A MODIFIED LSB MATCHING STEGANOGRAPHY TECHNIQUE FOR IMAGES TO HIDE EXECUTABLE FILES

Mohan Kumar P.[1], Shunmuganathan K.L.[2]

[1]Associate Professor, CSE Department, Jeppiaar Engineering College, Chennai.
[2]Professor and Head, .M.K. Engineering College, Chennai.
Email: [1]mohankumarmohan@gmail.com, [2]kls_nathan@yahoo.com

Abstract

Steganography is the science of hiding secret information in any cover data. It is mainly used for hidden communication between any people. The main aim of steganography concentrates on higher data payload ie., the maximum amount of secret data that can be hidden inside a particular cover media and the invisibility of the hidden data. This paper proposes a new algorithm for the same and the secret data we are hiding is an executable file.

Keywords: steganography, stego data, payload, image segments, cover image.

I. INTRODUCTION

The term steganography means the science of hidden communication. The way in which steganography differs from another secure data communication technique called cryptography is, the visibility of the data exchange. In cryptography, even though the actual data transaction may not be known to a third person, he may get a doubt that some abnormal or suspicious communication is taking place. But, in case of steganography, the hidden communication will never come to the notice of the eavesdropper. Because, the carrier signal we are using to hide the secret data is going to be innocent. So, we can call the technique as information hiding [1-4].

Another technique which is based on the information hiding strategy is digital watermarking. But, in case of digital watermarking, the important property of information hiding known as resistance to removal is preferred. So, in these applications, we are not worrying about imperceptibility but resistance to removal. This is mainly used in commercial applications like copyright protection of digital forms of media like video or image. Unlike image steganography, digital watermarking techniques mainly concentrate on keeping logos or any other symbols or images in the carrier data. And also it is made sure that those signals embedded are not able to be removed by any other person. There are a number of watermarking techniques have been explained in [5-7].

For a long period of time many researchers have been involved in developing new steganographic systems. Meanwhile, the development of steganalytic tools are also started growing. Steganalysis is a process of finding the existence of a secret data in a cover media [8]. Whenever a suspicious image is received, the main task of a steganalytic tool is to find the algorithm used for hiding secret data in the image. Most of the steganographic algorithm developers are also trying to crack their own algorithm using the existing steganalytic tools, so that the strength and weaknesses of their system may be found.

II. RELATED WORK

Generally, digital image steganography is a way to exchange secret data. So, the important components of a steganographic system include an embedding / extracting algorithm, secret key which is going to be shared by the sender and receiver of the secret data and also a communication channel which is considered to be more secure [9]. The general frame work for a steganographic system is shown as the figure.

This framework has been derived from the popular idea called prisoner's problem [9]. In this approach Alice and Bob were trying to exchange an

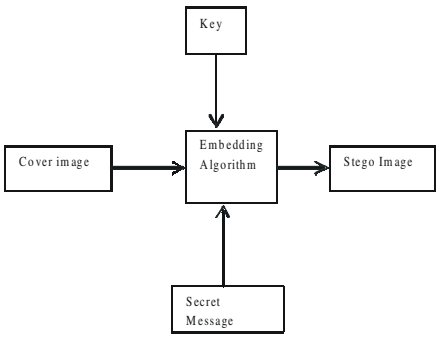

Fig 1: A simple image steganography scheme

escape plan without the knowledge of the warden. Some of the terms used in steganographic system are cover-media (the digital media which is used to hide secret data), secret data (the important data to be hidden) and stego-media (after embedding the secret message in the cover media). The hidden data cannot be detectable when we are performing the embedding phase randomly and also the level of independence between the secret message and cover as well as stego objects [10]. There are many other ways for providing more security includes the usage of encryption-decryption functions for embedding and extraction of secret data [11]. Since JPEG images are widely exchanged through internet, choosing JPEG image for sending secret message to the receiver will never be suspicious. And also, the redundancies that are appearing in JPEG images help us to hide more information securely. Methods for improving the hiding capacity of a JPEG image have been explained in [12].

The Least Significant Bits replacement method or LSB method is the very simple and a commonly used approach for developing steganographic system. Because the amount of space that an image can provide for hiding data will be more comparing with other algorithms. And also the implementation of this technique is also very easy. In this approach, the image pixel's LSB is replaced by one bit of secret data [13]. Spatial domain embedding technique is also known as image domain. The techniques that are following spatial domain embedding are embedding the secret message in the intensity of the cover image pixels. Spatial domain techniques include bit-wise methods that apply bit insertion and noise manipulation techniques [14]. The main disadvantage of LSB replacement is that, while hiding secret data in the image, some of the pixels will never be modified or replaced with the secret bits, since we are using pseudo random generator for placing the secret message bits. As a result, very simple steganalytic tool could trace the existence of the secret message.

