

AN EFFICIENT NEURAL NETWORK BASED ALGORITHM FOR DETECTING STEGANOGRAPHY CONTENT IN CORPORATE MAILS: A WEB BASED STEGANALYSIS

Ms. P. T. Anitha¹, Dr. M. Rajaram², Dr. S. N. Sivanandham³

¹Asst. Prof. /MCA, Karpagam College of Engineering, Coimbatore, Tamilnadu, India

²Vice Chancellor, Anna University of Technology, Tirunelveli, Tamilnadu, India

³Educational Advisor, Karpagam Group of Institutions, Coimbatore, Tamilnadu, India

Abstract

Steganography refers to information or file that has been concealed inside a digital media. Steganalysis is used to detect and/or estimate potentially hidden information from observed data with a little or no knowledge about the steganography algorithm and its parameters. Current trend in steganalysis seems to suggest two extreme approaches (a) little or no statistical assumption about the image under investigation. Statistics are learnt using a large database of training image and (b) a parametric model is assumed for the image and its statistics are computed for steganalysis detection. This research developed a new hybrid approach which comprises of neural network and S-DES encryption scheme which is used to detect the stego content in corporate mails. In this research work we implemented the combination of Compression, Encryption, Steganography to enhance the security of the data sent and Steganalysis methods which will detect the stego content in corporate emails. This method will be used to enhance the security measures of corporate mails.

Keywords: Steganalysis, S-DES, LSB, Stego, Steganography, information hiding, neural network

1. INTRODUCTION

The approach for high level secured communication is cryptography, which deals with encryption and decryption. The main difference between cryptography and steganography is the suspicion factor. When we implemented both the cryptography and steganography together, one can achieve a high level security.

Steganography refers to the science of invisible communication. The goal of steganography is to secure communication from an eavesdropper; Steganographic techniques strive to hide the very existence of the message itself from an observer. The simplest image Steganographic techniques essentially embed the message in a subset of the LSB (Least Significant Bit)

[1]. Cryptography and steganography are well known and widely used techniques that manipulate message in order to cipher or hide their existence. They are used to protect military images, E-mails, Credit card information, corporate data, personal files and etc. Cryptography encrypts the actual message that is being sent. This uses mathematical schemas and algorithm to scramble data into unreadable text.

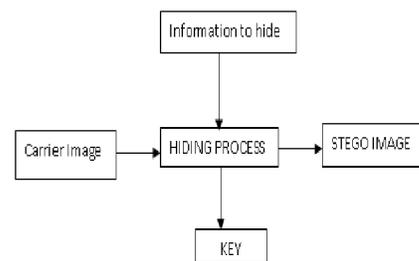


Figure 1: Process of Steganography

Steganalysis is the science of detecting hidden information. The goal of steganalysis is to break steganography. Steganalysis deals with three important attacks. (a) Visual attacks : one can identify the stego image with the naked eyes (b) Statistical attacks: they reveal the smallest alterations in an image statistical behaviour. It is further subdivided into (i) Passive attack: identifying the presence or absence of a covert messages or embedding algorithm used (ii) Active attacks: used to investigate embedded message length or hidden message location or secret key used in hidden process (c) Structural attacks: identifying the changes in the cover file.

2. PROPOSED METHOD

The advent of electronic mail and the internet has greatly enhanced communications throughout the world. The majority of internet users around the world use e-mail to

communicate. People can intercept email either maliciously or screen it.

This proposed method is used to detect the stego content in corporate emails. This method scans the corporate e-mails for the presence of stego content and it analyse the detection ratio of the stego contents.

2.1 EMAIL ATTACHMENT SCREENING

A new filtering algorithm is developed to screen the JPEG images from the Email attachments and this will store the same in the hard disk in a separate folder.

Steganography is used to hide the cipher text obtained in the above step into an image. The main purpose of steganography is encoding and decoding. The inputs of the steganography are Cover data, Plain text and the key value. An image file contains the binary representation of the colour or light intensity of each picture element comprising the image. Each pixel is represented by three bytes representing the intensity of three primary colours red, green, blue (RGB) respectively. A typical 640x480 pixel image using a palette of 256 colours would require a file about 307 KB in size whereas a 1024x768 pixel image would result in 2.36 MB file. JPEG uses lossy compression on the other hand. In this the expanded image is very nearly the same as the original but not an exact duplicate. Jpeg can be used for stego applications because it is more common to embed data in the GIF and BMP files.

