P-3517

COMPARISON OF VARIOUS STEGANALYSIS SCHEMES FOR MEDICAL IMAGING APPLICATION

By

RAJY XAVIER

Reg. No. 0920107017

of

KUMARAGURU COLLEGE OF TECHNOLOGY

(An Autonomous Institution affiliated to Anna University of Tech, Coimbatore)

COIMBATORE - 641006

A PROJECT REPORT

Submitted to the

FACULTY OF ELECTRONICS AND COMMUNICATION

In partial fulfillment of the requirements

for the award of the degree

of

MASTER OF ENGINEERING

IN

COMMUNICATION SYSTEMS

APRIL 2011

MAHATMA GANDHI CENTRAL LIBRARY
KCT
P 3517

BONAFIDE CERTIFICATE

Certified that this project report entitled "**COMPARISON OF STEGANALYSIS SCHEMES FOR MEDICAL IMAGING APPLICATION**" is the bonafide work of Rajy Xavier [Reg. no. 0920107017] who carried out the research under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form part of any other project or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

S. Nagarathinam 6/4/2011

Project Guide

Ms.S. Nagarathinam

6/4/11

Head of the Department

Dr. Rajeswari Mariappan

The candidate with university Register no. 0920107017 is examined by us in the project viva-voce examination held on ..21/04/2011....

D. M 21/4/11

Internal Examiner

21/4/11

External Examiner

GOVERNMENT COLLEGE OF TECHNOLOGY
COIMBATORE - 13
INDIA

IEEE

NCVIT 11

Dr./Prof./Mr./Ms.....RAJ.Y...XAVIER.....

of.....KUMARAGURU.....COLLEGE....OF.....TECHNOLOGY...COIMBATORE.....

has ~~participated~~/presented a paper titled.....ANALYSIS...OF...IMAGES...FOR...VARIOUS

.....SPECIFIC....AND....GENERIC.....STEGANALYSIS....SCHEMES.....

in the National Conference on Recent Advances in Computer Vision & Information Technology organized by the Department of Computer Science & Engineering and Information Technology on 7th March, 2011.

ORGANIZING SECRETARY | CHAIRPERSON | PATRON

SATHYABAMA UNIVERSITY

JEPPIAAR NAGAR, RAJIV GANDHI ROAD, CHENNAI - 600 119.

NATIONAL CONFERENCE ON
EMERGING TRENDS IN VLSI, EMBEDDED AND NANO TECHNOLOGIES.
" NC-EVENT 2011"

CERTIFICATE

This is to certify that Mr/Ms RAJY XAVIER of Kumaraguru College of Technology participated/presented a paper entitled comparision of Various Steganalysis Schemes for Medical Imagery application in the national conference on "NC-EVENT 2011" organised by the Department of Electronics & Communication Engineering held on January 27-28, 2011.

co-author(s):

MARIE JOHNSON
DIRECTOR

MARIAZEENA JOHNSON
DIRECTOR

KUMARAGURU COLLEGE OF TECHNOLOGY

(An Autonomous Institution Affiliated to Anna University of Technology, Coimbatore)

COIMBATORE - 641049, TAMIL NADU, INDIA.

CERTIFICATE

CITEL 2011

This is to certify that ~~Mr.~~/Ms. RAJY XAVIER, ME [COMMUNICATION SYSTEMS] KUMARAGURU COLLEGE OF TECHNOLOGY, COIMBATORE has ~~attended~~ / presented a paper titled ANALYSIS OF MEDICAL IMAGES FOR VARIOUS STEGANALYSIS SCHEMES in the 3rd National Conference on **COMMUNICATION, INFORMATION AND TELEMATICS (CITEL 2011)** on 3rd & 4th March 2011, organized by the Department of Electronics and Communication Engineering, Kumaraguru College of Technology, Coimbatore.

Dr. Rajeswari Mariappan	Dr. S. Ramachandran	Dr. J. Shanmugam
HOD - ECE	Principal	Director

ACKNOWLEDGEMENT

First I would like to express my praise and gratitude to the Lord, who has showered his grace and blessing enabling me to complete this project in an excellent manner. He has made all things beautiful in his time.

I express my sincere thanks to our beloved Director **Dr.J.Shanmugam**, Kumaraguru College of Technology , I thank for his kind support and for providing necessary facilities to carry out the work.

I express my sincere thanks to our beloved Principal **Dr.S.Ramachandran**, Kumaraguru College of Technology , who encouraged me in each and every steps of the project work.

I would like to express my sincere thanks and deep sense of gratitude to our HOD, **Dr.Rajeswari Mariappan**, Department of Electronics and Communication Engineering, for her valuable suggestions and encouragement which paved way for the successful completion of the project work. I also thank her for her kind support and for providing necessary facilities to carry out the work.

