International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Steganographic Method of Data Hiding using JPEG Images

Shabana Vathelil Subair[1], Fathima A Muhammadali[2]

[1]Final Year M. Tech. (Cyber Security), KMP College of Engineering, Perumbavoor, Kerala, India

[2]Assistant Professor, Department of Computer Science and Engineering, KMP College of Engineering, Perumbavoor, Kerala, India

Abstract: ***Steganography is an art of concealing the fact that communication is taking place, by hiding information in other information. Different formats can be used as medium, but digital images are the most popular in that, because of their frequency in the web. For hiding secret data in images there exist a large variety of steganographic methods, i.e. there exist complex method to simpler ones. Different applications may adopt different steganography techniques. Some applications may require complete invisibility of the secret information while some other may require a larger secret message to be hidden. This paper gives an overview of steganographic method in which we hide both text message and digital image in another digital image. To this end, the widely adopted framework for the development of the steganographic system is the minimal distortion embedding, in which a well-designed distortion function is the most important.***

Keywords: JPEG steganography, uniform embedding, cover image, stego image, discrete cosine transform, quantization

1. Introduction

Steganography technique transfer information over the public data transfer channels in such a way that an attacker is unable to identify the transmission of secret information on the background of a public communication. In general, efficiency of the secret information detection in any steganography system is related to the size of the hidden secret data.To this end, the framework of minimal distortion embedding is widely adopted in the development of the steganographic system, in which a well-designed distortion function is of vital importance.

Steganography is a data hiding technique where the 3rd party would be completely ignorant about any hidden information. Usually steganography is implemented using some cover medium which include an image, audio or video. No one would have the faintest of ideas that any information would be hidden within these mediums. The advantage of steganography, over cryptography alone, is that messages do not attract attention to themselves. because cryptography encodes the secret message in unreadable form for third party but steganography hides the secret message behind some text ,image, video etc. so that third party is unaware of secret hidden message.

In the digital world of today data encoding is widely accepted. This is because it can take advantage of the limited power of human vision. When dealing with digital images for use with Steganography, 8-bit and 24-bit per pixel image files are typical. Both have advantages and disadvantages, as we will explain below. 8-bit images are a great format to use because of their relatively small size. The drawback is that only 256 possible colors can be used which can be a potential problem during encoding. 24-bit images offer much more flexibility when used for Steganography. The large numbers of colors (over 16 million) that can be used go well beyond the human visual system (HVS), which makes it very hard to detect once a secret message, has been encoded. The other benefit is that a much larger amount of hidden data can be encoded into a 24-bit digital image as opposed to an 8-bit digital image. The one major drawback to 24-bit digital images is their large size (usually in MB) makes them more suspect than the much smaller 8-bit digital images (usually in KB) when sent over an open system such as the Internet. Digital image compression is a good solution to 24-bit digital images but it has drawback that the possibility of the uncompressed secret message to lose parts of its contents least significant bit (LSB) encoding. It is by far the most popular of the coding techniques used for digital images. By using the LSB of each byte (8 bits) in an image for a secret message, you can store 3 bits of data in each pixel for 24-bit images and 1 bit in each pixel for 8-bit images.

Steganography includes the concealment of information within computer files. In digital steganography, electronic communications may include steganographic coding inside of a transport layer, such as a document file, image file, program or protocol. Media files are ideal for steganographic transmission because of their large size. For example, a sender might start with an innocuous image file and adjust the color of every 100th pixel to correspond to a letter in the alphabet, a change so subtle that someone not specifically looking for it is unlikely to notice it.

To conceal the very existence of communication, the stego image has to be statistically undetectable from its cover counterpart. Therefore, the two conflicting objectives, i.e., undetectability and embedding payload, should be carefully considered when devising a steganographic scheme.

In a more generalized way, it can be said that steganography is a two-step process. In the first step, an analysis of the cover image is done to find the insignificant bits. It is expected that modifying these bits will not cause any observable changes in the cover medium. In the second step, these bits are replaced by message bits to create the stego image. Generally these insignificant bits are the LSB's of the

image. In JPEG images, modifying the LSB creates imperceptible distortions of the original image. Here the intention is to reduce these distortions and also enhance the undetectability and thereby improving the security.

