American Journal of Applied Sciences 9 (5): 766-771, 2012
ISSN 1546-9239
© 2012 Science Publications

A Hybrid Approach for Detecting Stego Content in Corporate Mail Using Neural Network Based Simplified-Data Encryption Standard Algorithm

[1]Anitha, P.T., [2]M. Rajaram and [3]S.N. Sivanandham
[1]Karpagam College of Engineering, Coimbatore 641032, Tamilnadu, India
[2]Anna University of Technology, Tirunelveli, Tamilnadu, India
[3]Karpagam Group of Institutions, Coimbatore 641032, Tamilnadu, India

Abstract: Problem statement: The major growth of information technology is based on the security measures implemented. Steganography is a method which is used to give high level security. **Approach:** Today, email management and email authenticity must be unquestionable with strong chains of custody, constant availability and tamper-proof security. A secure communication can be achieved through neural based steganography. Email is insecure. **Results:** This research developed an application which can check the Email content of corporate mails by S-DES algorithm along with the neural networks back propagation approach. A new filtering algorithm is also developed which can used to extract only the JPG images from the corporate emails. Experimental research shows that this algorithm is more accurate and reliable than the conventional methods. **Conclusion:** We anticipate that this study can also give a clear picture of the current trends in Steganography and the experimental results indicate this method is valid in steganalysis. This method can be used for internet/network security.

Key words: Steganalysis, steganography, information hiding, LSB, stegdetect, stego, S-DES

INTRODUCTION

Steganography is the study of techniques for hiding in an innocuous carrier so that the existence of the message is concealed. Cryptography is different from Steganography. Cryptography is mainly concerned with obscuring the content of a message but not its existence. Nowadays Steganographic techniques are also used on digital contents. The important aspect of Steganography is to detect that there is hidden information. This reverse process of Steganography is called as Steganalysis.

Steganalysis specifically aimed at making the difference between genuine documents and steganographied (Maitra, 2011). Steganalysis is used to analyze data to determine the presence of hidden information in it. Steganalysis techniques can be used to detect, extract, change or ultimately destroy the hidden information and can be applied to suspect data for Steganography, watermarking, or authentication purposes. Steganalysis is of growing interest to communities in law enforcing and counter-espionage (Chandramouli, 2002).

Cryptography and Steganography have many applications in computer science and other related fields: they are used to protect military messages, E-mails, credit card information, corporate data, personal files. The widespread use of Steganography inevitably leads to a need to detect hidden data. Steganalysis is detecting and ultimately extracting data hidden in an innocuous medium. Hiding information in image, audio and video as well as other data types on digital computers, is increasingly common. Thus tools for determining whether files containing any hidden information are important for cyber forensics personnel. Law enforcement agencies at the local and national levels need good software programs that do reliable jobs of identifying suspicious files on computers and websites (Chandramouli, 2002).

The goal of steganalysis is to detect and/or estimate potentially hidden information from observed data with little or no knowledge about the steganography algorithm and/or its parameters.

The messages embedded into an image are often imperceptible to human eyes. The embedding algorithm is used to hide secret images inside a cover or a carrier document. This is protected by a keyword so that only who possess the secret keyword can access the hidden message. The most well known Steganography technique is LSB.

Corresponding Author: Anitha P.T., MCA, Karpagam College of Engineering, Coimbatore 641032, Tamilnadu, India

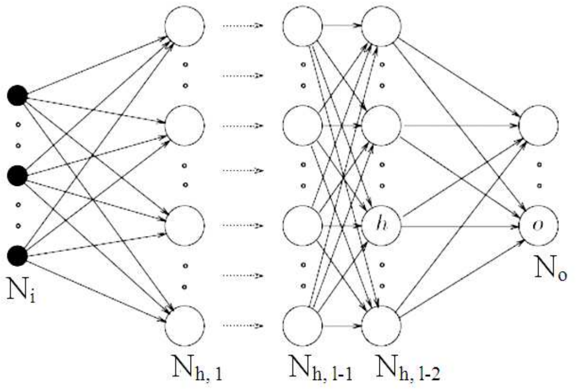

Fig. 1: Back propogation

Terminology: A message is the information hidden and may be plaintext, cipher text, images, or anything that can be embedded into a bit stream. Together the cover carrier and the embedded message create a stego-carrier. Hiding information may require a stego key which is additional secret information, such as a password, required for embedding the information. For example, when a secret message is hidden within a cover image, the resulting product is a stego-image (Geetha *et al*., 2009). A possible formula of the process may be represented as:

Stego-image = cover medium
+ secret key + Embedded message

Cryptanalyst applies cryptanalysis to decode or crack encrypted message whereas the steganalyst is one who applies steganalysis to detect the existence of the message. In Cryptography, the comparison is made between portions of the plain text and portions of the cipher text (Goth, 2005). But in steganography comparisons may be made between the cover image, embedded message and stego media.

Neural designs for steganography: Training the pattern that is formed from the original data and the selected bits from the graphical images produced the cipher text. This cipher text is embedded in the image and transmitted to the receiver. The strength of the algorithm lies in the design of the neural algorithm that trains the pattern. Through a secured channel the neural network algorithm has to be exchanged. A neural algorithm accepts “n” input pattern that has to be trained, equals 2^K+1, for an input pattern having k+1 bits. Out of k+1 bit one bit is selected from the secret message and remaining k bits from the carrier image. Input layer is designed with k+1 neurons. Depends on the problem domain the number of neurons are designed (Shaohui *et al*., 2003).

On the decryption side the input patters are formed by merging of cipher and selected bits from the image. The retrieval process returns back the secret message and the original cover image. In any performance evaluation work is the data set employed in the experiments. Our goal was to use a data set of images that would include a variety of textures, qualities and sizes. Obtaining images by using the file filtering scheme would provide us with such a data set. We chose the JPEG image format due to its wide use and popularity.

Image steganalysis: There are essentially three types of image formats: raw, uncompressed formats (BMP, PCX), palette formats (GIF) and lossy compressed formats (JPEG, Wavelet, JPEG2000). Only few current steganographic programs offer the capability to embed messages directly in the JPEG stream. It is a difficult problem to devise a steganographic method that would hide messages in the JPEG stream in a secure manner while keeping the capacity practical. Far more programs use the BMP, PCX, or the GIF palette-based format. The GIF format is a difficult environment for secure steganography with reasonable capacity.

Consequently, if the cover-image was initially stored in the JPEG format, the act of message embedding will not erase the characteristic structure created by the JPEG compression and one can still easily determine whether or not a given image has been stored as JPEG in the past. Actually, unless the image is too small, one can reliably recover even the values of the JPEG quantization table by carefully analyzing the values of DCT coefficients in all 8×8 blocks. After message embedding, however, the cover-image will become (with a high probability) incompatible with the JPEG format in the sense that it may be possible to prove that a particular 8×8 block of pixels could not have been produced by JPEG decompression of any block of quantized coefficients. This finding provides strong evidence that the block has been modified. It is highly suspicious to find an image stored in a lossless format that bears a strong fingerprint of JPEG compression, yet is not fully compatible with any JPEG compressed image. This can be interpreted as evidence for steganography. Presented in the Fig. 1 and 2 is an example of a hidden message inside a picture.

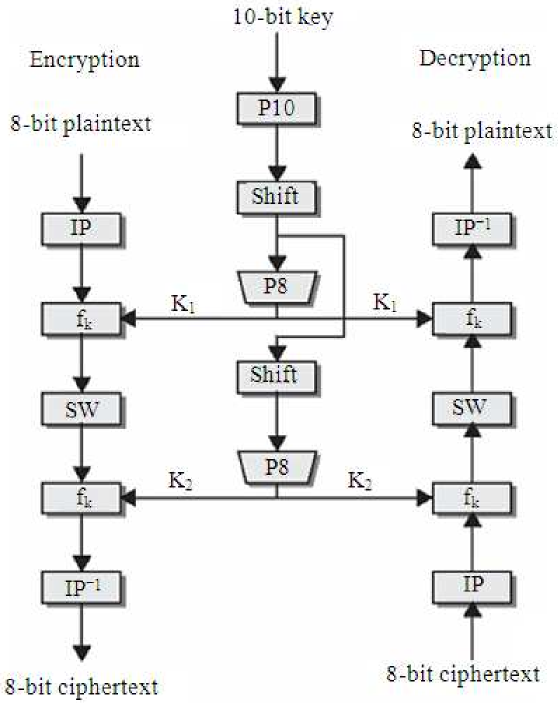

Fig. 2: Simplified DES scheme

MATERIALS AND METHODS

Proposed method: Our project was to write a program that could remove any hidden information from a JPEG image. The steganalysis we proposed is based on the observation of the filtered images of capturing algorithm we developed.