In particular, I wish to thank and everlasting gratitude to the project coordinator **D.Mohanageetha(Ph.D).**, Associate Professor, Department of Electronics and Communication Engineering for her expert counseling and guidance to make this project to a great deal of success.

I am greatly privileged to express my deep sense of gratitude to my guide **S.Nagarathinam M.E.** Assistant Professor (SRG), Department of ECE, Kumaraguru College of Technology throughout the course of this project work and I wish to convey my deep sense of gratitude to all the teaching and non-teaching of ECE Department for their help and cooperation.

Finally, I thank my parents and my family members for giving me the moral support and abundant blessings in all of my activities and my dear friends who helped me to endure my difficult times with their unfailing support and warm wishes.

ABSTRACT

Recently, there has been a lot of interest in the fields of Steganography and Steganalysis. Steganography involves hiding information in a cover (carrier) media to obtain the stego media, in such a way that the cover media is perceived not to have any embedded message for its unintended recipients. Steganalysis is the mechanism of detecting the presence of hidden information in the stego media and it can lead to the prevention of disastrous security incidents. In this paper, it is provided that, a critical review of the steganalysis algorithms available to analyze the characteristics of an image stego media vis-à-vis the corresponding cover media (without the hidden information) and understand the process of embedding the information and its detection. It is noteworthy that each of these cover media has different special attributes that are altered by a steganography algorithm in such a way that the changes are not perceivable for the unintended recipients; but, the changes are identifiable using appropriate steganalysis algorithms. This paper can also give a clear picture of the current trends in steganography so that it is possible to develop and improvise appropriate steganalysis algorithms.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

JPEG	-------	Joint Picture Experts Group
LSB	-------	Least Significant Bit
GIF	-------	Graphical Interface Format
BPCS	-------	Bit-Plane Complexity Segmentation
DCT	-------	Discrete Cosine Transform
BMP	-------	Bit Map Plane
IQM	-------	Image Quality Matrix

CHAPTER 1
INTRODUCTION

1.1 Steganography Overview

Steganography is a Greek word which means "covered writing" and can trace its origins as far back as 440 B.C.. In Histories written by Herodotus, he gives two examples of steganography. The first is of Demeratus, a Greek in the Persian court who sent warning of a forthcoming invasion by Xerxes by writing a message on a wooden pallet and then covering it in wax. The messenger was able to successfully smuggle the "blank" tablet to Sparta. A second example was that of Histiaeus who shaved the head of his most trusted slave and tattooed a message on his head. After the slave's hair grew in he was dispatched with the "hidden message".

As technology has evolved so has steganographic technique. Along with the printing press came the use of "invisible inks", usually crafted from organic materials such as milk, juices or urine. When heat is applied to the document the hidden writing becomes visible. Photography provided the opportunity to create microfilm(s) which could be smuggled in secret compartments in clothing and luggage. Microfilm was a popular medium during the Franco- Prussian War (1870 - 1871). By the turn of the century, photographic reductions made it possible to produce microdots, a picture that could be reduced to the size of a period. In the 21st century the governments began to use steganography to protect their currency from being counterfeited. They have employed special inks, dyes, embedded threads and micro strips which denote the face value of the bill. Steganography has seen its greatest growth and use with the growth of the Internet. The power of the Internet lies in its ability to transmit large quantities of data, very quickly, to a large audience.

1.1.1 Why do we need to hide information?

There are two major issues that drive the technology to hide information. In the first group are those who are trying to protect their intellectual property rights. With the high availability of information via the Internet it is becoming more difficult to protect intellectual property and enforce copyright laws. The use of digital watermarks provides a way to insert a copyright notice into a document or image. The watermark is often a small image or text that is

repeated frequently through out the document or image. A similar technique is to Embed a digital fingerprint or serial number. The advantage of a fingerprint is that it can be used to trace the copy back to the original and is a powerful tool for prosecuting copyright violators.

The second group of people who are interested in hiding information are those who wish to convey information in a covert manner and avoid observation by unintended recipients. In this case the hidden message is more significant than the "carrier" object that is used to transport it. Steganography is often compared to cryptography in its ability to restrict unauthorized access to information. Cryptography is used to encrypt or scramble the data in such a fashion that only the intended recipient can decrypt it. When transmitting an encrypted message it is obvious that some form of communication has occurred, even if the message cannot be read. Steganography is used to hide the very existence of the message.

