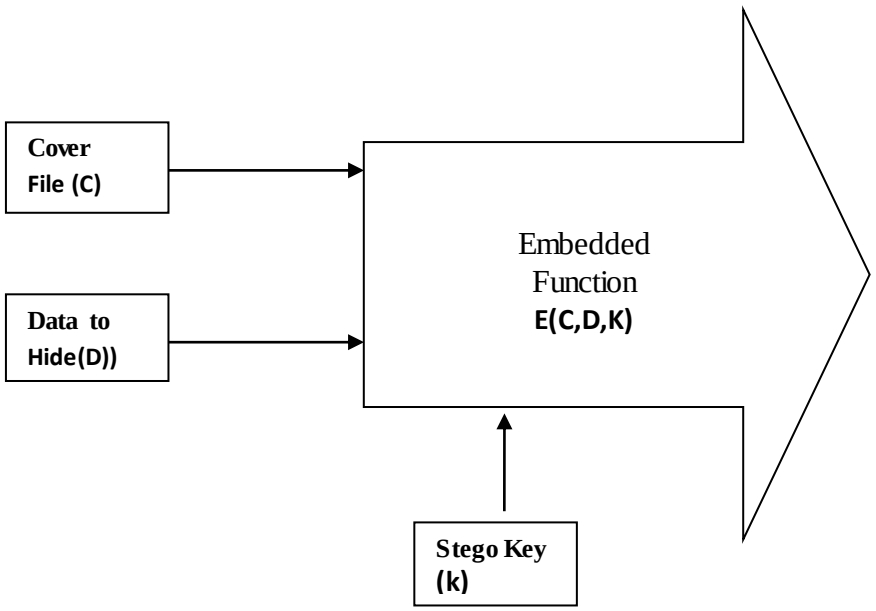


Figure1: **Steganography Diagram**

V. Steganography Applications

Steganography provide secure communication and help in storing of secret data. It can hide a secret message in another message, be it text, image, audio or whatever media you decide to hide the secret message in it. Other applications are TV broadcasting, video-audio synchronization, protection of data alteration, companies safe circulation of secret data, Access control system for digital content distribution, TCP/IP packets(in which unique ID can be embedded into an image to analyze the network traffic of particular users.

VI. Steganography Types

Steganography may be classified as pure, symmetric and asymmetric. While pure steganography does not need any exchange of information, symmetric and asymmetric need to exchange of keys prior sending the messages. Steganography is highly dependent on the type of media being used to hide the information. Medium being commonly used include text, images, audio files, and network protocols used in network transmissions . Image Steganography is generally more preferred media because of its harmlessness and attraction. Additionally exchange of greetings through digital means is on the increase through the increased used of the internet and ease of comfort and flexibility is sending them. Technology advancement in design of cameras and digital images being saved in cameras and then transfer to PCs [12] has also enhanced many folds. Secondly, the text messages hidden in the images does not distort the image and there are techniques which only disturb only one bit of an image who's effects is almost negligible on its quality. The major drawbacks of steganography is that one can hide very little information in the media selected.Some methods are following.

- Encoding secret message in text/documents
- Encoding secret message in audio
- Encoding secret message in images

6.1 Text Steganography: Text Steganography embed the secret data in text files through various techniques:

Format based Method

- **Random and statistical method**
- **linguistic method.**

6.1.1 Format based method: This method modifies the existing text in order to hide the steganographic text. Format Based Method involves the insertion of spaces, resizing the text, change the style of text to hide the secret message.

6.1.2 Random and statistical method: Random method hide the characters that are appeared in random sequence.Statistical methods [13]determine the statistics such as means, variance and chi square test which can measure the amount of redundant information to be hide within the text.

6.1.3 Linguistic Method: Linguistic method is a combination of syntax and semantics methods.Linguistic steganography considers the linguistic properties of generated and modified text, and uses linguistic structure as the space in which messages are hidden. Syntactic steganalysis is to ensure that structures are syntactically correct. Because the text is generated from the grammar, unless the grammar is syntactically flawed, the text is guaranteed to be syntactically correct.In Semantic Method you can assign the value to synonyms and data can be encoded into actual words of text.