This research study developed a frame work which contains the following tasks: Image separation from corporate mails using the newly developed capturing algorithm, Compression, encryption, hiding, decryption and decompression steps.

Hybrid algorithm: A new hybrid algorithm is developed by combining the S DES algorithm and Back propagation algorithm of neural network which will effectively detect the stego content in the images (Provos and Honeyman, 2003). The S_DES is the best known and most widely used cryptosystem for civilian applications. It was developed at IBM and adopted by the National Bureau of Standards in the mid 1970s and has successfully withstood all the attacks published so far in the open literature.

DES method of encryption: DES works by encrypting groups of 64 message bits, which is the same as 16 hexadecimal numbers. To do the encryption, DES operates on the 64-bit blocks using key sizes of 56-bits. The keys are actually stored as being 64 bits long, but every 8th bit in the key is not used (i.e., bits numbered 8, 16, 24, 32, 40, 48, 56 and 64).

However, we will nevertheless number the bits from 1 to 64, going left to right, in the following calculations. But, as you will see, the eight bits just mentioned get eliminated when we create sub keys.

The 64-bit key is permuted according to the following table, PC-1. Since the first entry in the table is "57", this means that the 57th bit of the original key K becomes the first bit of the permuted key K+. The 49th bit of the original key becomes the second bit of the permuted key. The 4th bit of the original key is the last bit of the permuted key. Note only 56 bits of the original key appear in the permuted key.

PC-
157 49 41 33 25 17 9
1 58 50 42 34 26 18
10 2 59 51 43 35 27
19 11 3 60 52 44 36
63 55 47 39 31 23 15
7 62 54 46 38 30 22
14 6 61 53 45 37 29
21 13 5 28 20 12 4

We now proceed through 16 iterations, for 1<=n<=16, using a function f which operates on two blocks--a data block of 32 bits and a key K_n of 48 bits--to produce a block of 32 bits. Let + denote XOR addition, (bit-by-bit addition modulo 2). Then for n going from 1 to 16 we calculate:

$$L_n=R_{n-1};\ R_n = L_{n-1} + f(R_{n-1},K_n)$$

The DES algorithm turns a 64-bit message blocks M into a 64-bit cipher block C. If each 64-bit block is encrypted individually, then the mode of encryption is called Electronic Code Book (ECB) mode. There are two other modes of DES encryption, namely Chain Block Coding (CBC) and Cipher Feedback (CFB), which make each cipher block dependent on all the previous messages blocks through an initial XOR operation.

The capturing algorithm: All JPEG files in the e-mail inbox are detected and filtered. This filtering concept helps us to minimize the seeking time of filtering the JPEG files. After filtering those files they are stored in a large database for further processing. A sample image is taken from the database as covert channel which is used to hide the secret information. For our experiments, we created a database containing more than 20000 JPG images obtained from corporate mails. For each image, we embedded a random binary stream of different lengths using S-DES algorithm. The proposed research analyzes the performance of the improved version of image steganalysis algorithms in corporate mails. A large database is used to store the images. The performance and the detection ratio are going to be measured in corporate mails.

Image to text encryption: In this approach, each byte (pixel) of all the three matrices (R, G, B matrices of payload) are encrypted using S-DES algorithm and an image comprised of encrypted pixels is formed. The key used to encrypt each pixel is of 10-bit length and is obtained from the pixels of key image. The pixel values of red, green and blue intensities of each pixel of key image are combined to get a 24-bit value. The first ten bits are selected as the key to encrypt the red intensity pixel of payload image. The middle ten bits will be the key to encrypt the green intensity pixel of payload and finally the last ten bits is the key to encrypt blue intensity pixel of payload image. So the size of key image must be same as that of payload. If not, then the key image will get resized. Each pixel (24-bit) of the key image is split into three keys (10-bit each).This encrypted data is represented as an image which is hidden in another image called carrier image using Steganography.