1.1.2 How do we hide information in the electronic age?

At the most fundamental level computers use binary, a combination of zeros and ones to represent text and graphics. The American National Standard Code for Information Interchange (ASCII) is the de facto standard for representing text and certain control characters. ASCII uses one parity bit and seven data bits to represent each character in the English language. For example an uppercase "A" is represented by 1000001. A digital image is composed of picture elements or "pixels." Each pixel contains information as to the intensity of the three primary colors, red, green and blue. This information can be stored in a single byte (8 bits) or in three bytes (24 bits). For example, in an 8 bit image white is represented by the binary value of 11111111 and black is 00000000. Current information hiding techniques rely on the use of a cover object (image, document, sound file, etc.) sometimes known as a carrier. The secret message is then broken down to its individual bits by a steganographic tool (stego-tool) and Embedded in the cover object. Many tools will utilize a password or passphrase which is necessary to extract the hidden message and is referred to as a stego-key. The result of this process is known as the stego-object.

Where can information be hidden? Almost anywhere on the Internet! The standard protocol suite used on the Internet is the Transmission Control Protocol / Internet Protocol (TCP/IP). The headers used to transfer data between computers allow the use of flags and certain reserved fields. With the appropriate tool, information can be inserted into these fields. The

advantage of this technique is that headers are rarely read by humans and thus makes an ideal place to hide data. The disadvantage of this method is that firewalls can be configured to filter out packets that contain inappropriate data in the reserved fields, thus defeating the steganographic transmission. Another popular technique for hiding information is to include extra spaces in documents. These spaces may contain hidden characters. Again this is a simple technique for hiding information and consequently is easy to detect and defeat. By opening such a document in a word processor the unusual spacing becomes readily apparent. Reformatting the document can remove the hidden message. The use of audio files can provide a good carrier for hidden messages. By their very nature sound files tend to be large in size and thus do not attract attention. In particular MP3Stego, can be used to hide information and maintain nearly CD quality sound.

The most relevant cover objects in use today are digital images because of their potential payload (hidden information). A typical image with 640 x 480 pixels and 256 colors (8 bit) can hide approximately 300 Kilobytes of information. A high resolution image, 1024 x 768 pixels and 24 bit color could hide approximately 2.3 Megabytes worth of data. Due to the potential large size of such files compression algorithms are used to reduce the image to a suitable size for sending across the Internet. There is a wide variety of compression algorithms available, but the three most common are Windows Bitmap (BMP), Graphic Interchange Format (GIF) and Joint Photographic Experts Group (JPEG). When choosing a cover image for use in steganography the first two compression algorithms, BMP and GIF are preferred because they offer "lossless" compression. The compressed image is an exact representation of the original. The JPEG compression algorithm uses floating point calculations to translate the picture into an array of integers. This conversion process can result in rounding errors which may eliminate portions of the image which are not visible to the naked eye. Although this rarely causes a noticeable change to the image it can significantly alter or destroy any information that was hidden in the image.

Embedding data into an image can be accomplished by either of two techniques, Image Domain tools or Transform Domain tools. Image Domain tools, also known as Bit Wise Methods, manipulate the Least Significant Bit (LSB) of the cover image. In this method the leftmost bit of each pixel in the cover image is replaced with one bit from the secret message. Because the LSB can only contain zeros and ones, approximately half the time the bit does not need to be altered in order to embed the data from the secret message. In a low resolution (small

number of pixels) image with 8 bit color the effects of manipulating the LSB can cause noticeable shifts in colors. As the resolution and depth of color increase in an image the impact of manipulating the LSB becomes less noticeable. Thus high resolution images are preferred for use as cover images. One exception to this rule are gray scale images. A gray scale image uses 8 bits to define 256 shades of gray between white and black. In a gray scale image pallet each shade represents an increment (or decrement) of 1 bit from the previous shade. Thus when the LSB is manipulated it is less likely to create a "new" or previously unused shade within the pallet. Most of the stego-tools available today utilize bit wise methods for hiding information. Some of the more popular Image Domain tools include; Hide and Seek, Mandelsteg, Steganos, StegoDos, S-TOOLS, and White Noise Storm. Transform Domain tools utilize an algorithm such as the Discrete Cosine Transformation (DCT) or wavelet transformation to hide information in significant areas of the image. Stegotools which utilize one of the many transform domain techniques are more robust, have a higher resilience to attacks against the stego-image such as compression, cropping and image processing.

1.2 Hiding Information in JPEG Images

"The JPEG image format uses a discrete cosine transformation (DCT) to transform successive 8x8 pixel blocks of the image into 64 DCT coefficients each. The least significant bits of the quantized DCT coefficients are used as redundant bits into which the hidden message is embedded. In some image formats, e.g. GIF, the visual structure of an image exists to some degree in all bit-layers of the image. Steganographic systems that modify the least-significant bits of these formats are often susceptible to visual attacks. This is not true for the JPEG format. The modification of a single DCT coefficient affects all 64 image pixels. For that reason, there are no known visual attacks against the JPEG image format."

