ISSN 0975 - 6299 Vol 2 / Issue 3 / Jul - Sept 2011

International Journal of Pharma and Bio Sciences.

REVIEW ARTICLE BIO INFORMATICS

SECURE CONNECTIVITY USING IMAGE STEGANOGRAPHY
THEME
CYBER SECURITY

PROF.MRS.SARITA DHAWALE.

Ashoka Institute of Management & Technology Nashik
Department of Computer Science.
University of Pune, Maharashtra.

***Corresponding author**

ABSTRACT

Providing Information security for client authentication over a network in a client/server environment is critical issue. Authenticated connection establishment between client and server requires password verification, in which client provides the password and server verifies it. Successful password verification initiates the client/Server to perform secured request/reply mechanisms. Hence there is need for confidential transmission of password in an unsecured network.

Steganography is the art and science of writing hidden messages in such a way that no-one apart from the sender (client) and intended recipient(Server) even realizes there is a hidden message. The term Steganography includes the concealment of digital information within computer files.

Generally a Steganography messages will appear to be something else: a picture, an article, a shopping list, or some other message.Steganography techniques can be used for providing confidentiality of password. This paper intends to give an overview of image steganography, its uses and techniques. It also attempts to identify the requirements of a good steganographic algorithm and briefly reflects on which steganographic techniques are more suitable for which applications. In this paper I would like to focus on the most secured connection establishment for client/Server Environment. Characters provided by the client as a password are appended in to an image using Steganography technique. Appended image containing password Is transmitted over network to reach the server. Server retrieves the original password decoding algorithm .Server verifies and authorizes the client for connection establishment. Even if the intruder steals the images over the network he/She will not be able to decode the password from the image.

KEYWORDS

Steganography, request/reply mechanism, Authenticated connection establishment, frequency on Internet, Cryptography, Watermark, Fingerprinting.

1. INTRODUCTION

In this Internet age, one of the most important factors of information technology and communication has been the security of information. Cryptography developed as a technique for securing the secrecy of communication and many different methods 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. This technique is steganography.

Steganography is the art and science of invisible communication. It hides information in other information, thus hiding the existence of the communicated information. The word steganography is derived from the Greek words “stegos” meaning “cover” and “grafia” meaning, “writing”, defining it as “covered writing”. In image steganography, it hides the information exclusively in images.

Today computers with digital data need steganography, as the carriers and networks being the high-speed delivery channels. Steganography differs from cryptography in the sense that cryptography focuses on keeping the contents of a message secret, but 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. Once the presence of hidden information is revealed or even suspected, the purpose of Steganography is partly defeated. The strength of steganography develops by combining it with cryptography.

Two other technologies 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.

In watermarking and fingerprinting, the fact that information hidden inside the files may be public knowledge – sometimes it may even be visible – while in steganography the imperceptibility of the information is crucial.

Research in steganography developed due to lack of strength in cryptographic systems. Businesses have also started to realise the potential of steganography in communicating trade secrets or new product information. Avoiding
Communication through well-known channels greatly reduces the risk of information leaked in transit. Hiding information in a photograph of the company picnic is less suspicious than communicating an encrypted file.

This paper intends to offer a state of the art overview of the different algorithms used for image steganography to illustrate the security potential of steganography for business and personal use.

2. Overview of Steganography

An overview of steganography includes explanation of important terms and concepts. Different kinds of steganography are explained later.

2.1 Steganography concepts

Although steganography is an ancient subject, Simmons proposes its modern formulation in terms of the prisoner’s problem, 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.

2.2 Different kinds of steganography

Steganography is suitable for all digital file formats, but the formats that are more suited are those with a high degree of redundancy. Figure 1 shows the four main categories of file formats used for steganography.