JPEG is the most commonly used format for digital communication and due to this the field of JPEG steganography is being researched on an extensive basis. The Joint Photographic Experts Group (JPEG) file format stores image data in compressed form as quantized frequency coefficients. The compressing steps performed by the JPEG compressor starts by cutting the uncompressed bitmap image into parts of 8 x 8 pixels. The 8 x 8 brightness values are transformed into 8 x 8 frequency coefficients by using Discrete Cosine Transform (DCT). After DCT, quantization rounds up the frequency coefficients to integers in the range of -2048 to 2047. Analysis of a discrete distribution of coefficients frequency of occurrence shows two characteristics viz. I) The coefficients degree of occurrence decreases with increasing absolute value and II) the decrease of the coefficient's frequency of occurrence decreases with the increasing absolute value, that is difference between two bars of the histogram in the middle is larger than on the margin.

2. Related Works

This section describes the various existing schemes which are compared in this paper.

2.1 Minimizing additive distortion in steganography using syndrome-trellis codes

A full practical method is given for minimising the additive distortion [1] by using general i.e. non-binary embedding operation. Each possible value of every stego item can be designated a scalar value which can express the distortion caused by the embedding done by changing the cover element with this value. Once the steganographer specifies the form of the distortion function, the proposed framework provides all essential tools for constructing practical embedding schemes working close to their theoretical bounds. The methods are not limited to binary embedding operations and allow the embedder to choose the amplitude of embedding changes dynamically based on the cover-image content. The distortion function or the embedding operation do not need to be shared with the recipient. In fact, they can even change from image to image. The framework can be thought of as an off-the-shelf method that allows practitioners to concentrate on the problem of designing the distortion measure instead of the problem of how to construct practical embedding schemes.

2.2 F5—A steganographic algorithm

F5 algorithm [2] which provides large steganographic capability and it can also deal very efficiently with visual and statistical attacks. F5 algorithm uses matrix encoding technique to increase the performance of embedding. It is known that the images provide limited steganographic capabilities, also many a times embedding work do not require the full capacity of the image. Thus, it can be said that some part might be left unused. Some of the prominent steganographic algorithms attempt to scatter the message over the entire cover element. This might cause them to have a bad time complexity. This may be the case when the algorithm tries to use up the capacity of the image completely. The task of straddling can be made easy if the exact capacity of the carrier element is known.

2.3 Statistically undetectable JPEG steganography: Dead ends challenges, and opportunities

To determine the steganographic capacity of JPEG images (the largest payload that can be undetectably embedded) with respect to current best steganalytic methods. Additionally, by testing selected steganographic algorithms we evaluate the influence of specific design elements and principles, such as the choice of the JPEG compressor, matrix embedding, adaptive content-dependent selection channels, and minimal distortion steganography using side information at the sender. In this the statistical detectability of current steganographic methods for JPEG images with several goals are investigated 1) to determine the maximal relative payload at which the methods become statistically undetectable, 2) to study the influence of various design elements, such as matrix embedding, adaptive selection rules, minimizing embedding impact using side information at the sender, and the type of the embedding modification, 3) to evaluate the influence of different JPEG compressors on steganalysis and 4) to present steganalysis results for the most promising candidate methods as well as their modifications [3].

2.4 Modified matrix encoding technique for minimal distortion steganography

All information hiding methods that modify the least significant bits [4] introduce distortions into the cover objects. Those distortions have been utilized by steganalysis algorithms to detect that the objects had been modified. It has been proposed that only coefficients whose modification does not introduce large distortions should be used for embedding. In Modified Matrix Encoding Technique for Minimal Distortion Steganography paper we propose an efficient algorithm for information hiding in the LSBs of JPEG coefficients. The algorithm uses modified matrix encoding [4] to choose the coefficients whose modifications introduce minimal embedding distortion. Author derive the expected value of the embedding distortion as a function of the message length and the probability distribution of the JPEG quantization errors of cover images. The experiments show close agreement between the theoretical prediction and the actual embedding distortion.

2.5 Less detectable JPEG steganography method based on heuristic optimization and BCH syndrome coding

A new Bose-Chaudhuri-Hochquenghem (BCH)-based data hiding scheme for JPEG steganography [5] is presented. Traditional data hiding approaches hide data into each block, where all the blocks are not overlapping each other. However, in the proposed method, two consecutive blocks can be overlapped to form a combined block which is larger

than a single block, but smaller than two consecutive non overlapping blocks in size. In order to embed more amounts of data into the combined block than a single block, the BCH-based data hiding scheme has to be redesigned. In this article, author propose a way to get a joint solution for hiding data into two blocks with intersected coefficients such that any modification of the intersected area does not affect the data hiding process into both blocks. Due to hiding more amounts of data into the intersected area, embedding capacity is increased. On the other hand, the nonzero DCT coefficient stream is modified to achieve better steganalysis and to reduce the distortion impact after data hiding. This approach carefully inserts or removes 1 or -1 coefficients into or from the DCT coefficient stream according to the rule proposed in this article.