1.3 Steganalysis Overview

With careful selection of an appropriate cover image and a stego-tool it is possible to create a stego-image that does not appear to be different within the limits of human perception. However, electronically each of these tools leaves a fingerprint or signature in the image that can be used to alert an observer to the presence of a hidden message. Discovering a hidden message is the first step in steganalysis and is considered an "attack" on the hidden information. Attacks

may come in several different forms depending on what information is available to the steganalyst. There are two other types of attacks against steganography. The first is the known message attack. In this case the steganalyst (one who does steganalysis) has a known hidden message and the corresponding stego-image. In this case the objective is to determine patterns that result from hiding the message. These patterns can then be used to analyze other stegoobjects in the future. The second attack is the chosen-message attack. In this case the steganalyst will create a message and use a known stego-tool to create a stego-image. This known stego image is then analyzed to determine patterns for later use against other stegoimages.

Stego-only attack	Only the stego-object is available.
Chosen stego attack	The stego-tool (algorithm) is known and the stego-object is available.
Known cover attack	The stego-object and a known original copy of the cover object are available.
Known stego attack	The stego-tool (algorithm) is known and both the stego-object and original cover are available.

Table 1.1 Types of attacks

In order to be effective at steganalysis one must have good pattern recognition skills. In some instances comparing stego-images prepared with Image Domain tools and their original cover images will result in detectable visual noise. Noise is defined as a pixel that stands out from the other pixels in its area or "neighborhood." For example a lone red pixel on a white field. Another visual clue to the presence of hidden information is padding or cropping of an image. The Hide and Seek tool can only produce images of a fixed size; 320x200, 320x400, 320x480, 640x400 and 1024x768. If an image does not fit into one of these sizes it is cropped or padded with black spaces. StegoDos has a similar problem. The majority of stego-images does not reveal visual clues when compared with their cover image and thus require a more detailed analysis in order to determine that information has been concealed. In the work with current steganographic tools , researchers discovered several possible electronic signatures. The simplest signature is an increase in the file size between the stego-image and the cover image. Most of the other

signatures manifest themselves in some form of manipulating the color palette of the cover image. These fingerprints can include a large increase or decrease in the number of unique colors. Another fingerprint is colors in a palette which increase incrementally rather than randomly. The exception, of course, is gray scale images, which do increase incrementally. The presence of a disproportionate number of shades of black in a gray scale image is another strong indicator. Once a stego-image has been discovered there are several steps that can be taken to disable or destroy the hidden message. Stego-images created with an Image Domain tool can be rendered useless (the hidden message can not be recovered) by simply converting the image to a JPEG format . Images created with Transform Domain tools require a more aggressive approach in order to disable the hidden information. Although they can survive any single image manipulation, multiple manipulations on the same image have defeated all of the known tools. Image manipulation includes techniques such as: cropping, removing portions of the image; rotating the image; blurring, decreasing the contrast between pixels; sharpening, increasing the contrast between pixels (opposite of blurring); adding or removing noise; resampling; converting between bit densities (gray scale, 8 bit, 24 bit); converting from digital to analog to digital (print the image then rescan it); adding bit wise messages; adding transform message.