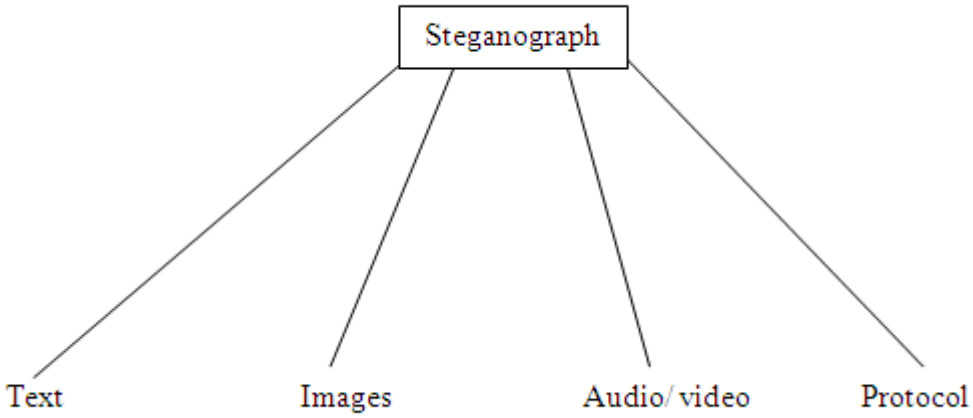

Figure 1
Categories of steganography

Historically, hiding information in text is 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. 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.

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.

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 referred 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 with its colour. These pixels display row by row horizontally. 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 greyscale images use 8 bits for each pixel and are able to display 256 different colours or shades of grey. 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 derive from three primary colours: red, green and blue. Eight bits represent each primary colour. 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 large amount of colours displays in the larger size file.

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 incorporated can reduce the image's file size. These techniques make use of mathematical formulas to analyse and condense image data, resulting in smaller file sizes. This process is compression. In images there are two types of compression: lossy and lossless.

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 approximation 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 remains the same 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 because excess image data is removed. Lossless compression keeps the original digital image intact without the chance of being lost, although it does not compress the image to such a small file size. Different steganographic algorithms developed for both these compression types.

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, image is transformed first 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 characterised 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 section, steganographic algorithms are explained in categories-according to image file formats and the domain in which they are performed.