But this problem of asymmetry can easily be avoided by an alternate scheme using a LSB matching scheme. In this technique, if the secret bit is not matching with the LSB of cover image, then ± 1 will be added randomly. By doing so, we can reduce the probability of increase or decrease in the pixel value modification can be avoided. So, we can eliminate the problem we faced in LSB replacement technique. Also, the steganalytic algorithms which can find the stego-images which were obtained from LSB replacement technique cannot find the stego-images we got from LSB matching.

There are several steganalytic algorithms found for finding stego-images which were got by LSBM (LSB matching) technique. In [15], the image is being taken and its two least significant bit planes are considered. The bit planes are split into 3×3 overlapped sub images. According to the number of gray levels those sub images are classified. In one sub image, the LSBM is applied and found that the alteration rate of cover image is higher than that of stego-image. In [16], the authors have compared the function of LSBM to a low pass filter through the histogram of the image. They found that the no. of high frequency components is very less comparing to the original cover image. But later in [17], this method is found that it will not be working well in case of gray scale images. As a remedy, the author has proposed techniques using down-sampled image and adjacency histogram instead of traditional histogram.

Instead of handling pixel values independently, the other technique proposed by Jarno in [18], is using a pair of pixels for embedding which is known as LSBM revisited (LSBMR). In this technique, the author has proposed an approach for data hiding, in which the LSB of the first pixel carries one bit of information, and a function of the two pixel values carries another bit of information. So, in this approach the changes that are made in the cover image are very few. Also, the modification rates of pixels have been greatly reduced. But all the techniques analyzed above are not taking care of the relationship between the pixel and its neighborhood.

There are many data embedding schemes analyzed which are taking the relationship of a pixel to its neighbor. In [19], a hiding scheme has been proposed by replacing the LSB of a cover according to the difference values between a pixel and its four touching neighbors. This method uses the edges of an image for hiding secret data. Although this method can achieve more visually imperceptible stego-images, the security performance is poor. Since the method just modifies the LSB of image pixels when hiding data, it can be easily detected by existing steganalytic algorithms.

The pixel Value differencing is another type of edge based data hiding scheme, which has been proposed in [20], in which the number of embedded bits is determined by the difference between a pixel and its neighbor. If there is large difference between the pixels, the number of secret bits that can be embedded will also be large. Also based on the experimental results, this approach can provide a larger embedding capacity.

Mostly, all the techniques discussed so far are using a random number generator which will spread the secret data throughout the image which may lead to lower embedding rates. But based on the experimental results, we have found that most of the approaches fail to provide security and stego-images of preferred quality. Also, the usage of JPEG images in internet now a days is very high, we have proposed an approach which is using the JPEG image as cover image. But the thing is, sufficient secret message bits can be embedded into cover image, so that we can achieve better secure payload. To make the JPEG image useful for embedding more data, an encoder scheme has been proposed in [12]. In this research, it is found that, it is necessary to enlarge the hiding capacity of JPEG2000 baseline system because the available redundancy is very limited. In addition, the bit stream truncation makes it difficult to hide information. These problems faced are being eliminated by introducing an encoder which eliminates bit truncation.

III. THE PROPOSED STEGANOGRAPHY

In this section, we shall present the proposed method, whose embedding and extracting procedures will be illustrated by the block diagrams shown in Figs. 2 and 3, respectively. In our scheme, the embedding procedure consists of three steps, including the pixel value prediction for reducing the distortion, the capacity estimation to achieve a higher payload, and finally the difference expansion that completes the embedding procedure.