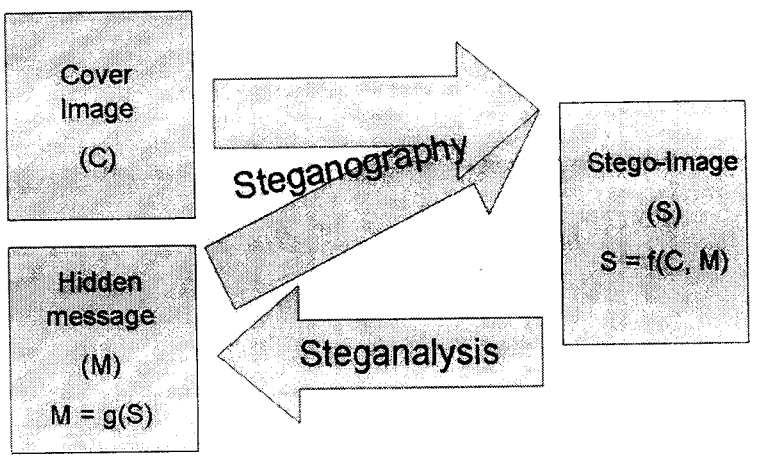


Figure 1.1 Basic Block Diagram of Steganalysis

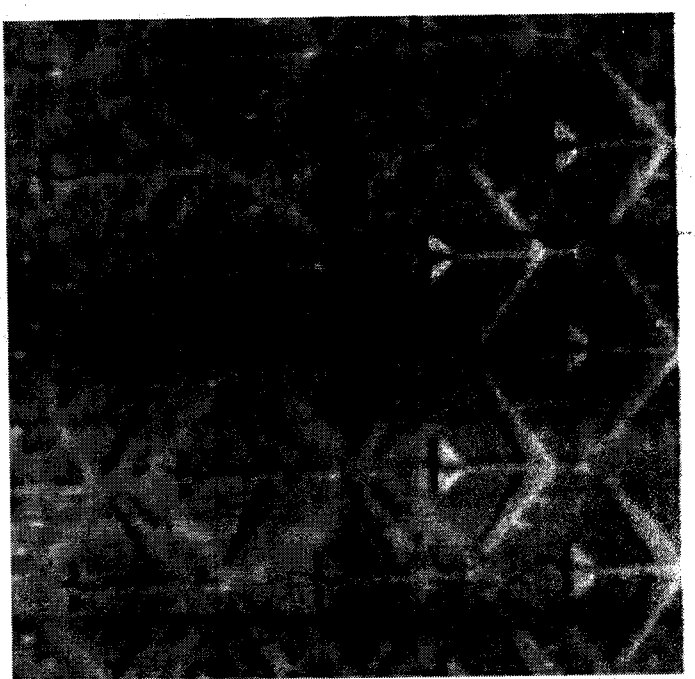
1.4 Current Events

If steganography is so easily detected and defeated who would use it? According to Ross Anderson, of Cambridge University, "There are about three or four generations of stego software. The stuff you can download is first generation and easily defeated." It is important to remember that steganography is only a single tool. Repeated use of the same tool will provide an unintended recipient with a large body of stego-objects which can facilitate the cracking of the stego-system, thus revealing all of your communications. There are several steps which can be taken to improve the security of your data, including encrypting messages before applying steganography and changing stego-tools periodically.

There is some concern that terrorists using steganography. In February 2001 Jack Kelly wrote two articles in USA TODAY which indicated the Osama Bin Laden and his organization, Al-Qaeda, as well as other known terrorist groups were using steganography to plan and implement terrorist acts. It was suggested that stego-images were being placed on auction sites such as e-Bay and Amazon as well as sports chat rooms and pornographic sites. Several other news agencies ran similar articles later in the month. None of the articles offered definitive proof, other than anonymous quotes from federal law enforcement agencies, indicating that stego-images had been found. There were several references to encrypted e-mail and files that had been recovered. Based on the allegations that terrorist organizations were using steganography, Niels Provos and Peter Honeyman, researchers at the University of Michigan, launched a project to determine the truth of the matter. In their technical report published on August 31, 2001, Provos and Honeyman outline the tools they used (Stegdetect, Stegbreak, Crawl and Disconcert) to launch an automated, statistical analysis of over 2 million JPEG images found on the e-Bay web site. As of this writing only one stego-image has been discovered. The image, sovereigntime.jpg (below) contained a gray scale image (below) of the "B-52 graveyard" at Davis-Monthan Air Force Base. These images were part of an ABC interview with an Internet security consultant who was demonstrating steganography. The authors conclude that based on their statistical analysis of images there is a small chance that they have not yet detected the stego-images that the terrorists are using. They believe it is more likely that as August 2001, there are no stego-images on the Internet.

sovereigntime.jpg

Figure 1.2 Image discovered by Provos and Honeyman



"B-52 graveyard" at Davis-Monthan Air Force Base

Figure 1.3 B-52 At Graveyard

1.5 Medical Image Steganography

In recent years, several architectures for secure storage and transmission of medical records have been proposed, both for real and semi-real time applications like blood glucose monitoring, secure telemedicine, and non-real time applications that involve maintaining and sharing medical records and databases. Most of these architectures, however, rely on some form of cryptography. Cryptographic techniques encrypt the medical records with a password and assume that only authorized parties have access to the password. While this does work most of the time, the encrypted data is prone to prying security thieves, who could decipher sensitive information like the patients' insurance service provider, medication history, etc.

Steganography provides an extremely effective alternative to this problem, hiding the very existence of sensitive data by concealing the data in a "carrier". In this paper, the case of cervical images and scanned medical documents as carriers for transferring covert information is considered. The objective of image steganography is to hide secret information in nondescript areas of the carrier image such that the changes made to the image are imperceptible, and the secret information itself is retrievable only by authorized, informed, personnel.

1.6 Technical Details

Since the rise of the Internet one of the most important factors of information technology and communication has been the security of information. Cryptography was created as a technique for securing the secrecy of communication and many different methods have been developed to encrypt and decrypt data in order to keep the message secret. Unfortunately it is sometimes not enough to keep the contents of a message secret, it may also be necessary to keep the existence of the message secret. The technique used to implement this, is called steganography.

The word steganography is derived from the Greek words "stegos" meaning "cover" and "grafia" meaning "writing"defining it as "covered writing". In image steganography the information is hidden exclusively in images.

