Advanced Image Steganographic Algorithms and Breaking strategies

Pooja Rawat Amit Kumar Pandey Shivpratap singh Kushwaha
Dept. of CSE
Institute of Technology & Management
Gwalior_INDIA

ABSTRACT

We all are familiar about Steganography. We know that Steganography is a skill of hiding any information with invisible existence inside other objects, which are known as cover objects. These cover objects can be protocol, audio, video, text, image etc. Today HTML pages and spam emails are also being used for Steganography. With the help of Steganography we can hide the information which is paramount and confidential for us. In comparison to the bulk of papers that examines either Steganography or Steganalysis techniques; this paper gives deeply description of both image Steganography breaking strategies of them. It also discusses that which image format is most appropriate or best for our algorithm and how can we perform compression on that. The paper can therefore be divided into many parts and all parts of this paper will be very beneficial for readers'.

General Terms

Algorithms, Capacity, Security etc.

Keywords

Steganography, JPEG vs other formats, Image compression.

1. INTRODUCTION

Steganography is a part of Network Security and going together well with cryptography. Where cryptography hides only the contents of the message, it hides the existence of the message. It also has big distinction with watermarking. In watermarking we only ensure that nobody is able to change or delete the content of watermarked data, it does not work on the existence of that data. But Steganography primarily fights for existence. The word Steganography is basically obtained from Greek dictionary, in which Stegano means "keep secret" and Graphy means "making words or writing". So making words secretly or secret writing stands for Steganography. A Symbolical Steganographic system is described or drew using the Prisoner's problem And Principle of Kerckhoffs which is also used for Steganography. We are informing the readers of this paper that a harmless looking digital image can hide an implacable terrorist plan. The strength of image Steganography makes it reliable for terrorist organizations. So we must have knowledge about breaking strategies of any Steganographic algorithm.The study of various attacks on Steganographic algorithms for detecting original message is called Steganalysis. It is good to say that Steganalysis is both a skill and a technique. The skill of Steganalysis plays a vital role in selecting the essential features of hidden message for testing while the technique helps in discovering new detection methods. As more techniques of Steganography are developing, the interest in Steganalysis field is increasing speedily. Researchers and developers of this field are preceding parallel with Steganography and Steganalysis. We also have to consider that Steganalysis is becoming successful so stealthiness of our important documents is also a subject of concern [1, 2].

In this paper, we are giving a description of Steganography and Steganalysis which are based on images.The remaining paper is systematized as follows. In section I we have given introduction. Section II will describe suitable image format and compression of that format. Section III will discuss about the two most important domains of Steganography i.e. spatial domain and transform domain. At last in section IV we will conclude about whole literature in a concise manner.

2. SELECTION AND COMPRESSION OF PROPER IMAGE FORMAT

2.1 Selection

For a successful Image Steganographic system we must select an appropriate image format, first. Image file formats are standardized means of organizing and storing digital images. There are many formats of an image are available, for examples- JPEG, TIFF, BMP, GIF, RAW[3] etc. Most probably, we use JPEG (Joint Photographic expert Group) because it works best for with photographs and complex images and JPEG images are plentiful on internet and also because of small sizes of images. PNG is also a good choice for storing images with sharp transitions that do not transform well into transform domain, because it works on lossless compression. But here we will select only JPEG format because it provides big compression ratio which is prior for a Steganographic system.

2.2 Compression

After selecting an image, we have to learn about compression process [4]. Here we have selected JPEG, so we will perform compression on it. There are two models for JPEG compression encoder and decoder [5]. Steps for encoding are as follows-

a) Take a image of RGB representation.
b) Coordinate YUV color
c) Perform chrominance Downsampling
d) Divide into 8x8 blocks and Forward DCT.
e) Quantization
f) Zigzag ordering and differential coding
g) Huffman Encoding
h) Resulting Bit-stream

Decoding procedure is an inverse of encoding. Steps for decoding are as follows-

a) Resulting Bit-stream from encoding an image.
b) Dequantization
c) De-Zigzag ordering and De-DC coding
d) Huffman Decoding
e) Perform Chrominance Upsampling
f) 8x8 Inverse DCT
g) Coordinate YUV color
h) Decoded Image(RGB)

Image compression aiming at lessening the redundancy of the image and save or pass data in satisfactory form. The primary motto of these types of system is to reduce the space as much as possible and the image obtained after decoding can be similar to the original image most possible.