3. Theoretical Background

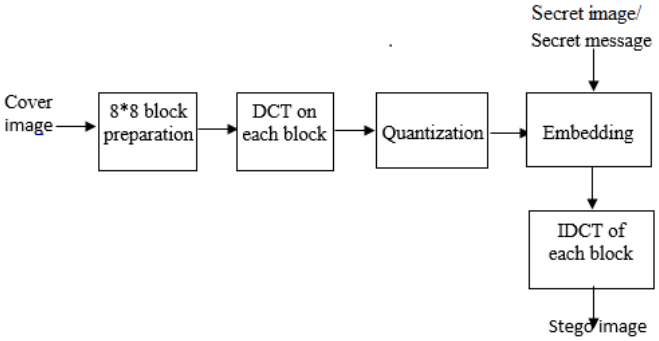
Steganography is a Image Data Hiding technique for increasing the embedding efficiency that is defined as an average number of bits embedded via per change on the cover. Hiding information inside an image without changing its visible properties, the cover source can be altered in noisy areas with many color variations, but the color values for existing and stego image are same. Reducing distortion between the cover object and the stego object is an important issue for steganography. The purpose of steganography is to send secret messages after embedding them into public digital multimedia.

3.1 JPEG Image File

JPEG files use data–loss compression. A redundant graphical information is discarded by this method without a significant impact on the picture. It is possible to achieve much better compression ratio that way than with the lossless compression. JPEG is a standard that prescribes a sequence of operations that are performed with visual data. These operations are: 1) subsampling of the image 2) transformation of a blocked representation of the image to a frequency domain representation using the discrete cosine transform 3) quantization of the blocked frequency domain data 4) coding of the frequency domain data.

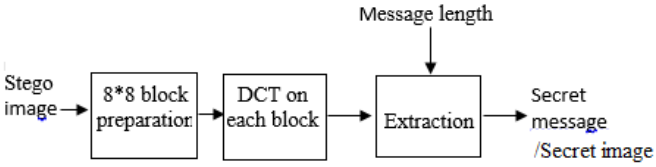
3.2 Proposed JPEG Steganographic Scheme

a) Data Embedding



The proposed system uses JPEG images to hide the data. In recent years, compressed JPEG images have become the most popular image on the internet, primarily because they take less space than other raw images, but also provide great visual quality with typical compression factors. The paper presents a high steganography capacity method based on modification JPEG quantization tables. This method uses a new 16×16 quantization table instead of the traditional JPEG steganographic methods using standard 8×8 quantization table. With our proposed quantization table, The DCT coefficients are quantized and embedded the secret messages. The quantized samples then applied the zigzag transformation. It is then encoded using Huffman coding. The data is embedded using LSB (Least Significant Bit algorithm).The experiment results show that our method has both larger steganography capacity and better stego-image quality.

b) Data Extraction



4. Conclusion

Minimal-distortion embedding framework is a practical approach to implement JPEG steganography with high embedding efficiency. In this paper, an efficient JPEG steganographic scheme which utilizes STC and UE strategy is presented. By uniformly –spreading" the embedding modifications to quantized DCT coefficients of all possible magnitudes, the average changes of statistics are possibly minimized, especially in the small coefficients, which leads to less statistical detectability, and hence, more secure steganography. Extensive experiments have been carried out to demonstrate the superior performance of the proposed scheme in terms of secure embedding payload against steganalysis.

Finally, it is worthwhile to note that, the inappropriate use of DC and zero AC coefficients in JPEG steganography may lead to additional block artifacts in stego image and decrease in the efficiency of JPEG compression, respectively. That is why the most existing JPEG steganographic schemes use only non-zero AC coefficients as possible cover elements to make the embedding naturally content-adaptive. According to some of our experiments, some DC and zero AC coefficients in the texture regions could indeed be incorporated to further data embedding without decreasing, even increasing the security performance. While the UED proposed in this paper could by no means tackle all of these issues, it raises a quite challenging open question, that is how to evaluate the embedding costs of all possible DCT coefficients (including DCs, zero and nonzero ACs) based solely on the coefficients in the DCT domain for JPEG steganography, which remains as the topic of our future research effort.