Back propagation method: The neural network back propagation approach is used to check for the discrepancy patterns and train itself for better accuracy by automating the whole process (Provos and Honeyman, 2007). This study used neural network to analyze object digital image based on three different types of transformation which are Domain Frequency Transform (DFT), Domain Coefficient Transform (DCT) and Domain Wavelet Transform (DWT) (Goth, 2005).

In this study, we only consider following transforms, DFT, DCT and DWT. Firstly we analysis object digital image according these three different kinds transforms in this method. The object image is transformed into transform domain data according these three transforms. Then calculate these transforms data's statistical features which can be exploited to detect hided information. The reason for selecting DFT, DCT and DWT is that most data hiding method operate in these domains. These selected features should be significantly impacted by the data hiding processing. But it is difficult to find those features, so we select neural network to process this problem, neural network has the super capability to approximation any nonlinear functions.

Fig. 3: Steganography based document

For these features which have more effected by data hiding process, neural network will assign larger weight coefficients and for these features which have less effected by data hiding process, neural network will assign less weight coefficients.

Let us denote the i-th DCT coefficient of the k-th block as d_k (i), $0 \le i \le 64$, k = 1, T, where T is the total number of blocks in the image. In each block, all 64 coefficients are further quantized to integers D_k(i) using the JPEG quantization matrix Q .

The quantized coefficients D_k(i) are arranged in a zig-zag manner and compressed using the Huffman coder. The resulting compressed stream together with a header forms the final JPEG file.

The decompression works in the opposite order. The JPEG bit-stream is decompressed using the Huffman coder and the quantized DCT coefficients D_k(i) are multiplied by Q(i) to obtain DCT coefficients QD_k, $QD_k(i) = Q(i)D_k(i)$ for all k and i. Then, the inverse DCT is applied to QD_k and the result is rounded to integers in the range 0-255.

To be more secure, the cipher text obtained can be hidden in another image instead of sending it along the channel directly. The payload image is encrypted with the same S-DES algorithm and hidden in the key image (Avcibas *et al*., 2003). The secret key used in S-DES algorithm is a plain text. Figure 3 shows the payload image and the image obtained by applying S-DES algorithm to the payload. This stego image is sent along the channel. There was a slight difference in the histogram of key image and stego image, but this difference is invisible to human eye. We were able to get back the payload image successfully using the decryption key and the decrypted payload was matching with the input payload without any error in any pixel value. The image obtained at receiving end.

From the measured statistics of training sets of images with and without hidden information, our destination is to determine whether an image has been hidden information or not. Artificial Neural Network has the ability to adapt, learn, generalize, cluster or organize data.

In this study, we will deal with Back Propagation Artificial Neural Network Neural network has an excellent capability to simulate any nonlinear relation, so we make use of neural network to classify images (Narayana and Prasad, 2010). In this study we take use of BP neural network to train and simulate images. This BP neural network uses three levels: Input level, Hidden level and Output level. In neural network, the important issue is the slow of convergence.

A typical Back Propagation ANN is as depicted below. The black nodes (on the extreme left) are the initial inputs. Training such a network involves two phases. In the first phase, the inputs are propagated forward to compute the outputs for each output node. Then, each of these outputs is subtracted from its desired output, causing an error (an error for each output node).

In the second phase, each of these output errors is passed backward and the weights are fixed. These two phases is continued until the sum of [square of output errors] reaches an acceptable value.

Training the network can be summarized as follows:

- Apply input to the network.
- Calculate the output.
- Compare the resulting output with the desired output for the given input. This is called the error.
- Modify the weights for all neurons using the error.
- Repeat the process until the error reaches an acceptable value (e.g., error < 1%), which means that the NN was trained successfully, or if we reach a maximum count of iterations, which means that the NN training was not successful.