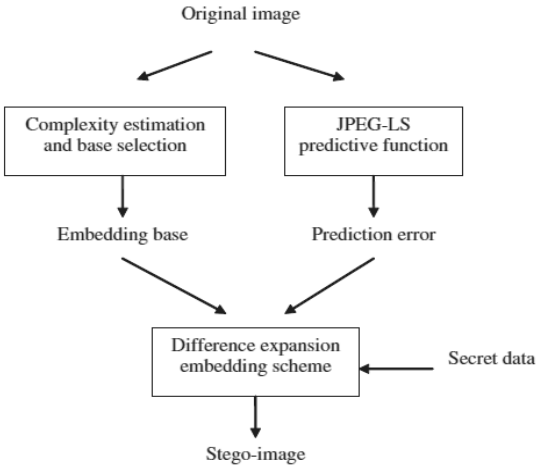

Fig 1. Embedding Phase

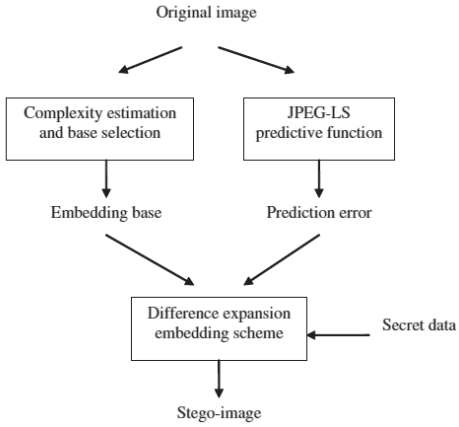

Fig 2. Extracting phase

A. The pixel value prediction phase

In general, the pixels in the top-most row and the left-most column of a cover image are preserved without any secret data hidden in them. Let *Pi*, *j* be the first pixel to hold the secret message, whose position is (2, 2) in the cover image. Three immediate neighbor pixels are *P*1, 2, *P*2, 1 and *P*1, 1. As defined in Section 2.1, the edge prediction technique can find a predictive value for a pixel by using its three adjacent pixels. Therefore, assume *P*0 *i*; *j* is the predictive value of pixel Pi,j and it can be computed by Eq. (1) where $x = Pi, j$, $a = Pi\_1, j$, $b = Pi, j\_1$ and $c = Pi\_1, j\_1$. Then, following raster scan order, predictive pixel values can be obtained for the next pixel and the next until the whole image is done.

B. The capacity estimation phase

The embedding capacity of a pixel is determined by the variance of its neighbor pixels; in other words, the embedding capacity can be different from pixel to pixel. This is because the local texture of each pixel is different. Pixels in smooth areas will be responsible for the embedding of more secret data because the pixel values can be predicted more accurately with the prediction errors close to zero. Then, the secret message will be transformed into various bases by using a multiple-base notational system.

C. The data embedding phase

Now that we have both the predictive pixel value *P*0 *i*; *j* and its base *bi*, *j*, it is time to hide the secret data into the pixel. Suppose *S* is the whole binary secret message whose size is *l*. *S* can be defined as $S = s1, s2, ..., sn$, where *n* is equal to *l*/8 and each symbol sk, $k = 1, 2, ..., n$, contains 8 bits of binary data. The following are the data embedding steps of the proposed scheme:

Step 1: As defined in Section 3.1, the top-most row and left-most column are not used for hiding data. Hence, for either $i = 1$ or $j = 1$ or both, the stego-pixel value *P*00 *i*;jequals *Pi*, *j*.

Step 2: Read the secret data *sk* and convert it into a decimal value *sd*, *k*. In order to judge whether the secret data *sd*, *k* is empty or not, we use a parameter u whose initial value is 1 to check whether the total hiding capacity is more than 8 bits.

Step 3: Compute the difference value *di*, *j* between the original pixel value *Pi*, *j* and the predictive value *P*0 *i*; *jby di*; *j* 1⁄4 *Pi*; *j*_*P*0 *i*; *j* The remainder value *ri*, *j* of the secret data in its decimal form *sd*, *k* is to be embedded with the base *bi*, *j* by *ri*; *j* 1⁄4 *sd*; *k mod bi*; *j* 10.

Step 4: According to the base *bi*, *j*, the original difference value *di*, *j* can be expanded *bi*, *j* times. After that, the secret message is concealed into the expanded value. Therefore, the new difference value *d*0 *i*; *j* can be obtained by *d*0.

Step 5: We can obtain a new stego-image pixel

However, in this step, there might be some overflow or underflow problems; that is, there can be times when the stego-pixel value *P*00 *i*; *j* goes beyond 255 or below 0. If so, the next step, namely step 6, has to be skipped, and the pixel will be preserved because it is not eligible to carry any secret data.