3.DOMAINS OF STEGANIOGRAPHY

Primarily, there are three types of Steganography: Spatial domain Steganography, Transform domain Steganography, Adaptive Steganography. In this section we will study about above three Steganography and their algorithms…..

3.1 Image Domain Steganography

We use Spatial Steganography when we do not want to apply any mathematical transformation before embedding. The techniques work on the principle of some special parameters like payload capacity, invisibility; file formats, Robustness against attacks etc. Spatial image Steganography is most carried on by the researchers because of its simplicity and easiness of mathematical analysis. Image domain includes direct insertion of bits and manipulation of noise and these are classified as simple systems. For Image domain Steganography we must use lossless image formats like- GIF (Graphical Interchange Format), 8-bit BMP (windows bitmap file format). Image domain Steganography encompasses many advanced Steganographic algorithms, which are explain below…

A. Direct Least Bit Substitution (DLSB)

We know that LSB substitution is one of the most common and ordinary technique of spatial domain. It can be like LSB in GIF, LSB in BMP and LSB in PNG [6].When we discuss about basic LSB substitution , in this LSB conveys about the smallest or can say right most bit of a given binary sequence for example, if we take a 8-bit binary number the LSB will be the 8th bit. “Hide & Seek” [7] is sequential and randomized algorithms which work on the principle of basic LSB substitution. Now what is DLSB???? It is an advanced version of LSB substitution. The embedding procedure of DLSB is based on the following equation-

$$Y_i = 2\lfloor (X_i/2) \rfloor + M_i$$

where,

$$\left\{ \begin{array}{l} Y_i = i^{th} \text{ pixel value after modification (Embedding)} \\ X_i = i^{th} \text{ pixel value before modification (Embedding)} \\ M_i = i^{th} \text{ Message bit (Given)} \end{array} \right.$$

The most advantageous character of DLSB is its simplicity. It affects the pixel values only by negative or positive values of 1 i.e. ±1.S-tool [8] is a Steganographic tool which is based on DLSB, which includes changing the LSB of each colour in a pixel of any image. Like S-tools, there are Steghide and Steganos, which are also based on DLSB. We can successfully steganalys DLSB by Chi-square attack, RS Steganalysis. We can also analyse DLSB with the help of “Pair of Values” in histograms.

B.OptimaPixel Adjustment Procedure(OPAP)

This algorithm is an enhanced or improved version of DLSB. DLSB is simple to apply but it is not a perfect technique because it introduces some distortions. So for reducing distortion, Chi-Kwon Chan and L.M Cheng introduced OPAP.It provides better stego-image [9] quality than DLSB. In OPAP first we hide the data then we apply OPAP on pixel values. We do not disturb the hidden data. Its procedure is as follows-

1. Primarily we take the least significant bits and substitute those bits with the data which is to be hidden.
2. Secondly we adjust those bits which are before the hidden bits. By adjusting those bits, we can minimize errors.
3. Take m bits to be substituted in each and every pixel.
4. D = pixel value (in decimal) after bit substitution.
5. P_i= pixel value (in decimal) of original image.
6. P_i’= pixel value (in decimal) of stego image.
7. Calculate difference between P_i and P_i’ i.e. (P_i-P_i’) denoted by P.
8. If $P \leq (2^m/2)$ then we will not adjust pixels anymore.
9. If (P_j<P_i’) then we calculate D as $D = D – 2^m$.
10. If (P_j>P_i’) then we calculate D as $D = D + 2^m$.
11. After this procedure, we will convert D in binary and substitute to pixels.

Attack by histogram and visual attack.