The program trains the network using JPEG images that are located in a folder. This folder must be in the following format:

- There must be one (input) folder that contains input images [*.jpg]
- Each image’s name is the target (or output) value for the network (the pixel values of the image are the inputs, of course)

RESULTS

Performance analysis and experiment results: The proposed method has been successfully implemented using MATLAB. Figure 3 represents the payload image that has to be concealed. The image pixels were encrypted using S-DES and converted to text form as described. The obtained cipher text is sent along the channel to the receiving end. Once the text is received at receiving end, it is then decrypted to get the image. For an intruder who attacks in the channel, the data looks like a plain text where the actual message passed is an image.

The cover image was taken from the image database. The image was originally in JPEG format in 680x480 resolutions. Since a BMP image was also required for the evaluation, a second image in BMP format was generated using the same JPEG image. Once both the cover images have been obtained, the proposed method generates the secret code for both the images were created. The encrypted image thus obtained was steganographically concealed in the carrier image.

The original cover file and its size before hiding the stgo content is:

100 KB (103,133 bytes)

The steganography document which is derived after hiding the secret information is:

107 KB (110,558 bytes)

The compression ratio and detection ratio of stego content is also analyzed. By analyzing the images in the sampled database the probability of occurrences of images with stego content in the corporate mails is zero.

DISCUSSION

Kekre *et al.* (2011) proposed a steganalysis of LSB encoding in color images. In which he introduced a powerful steganalysis technique that enables to detect the presence of pseudorandom binary images randomly spread in a color image. Shaohui *et al.* (2003) proposed a neural network based steganalysis in still images. In this he proposed a new method based on neural network to get statistical features of I mages to identify the underlying hidden data. Narayana and Prasad (2010) proposed two new mehods to detect the stego content by using the S-DES algorithm. Security is very important aspect in internet. Depending on the reliability of the steganalysis algorithms employed and the storage constraint one of two strategies, namely, coordinated search or random search can be chosen. It is seen that for a certain range of steganalysis reliability, both these methods give comparable performance.

CONCLUSION

Steganalysis is not as straight forward or convenient as steganography. It is interesting to detect the existence of hidden data resulting from any kind of embedding scheme known as the 'steganalysis'. From the above information that has been presented above we determine that when the cover image is in JPEG format, the detection results are very reliable and accurate. We demonstrated the performance of our detection method on a test database consisting of 50 grayscale images obtained from the email inbox. Today's truth the use of steganography and steganalysis is sure to increase and will be a growing hurdle for law enforcement and anti-terrorism activities.

REFERENCES

Avcibas, I., N. Memon and B. Sankur, 2003. Steganalysis using image quality metrics. IEEE Trans. Image Process., 12: 221-229. PMID: 18237902

Chandramouli, R., 2002. A Mathematical Approach to Steganalysis. Department of Electrical and Computer Engineering.

Goth, G., 2005. Steganalysis gets past the hype. IEEE Distribut. Syst. Online, 6: 1541-4922. DOI: 10.1109/MDSO.2005.22

Geetha, S., S.S.S. Sindhu and N. Kamaraj, 2009, Detection of stego anomalies in images exploiting the content independent statistical footprints of the steganograms. Inform. Slovenia, 33: 25-40. http://www.bibsonomy.org/bibtex/24570cb3c26c6ee8743097b064f378b87/dblp

Kekre, H.B., A.A. Athawale and S.A. Patki, 2011. Improved Steganalysis of LSB embedded color images based on stego-sensitive threshold close color pair signature. Int. J. Eng. Sci. Technol., 3: 836-842.

Maitra, I.K., 2011. Digital steganalysis: Review on recent approaches. J. Global Res. Comput. Sci.

Narayana, S. and G. Prasad, 2010. Two new approaches for secured image steganography using cryptographic techniques and type conversions. Department of Electronics and Communication, NITK, Surathkal, India.

Provos, N. and P. Honeyman, 2003. Hide and seek: An introduction to steganography. IEEE Security Privacy, 1: 32-44. DOI: 10.1109/MSECP.2003.1203220

Provos, N. and P. Honeyman, 2007. Detecting steganographic content on the internet. University of Michigan Ann Arbor, MI.

Shaohui, L., Y. Hongxun and G. Wen, 2003. Neural network based steganalysis in still images. Proceedings of the 2003 International Conference on Multimedia and Expo, Jul. 6-9, IEEE Xplore Press, pp: 509-512. DOI: 10.1109/ICME.2003.1221665