6.2 Audio Steganography: Embedding secret messages into digital sound is known as audio Steganography. Audio Steganography methods can embed messages in WAV, AU, and even MP3 sound files. there are three techniques that are used in audio steganography are:

- Low bit Encoding
- Phase Encoding
- Spread Spectrum Encoding

6.2.1 Low Bit Encoding: It is used in audio communications like mobile communications and VOIP. It performs to embed the data while pitch period prediction is conducted during low bit-rate speech encoding, thus maintaining synchronization between information hiding and speech encoding.

6.2.2 Phase Encoding: It split the original audio stream file into blocks and embeds the whole secret sequence into the phase spectrum of the first block. One drawback of the phase encoding method is that less message capacity because message is stored in only first block.

6.2.3 Spread Spectrum Encoding: It is a form of radio frequency communication. Data sent using the spread spectrum encoding is intentionally spread across as much of the frequency spectrum as possible. One Particular method of Spread Spectrum Encoding is DSSS(Direct Sequence Spread Spectrum) which spread signal by multiplying it by a certain maximal length pseudorandom sequence, which is known as chip. Then the calculation of start and end quanta is taken by the discrete, sampled nature of host signal for phase locking purpose. As a result, higher chip rate occur and you can hide maximum data in that chip

.

6.3 Image steganography: Images are cover object used for steganography. Image files are used for storing of digital images. An image file may store data in compressed, uncompressed format. In Image Steganography ,data hiding method can be classified into two categories. They are spatial domain and frequency domain. Spatial [14] involve direct manipulation of pixels in an image. Frequency domain techniques is based on modifying the Fourier transform of an image. Steganography algorithm operate on three types of images:Pallete based images(i.e.GIF images),Raw images(i.e,BMP format) and JPEG images. One of the most popular format used on the internet is JPEG(Joint Photographic Expert Group).It provide large compression ratio and maintain high image quality by measuring PSNR value.
In the JPEG compression, an image is divided into 8*8 blocks and then DCT is applied on each block. **Discrete Cosine Transformation** [15] is used for data compression. It is Similar to Fast Fourier Transform, DCT converts data (pixels, waveforms, etc.) into sets of frequencies. After that resultant DCT coefficient matrix is quantized using a quantization table. Quantization table is a matrix just contain DCT coefficients. Finally the inverse DCT of quantized coefficients is evaluated and finally jpeg image is obtained.

6.3.1 Jpeg-Jsteg In Steganography, there is JPEG data hiding tool jpeg-jsteg. It embed the data into least significant bits of the quantized DCT coefficients where values is not 0,1,-1.The main disadvantage of Jpeg-Jsteg has less capacity to hide the message.Also,Andreas Westfield noticed that by modifying low frequency coefficients cause a distortion detectable by a steganographic method.

VII. Proposed method

This paper proposed a new steganographic algorithm for hiding text files in images. Here we have also used an underlying compression algorithm with maximum compression ratio of 8 bits/ pixel. We have developed a system in java based on the proposed algorithm. Here we have tested few images with different sizes of text files to be hidden and concluded that the resulting stego images do not have any noticeable changes. Also we found that for .bmp images this algorithm works very efficiently. Hence this new steganographic approach is robust and very efficient for hiding text files in images.

As mentioned in the previous section, almost all steganography research done in the JPEG transformation domain which divides a given cover image into non-overlapping blocks of 8*8 pixels. Since the research to increase the message hiding capacity by proposing a new steganography method based on JPEG and quantization table is a continuous process.

VIII. Conclusion

In the past few years, the steganography is interested topic for image cover media. This paper provide an overview of steganography and introduce some techniques of steganography which help to embed the data. These techniques are more useful for detecting the stego images as well as the image media relating to security of of images and embed the data for complex image area and you can easily estimate the high embedding rate by using the quantitative steganalytic technique
Steganography will continue to increase in popularity over cryptography. As it gets more and more advanced as will the steganalysis tools for detecting it. At the time though most of the tools can detect the files hidden in any image.It is well accepted though, small sentences and one-word answers example a „yes" are virtually impossible to find. This could be an area for further advances as possible compression sizes decreases further. There also seems very little in terms of tools for hiding data in videos. There are some for audio, but this is still an area, which lags behind image

steganography. The future may see audio files and video streams that could possibly be decoded on the fly to form their correct messages.