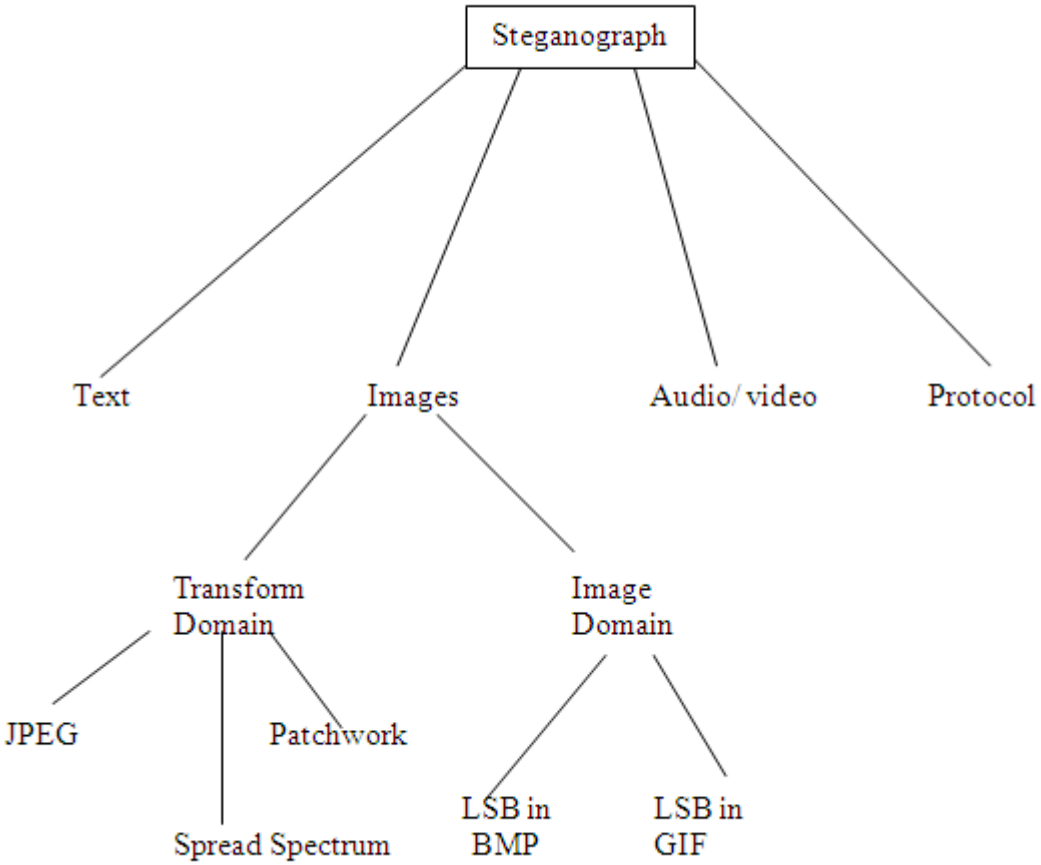

Figure 2
Categories of image steganography

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 is used, since each of them is 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 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. The human eye cannot perceive these changes -thus the message will be hidden successfully. 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. 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 LSB steganography, he will not know which pixel 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 might arouse suspicion, as they are not used in the Internet often. For this reason, LSB steganography 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 a single byte and the pixel data is an index to the colour palette. The colours of the palette are ordered from the most used colour to the least used colours to reduce lookup time.

LSB steganography uses GIF images, although extra care is necessary. 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. The index to the colour palette is changed. If adjacent palette entries are similar, there might be little or no noticeable change. If the adjacent palette entries were very dissimilar, the change would be evident. One possible solution is to sort the palette so that the colour differences between consecutive colours minimized. Another solution is to add new colours that are visually similar to the existing colours in the palette. This requires the original image to have less unique colours than the maximum number of colours (this value depends on the bit depth used). Using this approach, one should thus carefully choose the right cover image. Unfortunately any tampering with the palette of an indexed image leaves a very clear signature, making it easier to detect.

A final solution to the problem is to use greyscale images. In an 8-bit greyscale GIF image, there are 256 different shades of grey. The changes between the colours are very gradual, making it harder to detect.

3.2 Transform Domain

To understand the steganography algorithms used when embedding data in the transform domain, one must first explain the type of file format connected with this domain. The JPEG file format is the most popular image file format on the Internet, because of the small size of the images.

JPEG compression

To compress an image into JPEG format, the RGB colour representation is first converted to a YUV representation. In this representation the Y component corresponds to the luminance (or brightness) and the U and V components stand for chrominance (or colour) . According to research, the human eye is more sensitive to changes in the brightness (luminance) of a pixel than to changes in its colour. This fact is exploited by the JPEG compression by down sampling the colour data to reduce the size of the file. The colour components (U and V) are halved in horizontal and vertical directions, thus decreasing the file size by a factor of two.

The next step is the actual transformation of the image. For JPEG, the Discrete Cosine Transform (DCT) is used, but similar transforms are for example the Discrete Fourier Transform (DFT). These mathematical transforms convert the pixels in such a way as

to give the effect of "spreading" the location of the pixel values over part of the image . The DCT transforms a signal from an image representation into a frequency representation, by grouping the pixels into 8 × 8 pixel blocks and transforming the pixel blocks into 64 DCT coefficients each. A modification of a single DCT coefficient will affect all 64-image pixels in that block.

The next step is the quantization phase of the compression. Here another biological property of the human eye is exploited: The human eye is good at spotting small differences in brightness over a relatively large area, but not so good as to distinguish between different strengths in high frequency brightness. This means that the strength of higher frequencies can be diminished, without changing the appearance of the image. JPEG does this by dividing all the values in a block by a quantization coefficient. The results can be rounded off to integer values and the coefficients encoded using Huffman coding to reduce the size further.

JPEG steganography

Originally, it was thought that steganography would not be possible to use with JPEG images, since they use lossy compression, which results in parts of the image data being altered. One of the major characteristics of steganography is the fact that information is hidden in the redundant bits of an object. Since redundant bits are left out when using JPEG, it was feared that the hidden message would be destroyed. Even if one could keep the message intact, it would be difficult to embed the message without the changes being noticeable because of the harsh compression applied. However, properties of the compression algorithm have been exploited in order to develop a steganographic algorithm for JPEGs.