The idea and practice of hiding information has a long history. In Histories the Greek historian Herodotus writes of a nobleman, Histaeus, who needed to communicate with his son-in-law in Greece. He shaved the head of one of his most trusted slaves and tattooed the message onto the slave's scalp. When the slave's hair grew back the slave was dispatched with the hidden

message. In the Second World War the Microdot technique was developed by the Germans. Information, especially photographs, was reduced in size until it was the size of a typed period. Extremely difficult to detect, a normal cover message was sent over an insecure channel with one of the periods on the paper containing hidden information. Today steganography is mostly used on computers with digital data being the carriers and networks being the high speed delivery channels.

Steganography differs from cryptography in the sense that where cryptography focuses on keeping the contents of a message secret, steganography focuses on keeping the existence of a message secret. Steganography and cryptography are both ways to protect information from unwanted parties but neither technology alone is perfect and can be compromised. Once the presence of hidden information is revealed or even suspected, the purpose of steganography is partly defeated. The strength of steganography can thus be amplified by combining it with cryptography.

Two other technologies that are closely related to steganography are watermarking and fingerprinting. These technologies are mainly concerned with the protection of intellectual property, thus the algorithms have different requirements than steganography. These requirements of a good steganographic algorithm will be discussed below. In watermarking all of the instances of an object are "marked" in the same way. The kind of information hidden in objects when using watermarking is usually a signature to signify origin or ownership for the purpose of copyright protection. With fingerprinting on the other hand, different, unique marks are embedded in distinct copies of the carrier object that are supplied to different customers. This enables the intellectual property owner to identify customers who break their licensing agreement by supplying the property to third parties. In watermarking and fingerprinting the fact that information is hidden inside the files may be public knowledge – sometimes it may even be visible – while in steganography the imperceptibility of the information is crucial. A successful attack on a steganographic system consists of an adversary observing that there is information hidden inside a file, while a successful attack on a watermarking or fingerprinting system would not be to detect the mark, but to remove it.

Research in steganography has mainly been driven by a lack of strength in cryptographic systems. Many governments have created laws to either limit the strength of a cryptographic system or to prohibit it altogether, forcing people to study other methods of secure information

transfer. Businesses have also started to realize the potential of steganography in communicating trade secrets or new product information. Avoiding communication through well-known channels greatly reduces the risk of information being leaked in transit. Hiding information in a photograph of the company picnic is less suspicious than communicating an encrypted file.

1.7 Overview

To provide an overview of steganography, terms and concepts should first be explained. An overview of the different kinds of steganography is given at a later stage.

1.7.1 Steganography concepts

Although steganography is an ancient subject, the modern formulation of it is often given in terms of the prisoner's problem proposed by Simmons, where two inmates wish to communicate in secret to hatch an escape plan. All of their communication passes through a warden who will throw them in solitary confinement should she suspect any covert communication.

The warden, who is free to examine all communication exchanged between the inmates, can either be passive or active. A passive warden simply examines the communication to try and determine if it potentially contains secret information. If she suspects a communication to contain hidden information, a passive warden takes note of the detected covert communication, reports this to some outside party and lets the message through without blocking it. An active warden, on the other hand, will try to alter the communication with the suspected hidden information deliberately, in order to remove the information.

MAHATMA GANDHI CENTRAL LIBRARY KCT CBE-6 P 3517

1.7.2 Different kinds of steganography

Almost all digital file formats can be used for steganography, but the formats that are more suitable are those with a high degree of redundancy. Redundancy can be defined as the bits of an object that provide accuracy far greater than necessary for the object's use and display. The redundant bits of an object are those bits that can be altered without the alteration being detected easily. Image and audio files especially comply with this requirement, while research has also uncovered other file formats that can be used for information hiding. Figure 1.4 shows the four main categories of file formats that can be used for steganography.

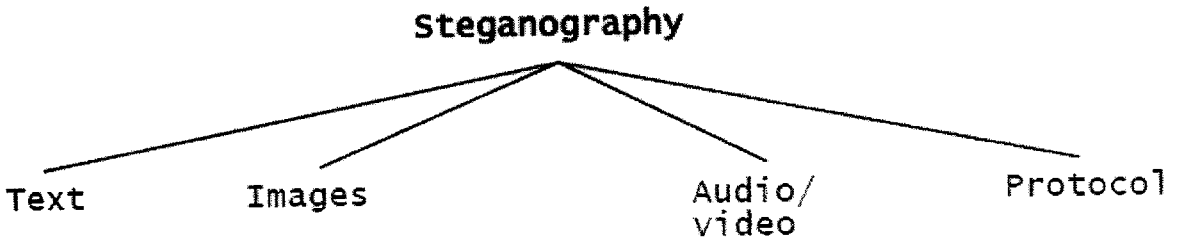


Figure 1.4 Categories of steganography

Hiding information in text is historically the most important method of steganography. An obvious method was to hide a secret message in every nth letter of every word of a text message. It is only since the beginning of the Text Images Audio/video Protocol Internet and all the different digital file formats that is has decreased in importance. Text steganography using digital files is not used very often since text files have a very small amount of redundant data.