REFERENCES:

[1] N.F. Johnson, S. Jajodia, Exploring steganography: seeing the unseen, IEEE Computer 31 (2) (1998) 26–34.
[2] J.C.Judge, Steganography: past, present, future. SANS Institute publication, <http://www.sans.org/ reading room/ whitepapers/ stenganography/ 552.php>, 2001
[3] F.A.P.Petitcolas, R.J.Anderson, M.G.Kuhn:"Information Hiding- A Survey", Process of IEEE, vol.87,no.7, pp.1062-1078, July, 1999.
[4]Artz D (2001). "Digital steganography: hiding data within data" Internet Computing. IEEE, 5(3): 75-80
[5] Derek Upham, Jsteg, http://zooid.org/ Paul/crypto/jsteg.
[6] Nassir Memon R. Chandramouli. Analysis of lbs. based image steganography techniques. In *Proceedings of IEEE ICIP*, 2001.
[7] R. Chandramouli, Nassir Memon, "Analysis of LSB Based Image Steganography Techniques", IEEE 2001.
[8] K. Sullivan, Z. Bi, U. Madhow, S. Chandrasekaran, and B. S. Manjunath, Steganalysis of quantization index modulation data hiding, Proc. of 2004 IEEE International Conference on Image Processing, vol. 2, pp. 1165-1168, 2004.
[9]Jar no Mielikainen, "LSB Matching Revisited", Signal Processing Letters, IEEE, Publication Date: May 2006 Volume : 13, Issue : 5, pp. 285- 287
[10] K. Gopalan. Audio steganography using bit modification. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '03), volume 2, pages 421–424, 6-10 April 2003.
[11] Xuping Huang, Ryota Kawashima, Norihisa Segawa, Yoshihiko Abe. "The Real-Time Steganograph Based on Audio-o-Audio Data Bit Stream",Technical report of IEICE, ISEC, vol.106 pp.15-22, September 2006.
[12] Haz Malik, Steganalysis of qim steganography using irregularity measure, Proc. of the 10th ACM workshop on Multimedia and security, ACM Press, pp. 149-158, 2008.
[13] A. Delforouz, Mohammad Pooyan, "Adaptive Digital Audio Steganography Based on Integer wavelet transform ", IEEE Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2007, 26-28 Nov 2007, pp 283-286.
[14] M. Goljan, J. Fridrich, and T. Holotyak, New blind steganalysis and its implications, IST/SPIE Electronic Imaging: Security, Steganography of Multimedia Contents VIII, vol. 6072, pp. 1-13, 2006.
[15] Y. Wang and P. Moulin, Optimized feature extraction for learning-based image steganalysis, IEEE Trans. Information Forensics and Security, vol. 2, no. 1, pp. 31-45, 2007.
[16] Jan Kodovsky and J. Fridrich, Inuence of embedding strategies on security of steganographic methods in the jpeg domain, Proc. of IST/SPIE Electronic Imaging: Security, Forensics, Steganography Contents X, vol. 6819, pp. 1-13, 2008.
[17] M.H. Shirali-Shahreza and M. Shirali-Shahreza. Text steganography in chat. In Proceedings of the Third IEEE/IFIP International Conference in Central Asia on Interne the Next Generation of Mobile, Wireless and Optical Communications Networks (ICI 2007), Tashkent, Uzbekistan, September 26-28, 2007.
[18] C.Y. Yang, .Color Image Steganography based on Module Substitutions,. Third International Conference on International Information Hiding and Multimedia Signal Processing Year of Publication: 2007 ISBN:0-7695-2994-1.

[19] Yincheng Qi, Jianwen Fu, and Jinsha Yuan, "Wavelet domain audio steganalysis based on statistical moments of histogram", Journal of System Simulation, Vol 20, No. 7, pp. 1912-1914, April 2008.

[20] Sutaone, M.S., Khandare, M.V, "Image based steganography using LSB insertion technique", IEEE WMMN, January 2008.

[21] H. Wu, H. Wang, C. Tsai and C. Wang, Reversible image steganographic scheme via predictive coding. 1 (2010), ISSN: 01419382, 35-43.

[22] J, Corporation, Steganography. http://www.webopedia.com/ TERM/S/steganography.html. 2005.

[23] B. Dunbar. A detailed look at Steganographic Techniques and their use in an Open-Systems Environment, Sans Institute, 1(2002).

[24]C.Christian. An Information-Theoretic Model for Steganography, Proceedings of 2nd Workshop on Information Hiding, MIT Laboratory for Computer Science. 1998.