One of these properties of JPEG is exploited to make the changes to the image invisible to the human eye. During the DCT transformation phase of the compression algorithm, rounding errors occur in the coefficient data that are not noticeable. Although this property is what classifies the algorithm as being lossy, this property can also be used to hide messages.

It is neither feasible nor possible to embed information in an image that uses lossy compression, since the compression would destroy all information in the process. Thus, it is important to recognize that the JPEG compression algorithm is actually divided into lossy and lossless stages

3.3 Image or Transform domain

As seen in Figure 2, some steganographic algorithms can either be categorised as being in the image domain or in the transform domain depending on the implementation.

Patchwork

Patchwork is a statistical technique that uses redundant pattern encoding to embed a message in an image. The algorithm adds redundancy to the hidden information and then scatters it throughout the image. A pseudorandom generator is used to select two areas of the image (or patches), patch A and patch B. All the pixels in patch A is lightened while the pixel in patch B is darkened. In other words the intensities of the pixels in the one patch are increased by a constant value, while the pixels of the other patch are decreased with

the same constant value. The contrast changes in this patch subset encodes one bit and the changes are typically small and imperceptible, while not changing the average luminosity.

A disadvantage of the patchwork approach is that only one bit is embedded. One can embed more bits by first dividing the image into sub-images and applying the embedding to each of them. The advantage of using this technique is that the secret message is distributed over the entire image, so should one patch be destroyed, the others may still survive. This however, depends on the message size, since the message can only be repeated throughout the image if it is small enough. If the message is too big, it can only be embedded once. The patchwork approach is used independent of the host image and proves to be quite robust as the hidden

message can survive conversion between lossy and lossless compression.

Spread Spectrum

In spread spectrum techniques, hidden data is spread throughout the cover-image making it harder to detect. A system proposed by Marvel et al. combines spread spectrum communication, error control coding and image processing to hide information in images Spread spectrum communication can be defined as the process of spreading the bandwidth of a narrowband signal across a wide band of frequencies. This can be accomplished by adjusting the narrowband waveform with a wideband waveform, such as white noise. After spreading, the energy of the narrowband signal in any one frequency band is low and therefore difficult to detect. In spread spectrum image steganography the message is embedded in noise and then combined with the cover image to produce the stego image. Since the power of the embedded signal is much lower than the power of the cover image, the embedded image is not perceptible to the human eye or by computer analysis without access to the original image.

4*. Evaluation of different techniques*

All the above mentioned algorithms for image steganography have different strong and weak points and it is important to ensure that one uses the most suitable algorithm for an application. All steganographic algorithms have to comply with a few basic requirements. The most important requirement is that a steganographic algorithm has to be imperceptible.

These requirements are as follows:

Invisibility **–** The invisibility of a steganographic algorithm is the first and foremost requirement, since the strength of steganography lies in its ability to be unnoticed by the human eye. The moment that one can see that an image has been tampered with, the algorithm is compromised

Payload capacity – Unlike watermarking, which needs to embed only a small amount of copyright information, steganography aims at hidden communication and therefore requires sufficient embedding capacity.

Robustness against statistical attacks **–** Statistical steganalysis is the practice of detecting hidden information through applying statistical tests on image data. Many steganographic algorithms leave a 'signature' while embedding information that could be detected easily through statistical analysis. To be able to pass by a warden without being detected, a steganographic algorithm must not leave such a mark in the image as be statistically significant.

Robustness against image manipulation – In the communication of a stego image by trusted systems, the image may undergo changes by an active warden in an attempt to remove hidden information. Image manipulation, such as cropping or rotating, can be performed on the image before it reaches its destination. Depending on the manner in which the message be embedded, these manipulations may destroy the hidden message. It is preferable for steganographic algorithms to be robust against either malicious or unintentional changes to the image.

Independent of file format – With many different image file formats used on the Internet, it might create suspicion if only one type of file format communicates between two parties. The most powerful steganographic algorithms thus possess the ability to embed information in any type of file. This also solves the problem of not being able to find a suitable image at the right moment, in the right format to use as a cover image.