Given the proliferation of digital images, especially on the Internet, and given the large amount of redundant bits present in the digital representation of an image, images are the most popular cover objects for steganography.

Here, we focus on hiding information in image. To hide information in audio files similar techniques are used as for image files. One different technique unique to audio steganography is masking, which exploits the properties of the human ear to hide information unnoticeably. A faint, but audible, sound becomes inaudible in the presence of another louder audible sound. This property creates a channel in which to hide information. Although nearly equal to images in steganographic potential, the larger size of meaningful audio files makes them less popular to use than images.

The term protocol steganography refers to the technique of embedding information within messages and network control protocols used in network transmission. In the layers of the OSI network model there exist covert channels where steganography can be used [12]. An example of where information can be hidden is in the header of a TCP/IP packet in some fields that are either optional or are never used. A paper by Ahsan and Kundur provides more information on this.

1.7.3 Image steganography

As stated earlier, images are the most popular cover objects used for steganography. In the domain of digital images many different image file formats exist, most of them for specific applications. For these different image file formats, different steganographic algorithms exist.

1.7.3.1 Image definition

To a computer, an image is a collection of numbers that constitute different light intensities in different areas of the image. This numeric representation forms a grid and the individual points are referred to as pixels. Most images on the Internet consists of a rectangular map of the image's pixels (represented as bits) where each pixel is located and its colour. These pixels are displayed horizontally row by row.

The number of bits in a colour scheme, called the bit depth, refers to the number of bits used for each pixel. The smallest bit depth in current colour schemes is 8, meaning that there are 8 bits used to describe the colour of each pixel. Monochrome and grayscale images use 8 bits for each pixel and are able to display 256 different colours or shades of gray. Digital colour images are typically stored in 24-bit files and use the RGB colour model, also known as true colour. All colour variations for the pixels of a 24-bit image are derived from three primary colours: red, green and blue, and each primary colour is represented by 8 bits. Thus in one given pixel, there can be 256 different quantities of red, green and blue, adding up to more than 16-million combinations, resulting in more than 16-million colours. Not surprisingly the larger amount of colours that can be displayed, the larger the file size.

1.7.3.2 Image Compression

When working with larger images of greater bit depth, the images tend to become too large to transmit over a standard Internet connection. In order to display an image in a reasonable amount of time, techniques must be incorporated to reduce the image's file size. These techniques make use of mathematical formulas to analyze and condense image data, resulting in smaller file sizes. This process is called compression.

In images there are two types of compression: lossy and lossless. Both methods save storage space, but the procedures that they implement differ. Lossy compression creates smaller files by discarding excess image data from the original image. It removes details that are too

small for the human eye to differentiate, resulting in close approximations of the original image, although not an exact duplicate. An example of an image format that uses this compression technique is JPEG (Joint Photographic Experts Group).

Lossless compression, on the other hand, never removes any information from the original image, but instead represents data in mathematical formulas. The original image's integrity is maintained and the decompressed image output is bit-by-bit identical to the original image input. The most popular image formats that use lossless compression is GIF (Graphical Interchange Format) and 8-bit BMP (a Microsoft Windows bitmap file).

Compression plays a very important role in choosing which steganographic algorithm to use. Lossy compression techniques result in smaller image file sizes, but it increases the possibility that the embedded message may be partly lost due to the fact that excess image data will be removed. Lossless compression though, keeps the original digital image intact without the chance of lost, although is does not compress the image to such a small file size. Different steganographic algorithms have been developed for both of these compression types and will be explained in the following sections.

1.7.3.3 Image and Transform Domain

Image steganography techniques can be divided into two groups: those in the Image Domain and those in the Transform Domain. Image – also known as spatial – domain techniques embed messages in the intensity of the pixels directly, while for transform – also known as frequency – domain, images are first transformed and then the message is embedded in the image.

Image domain techniques encompass bit-wise methods that apply bit insertion and noise manipulation and are sometimes characterized as "simple systems". The image formats that are most suitable for image domain steganography are lossless and the techniques are typically dependent on the image format.

Steganography in the transform domain involves the manipulation of algorithms and image transforms. These methods hide messages in more significant areas of the cover image, making it more robust. Many transform domain methods are independent of the image format and the embedded message may survive conversion between lossy and lossless compression.

In the next sections steganographic algorithms will be explained in categories according to image file formats and the domain in which they are performed.

