Concealing Arcanum Message in Any File Format behind Images to Ensconce the Existence of the Message without Knowing the Metadata of File

Amit Upadhyaya[1] **Dr. Rajesh Kumar Pathak**[2] **Dimple Jayaswal**[3]

[1]Ph.D. (Pursuing) CMJ University, Shillong, Meghalaya

[2]Ph.D. GNIET Engineering College Greater Noida, Utter Pradesh

[3]M.Tech. (Pursuing) RTU, Kota, Rajasthan

Abstract - ***Steganography is the science of hiding information within other information. For example, a watermark "hides" an image on a piece of paper. If you look at most paper currency at a low angle or if you hold it up to a bright light, you can see a ghostly image in the paper. When you look at the currency straight on in normal light, you cannot see the image. Because this example is so easy to understand, steganography is often called "watermarking."***

Today steganography is often used to hide copyright information in an image, movie and audio file. The information is carefully encrypted and hidden so you cannot easily find it. Later, if one think you have stolen his movie file, he can pull the hidden copyright information out of it to prove it is his not yours.

There are many techniques for hidings secrete file. To hide any file we must know the structure of the secrete file i.e. if we want to hide .txt file we must know the structure of .txt file. This becomes very difficult to know the structure of every file format. In my research I remove this headache. In this research we can hide any file format (.exe, .dll, .doc, .pdf, .mp3 etc.) without knowing the structure or metadata of file behind bitmap image.

Keywords— .Exe Hiding, Steganography, Bitmap, HVS.

--♦------------------------------------

INTRODUCTION

Steganography, the art of hiding messages inside other messages, has until recently been the poor cousin of cryptography. Now, it is gaining new popularity with the current industry demands for digital watermarking and fingerprinting of audio and video.

What are watermarking and fingerprinting? Through the use of advanced computer software, authors of images, music and software can place a hidden ``trademark'' in their product, allowing them to keep a check on piracy. This is commonly known as watermarking. Hiding serial numbers or a set of characteristics that distinguishes an object from a similar object is known as fingerprinting.

Steganography, from the Greek, means covered or secret writing, and is a long-practiced form of hiding information. Although related to cryptography, they are not the same. Steganography's intent is to hide the existence of the message, while cryptography scrambles (encrypt) a message so that it cannot be understood.

STEGANOGRAPHY UNDER VARIOUS MEDIA

Often, although it is not necessary, the hidden messages will be encrypted. This meets a requirement posed by the ``Kerckhoff principle'' in cryptography. [1] This principle states that the security of the system has to be based on the assumption that the enemy has full knowledge of the design and implementation details of the steganographic system. The only missing information for the enemy is a short, easily exchangeable random number sequence, the secret key. Without this secret key, the enemy should not have the chance to even suspect that on an observed communication channel, hidden communication is taking place.

STEGANOGRAPHY METHODS

STEGANOGRAPHY IN TEXT

The electronic marking techniques by Brassil et al. are to be applied to either an image representation of a document or to a document format file, such as PostScript or TEXTfiles. [9] The technique is applied by shifting code. The idea is that a codeword (such as a binary number, for example) is embedded in the document by altering particular textual features. By applying each bit of the codeword to a particular document feature, we can encode the codeword. It is the type of feature that identifies a particular encoding method. Brassil identifies three features

- Line-Shift Coding
- Word-Shift Coding
- Feature Coding

STEGANOGRAPHY IN IMAGES

Steganography is about exploiting the limited powers of the human visual system (HVS). Within reason, any plain text, cipher text, other images, or anything that can be embedded in a bit stream can be hidden in an image. Image steganography has come quite far in recent years with the development of fast, powerful graphical computers, and steganographic software is now readily available over the Internet for everyday users.

IMAGE ENCODING TECHNIQUES

Information can be hidden many different ways in images. Straight message insertion can be done, which will simply encode every bit of information in the image. More complex encoding can be done to embed the message only in ``noisy'' areas of the image, that will attract less attention. The message may also be scattered randomly throughout the cover image.

LEAST SIGNIFICANT BIT INSERTION:

The least significant bit insertion method is probably the most well known image steganography technique. It is a common, simple approach to embedding information in a graphical image file. Unfortunately, it is extremely vulnerable to attacks, such as image manipulation. A simple conversion from a GIF or BMP format to a lossy compression format such as JPEG can destroy the hidden information in the image.[5]

When applying LSB techniques to each byte of a 24-bit image, three bits can be encoded into each pixel. (As each pixel is represented by three bytes.) Any changes in the pixel bits will be indiscernible to the human eye. The main advantage of LSB insertion is that data can be hidden in the least and second to least bits and still the human eye would be unable to notice it.

REDUNDANT PATTERN ENCODING:

The idea behind Redundant Pattern Encoding is to paint a small message over an image many times. An advantage over this method is that it can withstand cropping. A disadvantage is that you can't paint large messages.