The levels at which the algorithms satisfy the requirements are defined as high, medium and low. A high level means that the algorithm completely satisfies the requirement, while a low level indicates that the algorithm has a weakness in this requirement. A medium level indicates that the requirement depends on outside influences, for example the cover image used. LSB in GIF images has the

potential of hiding a large message, but only when the most suitable cover image chosen.
The ideal, in other words a perfect; steganographic algorithm would have a high level in every requirement. Unfortunately, in the algorithms evaluated here, no algorithm satisfies all of the requirements. Thus, a trade-off will exist in most cases, depending on which requirements are more important for the specific application.

LSB in BMP – When embedding a message in a "raw" image, that has not been changed with compression (a BMP), there exists a trade-off between the invisibility of the message and the amount of information that can be embedded. A BMP is capable of hiding quite a large message, but the fact that more bits altered, results in a larger possibility that the altered bits can be seen with the human eye. The main disadvantage regarding LSB in BMP images is surely the suspicion that might arise from a very large BMP image being transmitted between parties, since BMP is not widely used anymore. Suggested applications: LSB in BMP is most suitable for applications where the focus is on the amount of information to be transmitted and not on the secrecy of that information.

LSB in GIF – The strong and weak points regarding embedding information in GIF images using LSB are more or less the same as those of using LSB with BMP. The main difference is that since GIF images only have a bit depth of 8, the amount of information that can be hidden is less than with BMP. GIF images are especially vulnerable to statistical – or visual attacks – since the palette processing done leaves a very definite signature on the image. This approach is dependent on the file format as well as the image itself, since a wrong choice of image can result in the message being visible. Suggested applications: LSB in GIF is a very efficient algorithm to use when embedding a reasonable amount of data in a greyscale image.

JPEG compression – The process of embedding information during JPEG compression results in a stego image. It has high level of invisibility, since the embedding takes place in the transform domain. JPEG is the most popular image file format on the Internet and the image sizes are small because of the compression, thus making it the least suspicious algorithm to use. However, the process of the compression is a very mathematical process, making it more difficult to implement. Suggested applications: The JPEG file format can be used for most applications of steganography, but is especially suitable for images that have to be communicated over an open systems environment like the Internet.
Patchwork – The biggest disadvantage of the patchwork approach is the small amount of information that can be hidden in one image. This property can be changed to accommodate more information but one may have to sacrifice the secrecy of the information. Patchwork's main advantage, however, is its robustness against malicious or unintentional image manipulation. Should a stego image using patchwork be cropped or rotated, some of the message data may be lost but since the message is repeatedly embedded in the image, most of the information will survive. Suggested applications: Patchwork is most suitable for transmitting a small amount of very sensitive information.

Spread spectrum – Spread spectrum techniques satisfies most requirements and is especially robust against statistical attacks, since the hidden information is scattered throughout the image, while not changing the statistical properties. Suggested applications: Spread spectrum techniques can be used for most steganography applications, although its highly mathematical and intricate approach may prove too much for some.

5. CONCLUSION

Although some steganographic techniques were discussed in this paper, there exists a large selection of approaches to hiding information in images. All the major image file formats have different methods of hiding

messages, with different strong and weak points respectively. Where one technique lacks in payload capacity, the other lacks in robustness. For example, the patchwork approach has a very high level of robustness against most type of attacks, but can hide only a very small amount of information. Least significant bit (LSB) in both BMP and GIF makes up for this. But both approaches result in suspicious files that increase the probability of detection when in the presence of a warden. Thus, for an agent to decide on which steganographic algorithm to use, he would have to decide on the type of application he want to use the algorithm for, and if he is willing to compromise on some features to ensure the security of others.

6. LIST OF REFERENCES

1. Dunbar, B., “Steganographic techniques and their use in an Open-Systems environment”,
2. Artz, D., “Digital Steganography: Hiding Data within Data”, IEEE Internet Computing Journal
3. Chandramouli, R., Kharrazi, M. & Memon, N., “Image steganography and steganalysis: Concepts and Practice”, Proceedings of the 2nd International Workshop on Digital Watermarking,