With high resolution digital images as carriers, detecting the presence of hidden information has also become considerably more difficult. From the measured statistics of a training set of images with or without hidden information the goal is to determine whether an image contains a message. Some formats are based on lossy compression such as JPEG/MPEG and others use efficient indexing for lossless representation such as GIF.

2.2 S-DES with NEURAL NETWORK

Neural networks back propagation algorithm with a single hidden layer of processing elements can model any continuous function to any degree of accuracy. Three layers back propagation neural network was trained to classify each block in the difference image into stego or non stego block.

The following figure shows the single layer back propagation network.

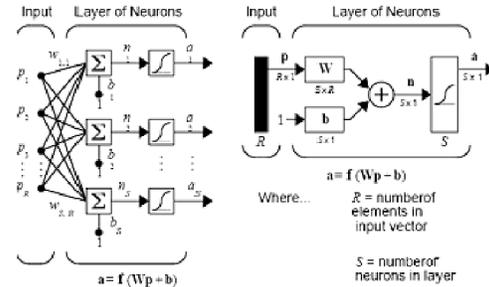


Figure 2: Single layer Back propagation neural network

Information in stego image can be retrieved from stego image can be retrieved by S-DES algorithm [2]. With our method, the use of neural network is the key technique. In this method the cover image is divided into group of blocks and all these blocks are generated by S-DES method. Now the embedder adjust a neural network weights with desired hidden bit code from the collection of both cover image and stego-image folder. This is common for both embedding and extracting the hidden information. We can use supervised learning of the neural network for learning. We embed the secret message within a cover image by using XOR neural network learning model.

The JPEG representation allows compression of the raster data to varying degrees. The original image cannot be exactly recreated since the algorithm losses some of the information. The DES algorithm extracting the JPEG image by converting the pixels into 8x8 blocks. Each block is compressed using Huffman encoding. The hidden message is compressed first and then encrypted before it is hidden. This will reduce the amount of hidden information.

The neural networks back propagation algorithm is used to indirectly hide the data into a graphical image, as it adds additional complexity for the hackers. Our approach hides indirectly the secured binary bits along with some selected graphical image bits, based on the neural network algorithm, to get cipher bits. The generated cipher bits are then placed in least significant bit (LSB) position of the transmitted graphical image [3]. In reverse process, this method regenerates the original data bits. The neural model used here is the multi-layer feed forward network.

3. ENCRYPTION and DECRYPTION using S-DES

ENCRYPTION

The S-DES algorithm encrypts the group of 64 bits of message which is the same as 16 hexadecimal numbers. For this S-DES uses 'keys' which is of 64 bits long.

Every 8th bit of the key is ignored by this algorithm and finally the key size is 56 bits. DES has 19 steps for encrypting the plain text. They are as follows:-

1. Initial permutation on 64 bit plain text.
2. 16 different iteration are performed which uses 16 different keys. The key size is 56 bits. Remaining bits are used for parity checking. Permutation is performed on the 56 bits and 48 bits are taken out to be used as key.
3. To perform iteration, 64 bit input is divided into 2 equal portion denoted by $L(i-1)$ and $r(i-1)$. For the value of i ranges from 1 to 16, the function f operates on two blocks. A data block of 32 bits and a key K_i of 48 bits to produce a block of 32 bits. Let the symbol \oplus denote XOR addition. Then for the value of i going from 1 to 16 we calculate

$$L_i = R_{i-1} \text{ and } R_i = L_{i-1} \oplus f(R_{i-1}, K_i)$$

The output generates two blocks $L(i)$ and $R(i)$ each of 32 bits long. The left part is simply the right part of the input. The right part of output is bitwise XOR of left part and function of right part of input and the key of the related iterations[1]