The position of the pixel will be recorded, which takes some additional memory. There are few these points in an image, but can be sent by another channel. The additional recorded data, which called side information, can be included into the file header of cover image or transmitted by public key cryptosystem. Moreover, the side information also can be embedded into the least significant bits (LSBs) of a pre-defined region. For instance, we can embed the side information into the LSBs of last line in cover image. Then, the original LSBs which have been modified will be embedded as secret messages. The embedding process appends original LSBs to the secret message. Hence, the pixels in last line will skip the Steps 1–6. If, on the contrary, no overflow or underflow problems occur, the data embedding algorithm is finished.

Step 6: If the data embedding in Step 5 is successful and the value of u is smaller than 255, then the secret message is not empty. Hence, we can hide the quotient of *sd*, *k* into the next pixel, and the parameter u must be multiplied by its base *bi*, *j* until $u > 255$.

In Step 6, if $u > 255$, it means the message *sk* has been embedded into the pixel. Hence, we read the next message $sk + 1$ and convert it into its decimal value *sd*, $k + 1$ with u set to be 1. After Step 6, we take the next pixel from the cover image and repeat Steps 3 through 6. The same thing goes on and on until all the pixels have been visited.

D. The data extraction and original pixel recovery phase

Here are the steps to take to extract the secret data from the stego-image and recover the original pixel values of the cover image. The detailed secret data extraction and original cover pixel recovery procedures are described as follows:

Step 1: Since the pixels in the top-most row and left-most column do not carry any secret data, we can readily restore them for $i = 1$ or $j = 1$, respectively. The stego-pixels *P i*; *j* will be processed in raster scan order. Then, the side information will be obtained from the file

header or decrypted by public key cryptosystem. These overflow/underflow points can be directly acquired from the recovered pixel values and skip all extracting process.

Step 2: For each stego-pixel $P\,i; j$ with its three adjacent pixels Pi_1, j, Pi, j_1 and Pi_1, j_1 having been restored already, the predictive value $P00\,i; j$ of $P00\,i; j$ can be obtained by Eq. (1). Then, the variance value $d2\,ij$ among the three adjacent pixels $Pi-1, j, Pi, j-1$ and $Pi-1, j-1$. can be computed by Eq. (8), and the base bi, j of $P00\,i; j$ can be derived by Eq. (9).

Step 3: Compute the new difference $d0$

Now we can extract the secret message, namely the remainder $ri; j\,1/4\,d0\,i; j\,mod\,bi, j$, and the original difference value di, j also can be derived by computing the quotient $bd0\,ij = bijc$. We can compute the sd, k value by converting ri, j with the base bi, j into its decimal value.

Step 4: Finally, the original cover pixel value Pi, j can be recovered by $Pi; j\,1/4\,P0\,i; j\,di; j$.

IV. RESULTS AND DISCUSSIONS

The proposed scheme has been conducted on six cover images with either smooth andlor complex contents. Our experimental results were obtained from three cases: case I: $B=3$; case 2: $B=2$; and case 3: $B=2$ but the modifications of $b4$ bits in LH and HL subbands were moved to the bz and bl bits in HH subband, respectively, to increase transparency. When we examine the introduced changes of bits during message embedding, we can find the following attributes. Statistically, the b4bsbzbl ofnearly 40% and 38% high-frequency coefficients would be assigned with a digit no larger than 3 (= 0011) and 6 (= OllO), respectively.

This implies that stego images implicitly contaminated with small noises can be mimiced easily in particular if cover images are carefully selected to the texture-like or edge-like. If we further check the relationship between perceptual quality (measured in PSNR) and two security criteria with respect to the three cases on the resultant six cover-stego image pairs, then we have the following results, as shown in Table 2.

First, kurtosis and relative entropy are respectively almost the same for the three experimental scenarios. This implies that the introduced message really obey behaviour of a Gaussian noise. Second, information embedded with smaller capacity certainly achieves less distortion but seems to not affect the two security measurements. Under the prerequisite of this dual security criterion, we can conclude that the proposed image steganographic scheme really achieve a pre-defined security criterion. Finally, we emphasize that these results should depend on the contents of images.

V. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a new image steganographic scheme under the norm of a dual security criterion. One is defined in the spatial domain that a hidden message can mimic a Gaussian noise (naturally appears in image acquisition systems) if its kurtosis is equal to 3. The other one is defined in the frequency domain based on relative entropy. Most importantly, the dual security criterion has been derived to be correlated with perceptual quality. In order to retain lossless message retrieval our future work will investigate the use of integer wavelet transform due to its lossless characteristic without rounding errors.

REFERENCES

[1] N. Provos and P. Honeyman, “Hide and seek: An introduction to steganography,” *IEEE security and Privacy Mag.*, vol. 1, no. 3, pp. 32–44, 2003.