C.Pixel Indicator Technique (PIT)

Pixel indicator technique is also a modified version of DLSB and it works primarily on security level than on capacity level. Mostly, we apply Pixel indicator technique on RGB images. We know that PIT is an enhanced version of DLSB and an improvement over OPAP but it is also taking many advantages from previous Steganographic algorithms. In the working of PIT, we take two bits as least significant and these bits can be from Red, Green or Blue. By the selection of two bits from any colour channel, we can indicate the existence of hidden data in the remaining channels. Bits are chosen from R to B always like RGB, RBG, GBR, GRB, BRG, and BGR. In the process of Pixel Indicator Technique there are two sub processes, one is construction and other is recovery [10][11].there are two measures of Pixel Indicator technique, one is security and other is capacity like other algorithms. It works for the betterment of security but its security level is as same as OPAP i.e. medium. For capacity measurement we have to calculate number of bits per pixel which we can embed with minimum distortion or no distortion. In PIT number of embedded bits must not exceed 3.

By Histogram attack and visual attack.

D.Pixel Value Differencing (PVD*)*

Pixel value differencing much better than previous methods but its working is based on situation of pixels, whether pixels are on flat area or on edge. It is base on the concept that human vision is influenced by delicate changes in the flat areas than higher changes on edges. It provides more capacity for embedding with minimum noticeable changes. Primarily, Pixel value differencing algorithm was developed to hide data into 256 bit gray valued images. In the working of Pixel Value Differencing, we divide the cover image into blocks of consecutive pixels and these blocks must not be overlapped. Procedure of PVD is as follows-

a) First we take a cover image and divide it into blocks of consecutive pixels.
b) Assume gray value difference of these blocks in a range from -255 to +255.

c) For embedding message in these blocks gray value may be from 0 to 255.
d) If gray value range from 0 to 255 then we can embed bits directly and can produce stego image.
e) If gray value does not range from 0 to 255 then take some bits from secret message according to the range from -255 to +255.
f) Then we change the gray value difference with new gray value difference which is based on bit values to be embedded and lower bound of range -255 to +255.
g) Eventually perform inverse function on gray value difference of the two pixel s by using new gray value difference.
h) After these steps, we can directly embed bits and can produce stego image.

Basically the working process of PVD is divided into two sub process, one is embedding and other is extraction [12]. We can break it by the differences of pixel pairs and Chi-square attack.

E. Selected least Significant Bit Substitution (SLSB*)*

SLSB embeds bits into single colour channels of the pixels. It is a novel Steganographic algorithm improving the LSB algorithm. In spatial domain there are three types of algorithm: Non-filtering Algorithm, Randomized Algorithm, and Filtering Algorithm [13].SLSB is defined as filtering algorithm because it uses a default filter for filter cover image. This filter is applicable only on most significant bits and less significant bits are remained for embedding. This filtering process provides an area for embedding. In this algorithm, we hide information only in one colour and then perform a new process, which is called LSB Match Adaption. By LSB match Adaption, we find the difference between primary colour and stego colour. By this difference we can imply reduction in distortion for betterment of image quality. It is a modification over LSB but it is able to produce more desired results then LSB.As mentioned before it develops the new LSB match Adaptation method to reduce distortion. The working process of SLSB is as follows-

a) First we take information, which is to be hidden and compress this information with the help of compression techniques.
b) Then we take cover image; select a colour from cover image; perform pixel filtering by using default filter; embed compressed information into it.
c) After embedding perform bit replacement and LSB match Adaptation.
d) Eventually we get stego image.

Attack By statistical attack.

3.2 Transform Domain Steganography

Transform domain techniques are developed for hiding larger data than image domain with better security, better invisibility and for lossy compression. Unlike spatial domain in transform domain algorithms, we perform some mathematical transformations before embedding. The working principle of transform domain is based on Discrete Cosine Transform (DCT), Discrete Fourier Transform (DFT), Discrete Wavelet Transform (DWT) and Singular Value Decomposition (SVD) [14]. These are also called versions of transform Domain Steganography. In transform domain, transform coefficients are selected in a way such that hidden information is imperceptible to our visual system. It is also called JPEG Steganography because in this Steganography we use JPEG image format as it provides lossy compression. JPEG images may make transform domain more secure because ii makes it difficult to recognize the presence or real nature of image. There are many advanced transform domain Steganographic algorithms, in which some are as follows-

A.JSteg

JSteg is developed by D. Upham and is a Xerox of Hide & Seek algorithm. It performs embedding in a sequential form. It is approved as first Steganographic tool which can be used commercially. The algorithm chooses Discrete Cosine transform for embedding image blocks. The strengths of JSteg algorithm are embedding in sequential form and not any secret key. The process of JSteg algorithm is as follows-

a) First we take message which is to be hidden and cover image.
b) Divide cover image into image blocks and generate DCT coefficients for image blocks by performing Discrete Cosine Transformation.
c) Start from first DCT coefficient, if DCT ≠ 0 and DCT ≠ 1 then go to LSB of message and replace DCT LSB with message LSB.
d) When required replacement is achieved then stego image is produced.