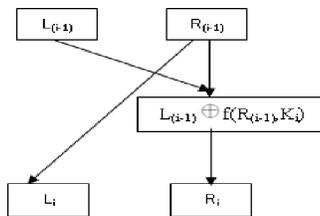


Figure 3: Sub blocks process

The 48 bit R_{i-1} is XORed with $K_{(i)}$ and stored in a temporary buffer so that R_{i-1} is not modified. This result is now split into 8 segments of 6 bits each. This forms the index into the S-boxes i.e. Substitution boxes are a set of 8 two dimensional arrays, each with 4 rows and 16 columns. The numbers in the boxes are always 4 bits in length so their value ranges from 0-15[5]. The S-boxes are numbered $S(1)$ to $S(8)$.

We now calculate $S1(B1)S2(B2)S3(B3) S4(B4) S5(B5) S6(B6) S6(B7) S8(B8)$

Where $S_i B_i$ refers to the output of the i^{th} S-box. For the first round we obtain as the output of the eight S-boxes are :

011000 010001 011110 111010 100001 100110 010100
 100111

4. The result from the previous stage is now passed into the P permutation.

$F = P(S1(B1)S2(B2)...S8(B8))$ The permutation P yields a 32 bit output from 32-bit input by permuting the bits of the input block. This is repeated 16 times.

5. After 16 iterations the 32 bits of left and 32 bits of right are swapped.
6. Then the reverse permutation is applied. Now the cipher text is obtained.

DECRYPTION

In this the Stego data is the input. The decryption process decodes the stego data. Final result is the cipher information. This is achieved by the S-DES algorithm, Then the cipher text is decrypted by using the same algorithm[4]. The extraction process starts from reading of stego-image and extraction of key information. The total length of the message which will embed with a stego-image. Only by knowing the proper network weights, extractor can induce the structure of the network and only proper network weights are able to output the proper hidden information.

4. DATA SET AND NETWORK CONFIGURATION

A data set contained 20 stego and non stego information from different images has been used to train and test the neural network. The experiments were performed using JAVA programming on Pentium IV, 2.8 GHz with 512 MB memory. To increase the classification capacity, the neural network had three layers, first layer, the input layer contains 60 neurons, the second layer, the hidden layer contains 20 neurons, and the third layer, output layer contains 1 neuron. The sum squared error goal was 0.05. This is performed on block by block basis. Ten images were used to train the neural network and ten images have been used for testing purposes. The following tables show the experimental results for different image in the training set and the results for test data.

IMAGE	TOTAL NUMBER	Correct Detection	DETECTION %
Hidden Images	42	35	83.33
No Hidden Images	40	32	80

Table 1: Test Results

5. IMPLEMENTATION RESULT

The proposed method is implemented successfully using JAVA. The figure 4.1 represents the payload image that has to be concealed. The image pixels were encrypted using S-DES and it is converted to text form. The resultant cipher text is sent along the channel to the receiving end. The cipher text is obtained by applying S-DES algorithm to payload image. Once the text is received it is then decrypted to get the image. For an intruder who attacked to retrieve the secret information, the data looks like a plain text. When the cover image and payload image were compared no pixel differences were found.

4. DISCUSSION AND CONCLUSIONS

From this research work we notice that the accuracy of the detection algorithm doesn't depend on the power of the network only but it also depends on the properties of the image and the embedding messages. This paper introduced the concept of combination of cryptography and steganography. It also proposed a new algorithm to overcome steganalysis. This method provided a higher similarity between cover image and stego image that yields a better impeachability. Further works are necessary to increase the accuracy if the positive detection of Steganographic content and decrease the false ones. From our experiments the proposed method is promising. The back propagation method is very slow when it is used for training images. Because of the lack of knowledge about embedding, it is very hard to select the suitable features that strongly discriminating stego from non stego images.