References

[1] T. Filler, J. Judas, and J. Fridrich, "Minimizing additive distortion in steganography using syndrome-trellis codes," IEEE Trans. Inf. Forensics Security, vol. 6, no. 3, pp. 920–935, Sep. 2011.
[2] A. Westfeld, "F5—A steganographic algorithm," in Proc. 4th Inf. Hiding Conf., vol. 2137. 2001, pp. 289–302.
[3] J. Fridrich, T. Pevný, and J. Kodovský, "Statistically undetectable JPEG steganography: Dead ends challenges, and opportunities," in Proc. 9th ACM Workshop Multimedia Security, Dallas, TX, USA, Sep. 2007, pp. 3–14.
[4] Y. Kim, Z. Duric, and D. Richards, "Modified matrix encoding technique for minimal distortion steganography," in Proc. 8th Inf. Hiding Conf., vol. 4437. Jul. 2006, pp. 314–327.
[5] V. Sachnev, H. J. Kim, and R. Zhang, "Less detectable JPEG steganography method based on heuristic optimization and BCH syndrome coding," in Proc. 11th ACM Workshop Multimedia Security, Sep. 2009, pp. 131–140.
[6] T. Filler and J. Fridrich, "Design of adaptive steganographic schemes for digital images," Proc. SPIE, vol. 7880, p. 78800F, Jan. 2011.
[7] C. Wang and J. Ni, "An efficient JPEG steganographic scheme based on the block entropy of DCT coefficients," in Proc. IEEE ICASSP, Kyoto, Japan, Mar. 2012, pp. 1785–1788.
[8] J. Kodovský and J. Fridrich, "Calibration revisited," in Proc. 11th ACM Workshop Multimedia Security, New York, NY, USA, Sep. 2009, pp. 63–74.
[9] J. Kodovský, J. Fridrich, and V. Holub, "On dangers of overtraining steganography to incomplete cover model," in Proc. 13th ACM Workshop Multimedia Security, New York, NY, USA, Sep. 2011, pp. 69–76.
[10] V. Holub and J. Fridrich, "Digital image steganography using universal distortion," in Proc. 1st ACM Workshop Inf. Hiding Multimedia Security, 2013, pp. 59–68.
[11] J. Kodovský and J. Fridrich, "Steganalysis of JPEG images using rich models," Proc. SPIE, vol. 8303, p. 83030A, Jan. 2012.
[12] J. Kodovský, J. Fridrich, and V. Holub, "Ensemble classifiers for steganalysis of digital media," IEEE Trans. Inf. Forensics Security, vol. 7, no. 2, pp. 432–444, Apr. 2012.
[13] C. Chen and Y. Q. Shi, "JPEG image steganalysis utilizing both intrablock and interblock correlations," in Proc. IEEE Int. Symp. Circuits Syst., Mar. 2008, pp. 3029–3032.
[14] L. Guo, J. Ni, and Y. Q. Shi, "An efficient JPEG steganographic scheme using uniform embedding," in Proc. 4th IEEE Int. Workshop Inf. Forensics Security, Tenerife, Spain, Dec. 2012, pp. 169–174.
[15] P. Bas, T. Filler, and T. Pevný, "Break our steganographic system—The ins and outs of organizing boss," in Proc. 13th Inf. Hiding Conf., 2011, pp. 59–70.
[16] N. Provos, "Defending against statistical steganalysis," in Proc. 10th USENIX Security Symp., Washington, DC, USA, 2001, pp. 323–335.
[17] D. Freedman, Statistical Models: Theory and Practice. Cambridge, U.K.: Cambridge Univ. Press, 2009.
[18] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," ACM Trans. Intell. Syst. Technol., vol. 2, no. 3, pp. 27:1–27:27, 2011.

Author Profile

Shabana Vathelil Subair received the B.Tech degree in Computer Science & Engineering from Mahatma Gandhi University, Kottayam in 2012 and currently pursuing final year M.Tech degree in Computer Science and Engineering with specialization in Cyber Security from KMP College of Engineering, Perumbavoor.

Fathima A Muhammadali received B.Tech degree in Computer Science & Engineering from Mahatma Gandhi University Kottayam in 2007 and received M.Tech in Computer Science & Engineering from Mahatma Gandhi University Kottayam in 2014 and currently working as assistant professor in KMP College of Engineering, Perumbavoor in Computer Science and Engineering Department.