Easily by Chi-square attack.

B. Outguess

It is a modification over JSteg algorithm and it is proposed by N. Proves. There are two versions of Outguess: Outguess 0.1 and Outguess 0.2.The working of Outguess is based on Pseudo Random Number Generator (PRNG) which is used to find the situation of embedding bits and their frequencies. The embedding process of Outguess 0.1 encompasses the principles of Hide & Seek algorithm (Randomized) and JSteg algorithm. The algorithm works as follows-

a) First we take a cover image, divide it into blocks and then convert the blocks into DCT coefficients.
b) Then by using Pseudo Random number Generator we shuffled the coefficients randomly.
c) Then we embed given information as same as in JSteg.
d) Then perform the inverse function on the shuffled coefficients.
e) Finally, image is converted into spatial domain and stego image is produced.

After Outguess 0.1, N.Proves developed a new version of Outguess, which is called Outguess 0.2.it is more secure and much qualitative approach of Outguess. Not affected by visual attack, histogram attack and Chi-square attack but steganalys by Blockiness.

C. F3 Algorithm

After Outguess 0.2, a more secure algorithm is developed by A.Westfeld, which is called F3. The concept behind it is that its embedding process is not as same as Outguess 0.1 and JSteg. It does not avoid embedding bits in DCT coefficients, which are equal to 1 but it avoids DC coefficients and DCT coefficients equal to zero.It does not support overlapping of bits. If LSBs of DCT coefficients does not match, then it decrements their values. After embedding, LSBs of non-zero coefficients match with the LSBs of given message. If in one time the embedding process is not supposed to be perfect then we perform re-embedding, this process is defined as shrinkage. F3 is more secure but it also has some weaknesses, which may be removed in next version.we can easily break F4 by converting stego image into quantized DCT coefficients.

D.F4 Algorithm

Re-Embedding is a weakness of F3 algorithm because more zeros are embedded than ones in a result of re-embedding and coefficients of JPEG images have odder values than even. These two points reduces the capability of F3 algorithm. For eliminating these, F4 is proposed, which is an enhanced version of F-Series.it provide a better working than F3 by also taking in account the negative coefficients. Like positive coefficients, are also of two types: negative and positive. The values are as follows: even-negative and odd-positive coefficients have values equal to 1; odd-positive and even-negative coefficients have values equal to 0.But this algorithm was not as much better as researchers were thinking.we can easily break F4 by converting stego image into quantized DCT coefficients.(little bit same as F3).

F.F5 Algorithm

it is developed by A.Westfeld in 2001[15]. As we have studied in previous algorithms, capacity and security both are opposite of each other. If any algorithm is providing desirable security then it does not provide desirable embedding capacity other and if any algorithm is providing required capacity then it does not provide better security. But character of F5 is opposite than all previously mentioned algorithms. It provides desirable capacity and desirable security in parallel. Like other algorithms of F-series, it does not support overlapping of blocks; it only increment or decrement the values of DCT coefficients as required. It introduces two new mechanisms: Matrix Encoding and Permutation Straddling. The working of F5 is as follows-

a) It takes cover image, quality factor, hidden message contained in a file, user password, PRNG for user password.
b) Find the RGB representation of cover image
c) Evaluate quantization table according to quality factor.
d) Perform compression and store DCT coefficients after quantization.
e) Calculate approx embedding capacity which is equal to hDCT-(hDCT/64) –h(0)-h(1)+0.49h(1)

WHRE ,

hDCT = Number of all DCT coefficients

h(0) = number of AC coefficients value equal to zero.

h(1) = number of AC coefficients value equal to 1

(hDCT/64) = number of DC coefficients

-h (1)+0.49h(1) = loss due to re-embeddig

Then uses PRNG for generate random order. In this step we ignore DC coefficients and coefficients equal to zero.

f) The secret message is partitioned into segment of n bits, and bits are embedded in a block of 2n-1 bits.
g) If message size is perfect for calculated embedding capacity, then embedding procedure is continue otherwise error occurs. We remove that and continue embedding.

it is protected from Blockiness, Chi-square attack, Histogram attack. But we can steganalys it by calculating original histogram of cover image from stego image.