References

- [1] Dhawal Seth, L. Ramanathan, Abhishek Pandey, Vellore Institute of Technology Vellore, Tamil Nadu, India. "Security Enhancement: Combining Cryptography and Steganography", International Journal of Computer Applications (0975 – 8887) Volume 9– No.11, November 2010.
- [2] Imran Khan, "An Efficient Neural Network based Algorithm of Steganography for image", International Journal of Computer Technology and Electronics Engineering (IJCTEE) Volume 1 , Issue 2(63-67)
- [3] H S Majunatha Reddy, & K B Raja, "High Capacity and Security Steganography using Discrete Wavelet Transform", International Journal of Computer Science and Security (IJCSS), Volume (3): Issue (6) 462 (462-472).
- [4] Jessica Fridrich, Miroslav Goljan and Rui Dub, "Steganalysis based on JPEG compatibility," Special session on Theoretical and Practical Issues in Digital Watermarking and Data Hiding, SPIE Multimedia Systems and Applications IV, Denver, CO, August 20-24, 2001, pp. 275-280.
- [5] Dhawal Seth, L. Ramanathan, and Abhishek Pandey, "Security Enhancement: Combining Cryptography and

Steganography", International Journal of Computer Applications (0975 – 8887), Volume 9– No.11, November 2010(3-6).

First Author



Ms. P T Anitha received B.Sc. Computer Applications and Master of Computer Applications degree from Bharathiar University in 1993 and 1996 respectively. At present working as an Assistant Professor in the department of MCA, Karpagam College of Engineering, Coimbatore. I am Pursuing Doctorate degree in computer Science under the guidance of Dr. M. Rajaram, Vice Chancellor of Anna University of Technology, Tirunelveli, Tamilnadu, India. My area of research is Steganalysis. Four papers are published in International conferences, 2 papers in International Journals and 11 in national conferences. My area of research is Steganalysis. Currently I am working to improve the performance of the steganalysis algorithms used in corporate E-mails.

Second Author



Dr. M. Rajaram, M.E., Ph.D., is a Professor and Head in Electrical and Electronics Engineering and Computer Science and Engineering in Government College of Engineering, Tirunelveli. He received B.E Degree in Electrical and Electronics Engineering from Madurai University, M.E and PhD degree from Bharathiyar University, Coimbatore, in 1981, 1988 and 1994 years and his research interests are Computer Science and engineering, electrical engineering and Power Electronics. He is the author of over 120 Publications in various International and National Journals. 7 PhD scholars and 10 M.S (By Research) Scholars have been awarded under his supervision. At present, he is supervising 12 PhD Scholars. Further Dr. Rajaram has become the Vice-Chancellor of Anna University of Technology, Tirunelveli, Tamilnadu, India.

Third Author



Dr. S. N. Sivanadam completed his B.E (Electrical and Electronics Engineering) in 1964 from Government College of Technology, Coimbatore and M. Sc. (Engineering) in power system in 1966 from PSG College of Technology, Coimbatore (University Second Rank). He acquired Ph.D. in Control Systems in 1982 from Madras University. He has received the Best Teacher Award in the year 2001 and the Dhakshina Murthy Award for teaching Excellence from PSG College of Technology. He received the CITATION for best Teaching and Technical contribution in the year 2002, Government College of Technology, Coimbatore. He has teaching experience (UG and PG) of over 44 years. The total number of undergraduate and postgraduate projects guided by him for both Computer Science and

Engineering and Electrical and Electronics Engineering is around 950. Formerly he was a Professor and Head for the departments EEE and CSE, PSG College of technology, Coimbatore. Further he was a coordinator for seven government funded projects. Dr. Sivanandam has co-authored 14 books. He has delivered around 100 special lectures of different specializations in Summer/Winter schools and also in various Engineering Colleges. He has guided 32 Ph.D. research works and at present 10 Ph.D. research scholars are working under him. The total number of technical publications credited to him in various National and International journals and Conferences is around 750. He has chaired 12 International and 12 National Conferences. He is a member of various professional bodies like IE (India), ISTE, CSI, ACS, SSI and IEEE. He is a Technical Advisor to various reputed industries and reputed engineering Institutions. His research areas include Modelling and Simulation, Neural Networks, Fuzzy Systems and Genetic Algorithms, Pattern Recognition, Multi-dimensional System Analysis, Linear and Non-Linear Control Systems, Signal and Image Processing, Power Systems, Numerical Methods, Parallel algorithms, Data mining and Database Security.