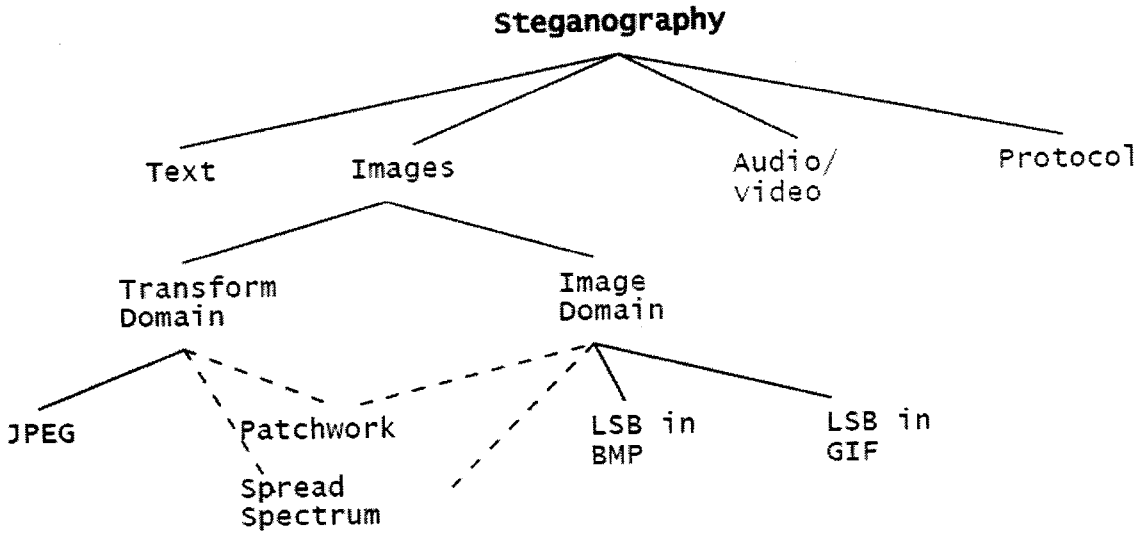


Figure 1.5 Categories of image steganography

1.7.3.3.1 Image Domain

Least Significant Bit

Least significant bit (LSB) insertion is a common, simple approach to embedding information in a cover image. The least significant bit (in other words, the 8th bit) of some or all of the bytes inside an image is changed to a bit of the secret message. When using a 24-bit image, a bit of each of the red, green and blue colour components can be used, since they are each represented by a byte. In other words, one can store 3 bits in each pixel. An 800 × 600 pixel image, can thus store a total amount of 1,440,000 bits or 180,000 bytes of embedded data. For example a grid for 3 pixels of a 24-bit image can be as follows:

(00101101 00011100 11011100)
(10100110 11000100 00001100)
(11010010 10101101 01100011)

When the number 200, which binary representation is 11001000, is embedded into the least significant bits of this part of the image, the resulting grid is as follows:

(00101101 00011101 11011100)
(10100110 11000101 00001100)
(11010010 10101100 01100011)

Although the number was embedded into the first 8 bytes of the grid, only the 3 underlined bits needed to be changed according to the embedded message. On average, only half of the bits in an image will need to be modified to hide a secret message using the maximum cover size. Since there are 256 possible intensities of each primary colour, changing the LSB of a pixel results in small changes in the intensity of the colours. These changes cannot be perceived by the human eye - thus the message is successfully hidden. With a well-chosen image, one can even hide the message in the least as well as second to least significant bit and still not see the difference.

In the above example, consecutive bytes of the image data – from the first byte to the end of the message – are used to embed the information. This approach is very easy to detect [4]. A slightly more secure system is for the sender and receiver to share a secret key that specifies only certain pixels to be changed. Should an adversary suspect that LSB steganography has been used, he has no way of knowing which pixels to target without the secret key.

In its simplest form, LSB makes use of BMP images, since they use lossless compression. Unfortunately to be able to hide a secret message inside a BMP file, one would require a very large cover image. Nowadays, BMP images of 800 × 600 pixels are not often used on the Internet and might arouse suspicion. For this reason, LSB steganography has also been developed for use with other image file formats.

LSB and Palette Based Images

Palette based images, for example GIF images, are another popular image file format commonly used on the Internet. By definition a GIF image cannot have a bit depth greater than 8, thus the maximum number of colours that a GIF can store is 256. GIF images are indexed images where the colours used in the image are stored in a palette, sometimes referred to as a colour lookup table. Each pixel is represented as a single byte and the pixel data is an index to the colour palette. The colours of the palette are typically ordered from the most used colour to the least used colours to reduce lookup time.

GIF images can also be used for LSB steganography, although extra care should be taken. The problem with the palette approach used with GIF images is that should one change the least significant bit of a pixel, it can result in a completely different colour since the index to the colour palette is changed. If adjacent palette entries are similar, there might be little or no