[2] J. Fridrich, “Applications of Data Hiding in Digital Images,” Tutorial for the ISSPA, pp. 22-25, Aug. 1999

[3] F.A.P. Petitcolas, R.J. Anderson and M.G. Kuhn, “Information hiding—A survey,” *Proc. IEEE*, vol. 87, no. 7, pp. 1062–1078, Jul.1999.

[4] W. Bender, D. Gruhl, N. Morimoto and A. Lu, “Techniques for data hiding,” *IBM Syst. J.*, vol. 35, no. 3–4, pp. 313–336, 1996.

[5] M. Swanson, M. Kobayashi and A. Tewfik, “Multimedia data embedding and watermarking technologies,” *IEEE Proceedings*, vol. 86, No. 6, pp 1064-1087, June 1998.

[6] I. Cox, J. Kilian, T. Leighton and T. Shamoon, “Secure spread spectrum watermarking for multimedia,” *Proc. First Int. Workshop Information Hiding,* R. Anderson, Ed., Cambridge, U. K.: Springer-Verlag, May/June 1996, pp. 183-206.

[7] I.J. Cox, M.L. Miller and J.A. Bloom, Digital Watermarking. Morgan Kaufmann, 2002.

[8] Tomá¡s Pevn and Jessica Fridrich, Multiclass Detector of Current Steganographic Methods for JPEG Format, IEEE Transactions On Information Forensics And Security, Vol. 3, No. 4, December 2008, 635-650.

[9] G. Simmons," The prisoner's problem and the subliminal channel, CRYPTO, pp: 51-67, 1983.

[10] J. Zollner, H. Federrath, "Modelling the security of steganographic systems", 2nd information hiding workshop, pp: 345-355, 1998.

[11] N.J. Hopper, J. Langford, L. Von Ahn, "Provably secure steganography", Advances in cryptology: CRYPTO 2002.

[12] A high capacity Steganography scheme for JPEG 2000 Baseline system, Liang Zhang, Haili wang, Renbiao Wu, IEEE transactions on image processing, 18(8), 2009.

[13] T. Morkel 1, J.H.P. Eloff 2, M.S. Olivier , AN OVERVIEW OF IMAGE STEGANOGRAPHY http://mo.co.za/open/stegoverview.pdf

[14] Johnson, N.F. & Jajodia, S., "Steganalysis of Images Created Using Current Steganography Software", *Proceedings of the 2nd Information Hiding Workshop*, April 1998.

[15] F. Huang, B. Li and J. Huang, "Attack LSB matching steganography by counting alteration rate of the number of neighbourhood gray levels," in *Proc. IEEE Int. Conf. Image Processing*, Oct. 16–19, 2007, vol. 1, pp. 401–404.

[16] J. Harmsen and W. Pearlman, "Steganalysis of additive-noise modelable information hiding," *Proc. SPIE Electronic Imaging*, vol. 5020, pp. 131–142, 2003.

[17] A.D. Ker, "Steganalysis of LSB matching in grayscale images," IEEE Signal Process. Lett., vol. 12, no. 6, pp. 441–444, Jun. 2005.

[18] J. Mielikainen, "LSB matching revisited," IEEE Signal Process. Lett., vol. 13, no. 5, pp. 285–287, May 2006.

[19] K. Hempstalk, "Hiding behind corners: Using edges in images for better steganography," in Proc. Computing Women's Congress, Hamilton, New Zealand, 2006.

[20] D. Wu and W. Tsai, "A steganographic method for images by pixelvalue differencing," Pattern Recognit. Lett., vol. 24, pp. 1613–1626, 2003.

ACKNOWLEDGEMENT

We take immense pleasure in thanking our chairman Dr. Jeppiaar M.A, B.L, Ph.D, the Directors of Jeppiaar Engineering College Mr. Marie Wilson, B.Tech, MBA, (Ph.D), Mrs. Regeena Wilson, B.Tech, MBA, (Ph.D) and the principal Dr. Sushil Lal Das M.Sc(Engg.), Ph.D for their continual support and guidance. We would like to extend our thanks to my guide, our friends and family members without whose inspiration and support our efforts would not have come to true. Above all, we would like to thank God for making all our efforts success.

P. Mohan Kumar B.E., M.E., (Ph.D) works as Associate Professor in Jeppiaar Engineering College and he has more than 8 years of teaching experience. His areas of specializations are Network security, Image processing and artificial intelligence.