4. FUTURE CHALLENGES

- Improving steganographic capacity.
- Improving steganographic security.
- Relate capacity and security mathematically.
- Decreasing embedding distortions.
- Reduce re-embedding theoretically.
- Work on selection image format.
- Enhance steganalys capability.
- Enhance F5 theoretically.
- Embedding in different image representation

5. CONCLUSION

Image Steganography is like a new era in the field of Data Hiding. There are many researchers, which exploring new things day by day and still there are lots of things to be explored. In this paper, we have presented the two main domains of Steganography i.e. image domain and transform domain. Both domains include various Steganographic algorithms. We have also presented them and how can we break them. In Future we can elaborate them more efficiently. We can mathematically present the evaluation parameters of algorithms.

6. REFERENCES

[1] T Morkel, J.H.P Eloff, M.S Olivier, "an overview of image Steganography". Proceedings of the Fifth Annual Information Security South Africa Conference (ISSA2005), 2005.

[2] Krenn, R., "Steganography and Steganalysis", **http://www.krenn.nl/univ/cry/steg/article.pdf**

[3] Paula Aguilera, "Comparison of different image compression formats". ECE 533 Project Report.

[4] Pao-Yen Lin, "Basic Image Compression Algorithm and Introduction to JPEG Standard".

[5] Wei-Yi Wei, "An Introduction to Image Compression.

[6] V.Lokeshwara Reddy, Dr. A.Subramanyam, Dr. P.Chenna Reddy, "Implementation of LSB Steganography and its Evaluation for Various File Formats". Int. J. Advanced Networking and Applications Volume: 02, Issue: 05, Pages: 868-872(2011).

[7] Philip Bateman, Dr. Hans Georg Schaathun, "Image Steganography and Steganalysis". University of Surrey, Department of Computing.

[8] Abbas Cheddad, Joan Condell, Kevin Curran and Paul Mc Kevitt , "Enhancing Steganography in Digital Images". Canadian Conference on Computer and Robot Vision. 978-0-7695-3153-3/08.2008 IEEE.

[9]Roy, Ratnakirti ; Changder, Suvamoy ; Sarkar, Anirban ;Debnath, Narayan C,"Evaluating Image Steganography techniques: future research challenges".Computing ,Management and Telecommunications (ComManTel), 2013 International Conference.

[10] Adnan Abdul-Aziz Gutub, "Pixel Indicator Technique for RGB Image Steganography". Journal of emerging

technologies in web intelligence, Volume 2, no. 1, February 2010.

[11] Adnan Gutub, Mahmoud Ankeer, Muhammad Abu-Ghalioun, Abdulrahman Shaheen, Aleem Alvi, "Pixel Indicator High Capacity Technique For RGB Image Based Steganography".

[12] Tasnuva Mahjabin, Syed Monowar Hossain, Md. Shariful Haque, "A Block Based data Hiding Method in Images using Pixel Value Differencing and LSB Substitution Method". 978-1-4673-4836-2/12.IEEE 2012.

[13] Juan Joś Roque, Jesús María Minguet, "SLSB: Improving the Steganographic Algorithm LSB".

[14] Yambem Jina Chanu, Kh. Manglem Singh, Themrichon Tuithung, "Image Steganography and Steganalysis: A Survey". International Journal of Computer Applications (0975-8887). Volume 52- No. 2, August 2012.

[15] Medha Kulkarni, Dr. J. W. Bakal, "Hide and Seek in JPEG Images" .International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 **www.ijera.com** Vol. 2, Issue 4, July-August 2012, pp.1634-163