SPREAD SPECTRUM:

Spread Spectrum steganography scatters an encrypted message throughout an image (not just in the LSB). To decode the message, the recipient needs the algorithm, crypto-key and stego-key. This method is still vulnerable to destruction from compression and image processing.

THE RESEARCH

COMPUTER STEGANOGRAPHY IS BASED ON TWO PRINCIPLES:

The first one is that the files that contain digitized images can be altered to a certain extend without loosing their functionality unlike other types of data that have to be exact in order to function properly.

The other principle deals with the human inability to distinguish minor changes in image colour, which is especially easy to make use of in objects that contain redundant information, be it 16-bit sound, 8-bit or even better 24-bit image. Speaking of images, changing the value of the least significant bit of the pixel color won't result in any perceivable change of that color**.**

The research deals with hiding data in the BMP files using pretty good privacy. It accepts the secrete file name with extension, whose data is to be hidden, and the image file name with extension, which should be a BMP file, in which the data has to be hidden.

TECHNICAL:

The program works only on real data in the data file and leaves unused space untouched. This allows the data file to be converted into other picture and leave the data intact.

The program spreads the secret data evenly over the whole image file. With real quantitative data it is able to use more than one bit per data point. It has built-in limits how many bits it may use: with palletized SVGA BMP files 1 bit, with true grey-scale BMP files 2 bits and with true color BMP files 6 bits per pixel (2 bits per channel). The limits should guarantee that no change of the image file contents will be evident to the normal viewer.

256 colours bitmaps often have only 100 or less colors. The remaining palette entries are still present but unused. Conceal may use some of them as partners for used colors whose first assigned partner lacks enough similarity. For this purpose the color in question is copied to the unused entry, resulting in two (nearly) identical entries; only 'nearly' because partner colors always differ in the least significant bit of one channel. Graphic programs or converters that rearrange the palette will respect the small difference, but scanning programs may search for those nearly identical entries. Conceal uses unused entries only on rare occasions.

Finding a suitable partner for each of the entries of a 256 color palette needs a measure of similarity. As distances in the RGB color space do not correlate very well with the human perception, the RGB coordinates are converted into the CIE L*a*b* coordinates as their distances give a much better correlation. Conversion formulas for the proposed sRGB standard by HP/Microsoft are used, taking also in account the nonlinearity of the typical monitor.

WORKING:

The program assumes that all integers are stored in the big Endean format, which is true for all Intel processors.

The graphical user interface is provided for easy access. The first screen gives you the option to conceal or retrieve the data or else to test the image file. If the conceal or retrieve option is chosen the program asks for the image and secrete file and then does the steganographic operation. The test option gives the information about the given image file. The given information includes the type of file and the amount of data that can be stored in that file.

PROCEDURE OR ALGORITHM FOR STEGANOGRAPHY

Most steganographic algorithms use a combination of various techniques to perform the subtasks of hiding the secret message in a cover file. A steganography program needs to do the following things: Finding the appropriate bits in a cover file that could be used to conceal the secret message there.

1. CHANGE THE EXTENSION OF FILE AT RUN TIME OF PROGRAM:

This is the first step of my algorithm. It is simple but very important step of this algorithm. This is the back bone of this research. In this to conceal file of any format behind image file, we change the extension of that file at run time and convert that file to text file, so that the secrete file work as text file for rest of the program. When we retrieve file from image file we convert that intermediate text file into original file format. But in this we have to mention the file and image name with their extension.

2. FINDING THE APPROPRIATE BITS:

The appropriate bits should be the redundant bit. The step of finding redundant bits in the cover file is done implicitly by most programs as they assume that the least significant bits are redundant and can be replaced. This technique however does not use the fact that the sender can analyse the cover image before embedding data in it.

A technique that analyzes the cover image to determine which bits could potentially be replaced is BPCS Steganography[7]. BPCS steganography is the technique which not only uses the least significant bits for embedding data but to use all regions in an image that are not shape-informative. To determine these regions the image is split into the single bit planes and each of these bit planes is analyzed individually. For every 8x8 pixel block in every bit plane a test is performed that determines the complexity of the image information in this block. If this complexity is above a certain threshold, random looking data, i.e. an encrypted secret message can be embedded in this block without significantly altering the image.

3. CHOOSING THE APPROPRIATE BIT FROM COVER FILE:

After the redundant bits of a cover file have been determined it is necessary to choose a subset of them to actually implant data there, as the number of redundant bits will in general not be equal to the number of bits required to implant the secret message.

To implant the secret message let n be the no of bits that are needed. The simple approach to choose the cover bits is to use the first n redundant bits of the cover file and embed the message there. This technique can usually be detected by visual attacks.

When choosing the first n redundant bits as cover bits the redundant bits at the beginning of the cover file are more likely to be chosen than the bits at the end of the cover file. However, it desirable that each cover bit is chosen with the same probability because these results in an equal utilization of image regions as cover bits. A technique which accomplishes this goal has been proposed in A pseudo-random permutation that depends on the secret key is applied to the positions of the redundant cover bits. The bits that are used as cover bits are given by the first n positions of the redundant cover bits after the permutation has been applied. This method provides an equal spreading of the cover bits among the redundant bits of the cover file. Using this method makes it harder to detect implanted data with visual attacks and statistical attacks.

1. IMPLANTING DATA:

There are different techniques to actually implant the secret message in the cover bits that have been chosen from the cover file.

The most common used technique is to overwrite the cover bits with bits of the secret message. As the cover bits usually are the least significant bits this results in replacing the least significant bits of some pixels with bits of the encrypted message. The advantage of this technique is that it provides a rather big capacity as every cover bit can contain one bit of the secret message. But the disadvantage is that it can be detected with visual and statistical attacks.

A technique that can implant more secret bits than two for every change in the cover file is called matrix-encoding. A tupple (series) of n bits is selected from the cover bits to implant k bits of the secret message there by changing at most d bits. For example if k is 4, n is 6 and d is 2 we can embed four bits of the secret message in six cover bits by changing at maximum only two of the cover bits. This can be accomplished by encoding the value of the first secret bit as the parity, i.e. the result of the XOR operation of the first and the third cover bit and the value of the second secret bit as the parity of the second and the third cover bit. Using this technique you can embed 2.67 bits per change to the cover file on average. If higher values for n and k are chosen this rate can be further increased, which reduces detect ability. The drawback of increasing this rate is that the capacity of the cover file will become smaller.

I have used another technique for implanting data, which is used in the algorithm F5. When embedding data in a .bmp file instead of overwriting the least significant bits the absolute value of the DCT coefficients can be decremented. This approach has the advantage that it preserves the statistical properties of the bmp file. The drawback is that this method only works on bmp files because it assumes a certain statistical distribution of the cover data that is commonly found in bmp files.

ADVANTAGES AND DISADVANTAGES OF ALGORITHM

ADVANTAGES:

- This algorithm is made for all file formats such as MP3, DLL, EXE, PDF, DOC, AVI etc.
- Knowing Metadata of file is not required, which decreases the complexity of program.
- When embedding data in a bmp file instead of overwriting the least significant bits the absolute value of the DCT coefficients can be decremented. This approach preserves the statistical properties of the BMP file.
- Algorithm is more secure than traditional one's.
- Difficult to steganalysis.

DISADVANTAGES:

- Algorithm use only BMP files as cover media because it assumes a certain statistical distribution of the cover data that is commonly found in bmp files.
- Size of file is depending on BMP file size. Larger the BMP file size, larger file we can conceal.
- This research work uses windows platform only.
- We have to give file name and BMP image name with their extension.

CONCLUSION

However, steganography has its place in security. It is no way can replace cryptography, but is intended to supplement it. Its application in watermarking and Fingerprinting, for use in detection of unauthorized, illegally copied material is continually being realized and developed.

Also, in places where standard cryptography and encryption is outlawed, steganography can be used to convert data transmission. Steganography, formerly just an interest of the military, is now gaining popularity among the masses. Soon, any computer user will be able to put his own watermark on his artistic creations.

Although steganography is an interesting research field, it falls short of the mark. Because the secret message can be damaged with manipulation attacks, current steganographic algorithms are fairly useless for digital watermarking. Additionally, all known image based stegosystems can be detected with statistical analysis. I used previous researches as base to my research for developing a more advance algorithm and program which conceal any file format without knowing the metadata of file behind BMP image file. My research can be used as prototype or base for developing much more powerful algorithms and programs. Perhaps one day an undetectable or unalterable stegosystem will be developed. Or maybe the mythical “public-key” stegosystem will be devised.

REFERENCES

[1]. Neil F. Johnson, Sushil Jajodia, George Mason University, "Exploring Steganography: Seeing the Unseen", IEEE Computers, February 1998.

[2]. Ross J. Anderson, Fabien A.P. Petitcolas, "On the limits of steganography"

[3]. Eiji Kawaguchi, et al: A Model of Anonymous Covert Mailing System Using Steganographic Scheme, in INFORMATION MODELLING AND KNOWLEDGE BASES XIV, H. Yaakkola et al (Eds), IOS Press.

[4]. Kahn security table by David Kahn.

[5]. Neil F. Johnson. Steganography. Technical Report. November 1995

[6]. Bitmap concept from the program 'Spyder' by Lucas Natraj and Philip Tellis.

[7]. www.support.microsoft.com

[8]. Eiji Kawaguchi and Richard O. Eason in 1997.

[9]. Electronic Marking and Identification Techniques to Discourage Document Copying Jack T. Brassil, Senior Member, IEEE, Steven Low, Member, IEEE, Nicholas F. Maxemchuk, Fellow, IEEE, and Lawrence O’Gorman, Senior Member, IEEE