American Journal of Applied Science 2 (3): 648-654, 2005
ISSN 1546-9239

An Investigation into the Use of the Least Significant Bit Substitution Technique in Digital Watermarking

Kevin Curran, Xuelong Li and Roisin Clarke
Internet Technologies Research Group
University of Ulster, Magee Campus, Northland Road, Northern Ireland, UK

Abstract: Digital imaging companies lose revenue each year to people who are illegally copying and using their images. One prevention mechanism is to digitally encode images making it difficult for others to copy. This could be done using digital fingerprinting or simply adding a visible watermark. The information is hidden in a way that should not attract attention, but make it very difficult to make an exact replica of the image. The information is encoded within a host image, so that the actual appearance of the image does not change, but within the image there is a watermark or secret message, which prohibits the attacker from making an exact copy. The objective with Steganography is not to change the actual message, or make it difficult to read, as cryptography does, rather to hide the existence of the message without distorting the carrier or the actual information. This study presents the results of implementing a Least Significant Bit (LSB) digital watermarking system.

Key words: Digital Watermarking System, Least Significant Bit, Secret Message Transmitting

INTRODUCTION

Steganography is derived from the Greek word Steganos which means *covered* or *secret*, and *graphy* meaning written or drawn. The art of Steganography originated from a Greek man named Histiaeus, who was a prisoner of a rival king. He needed a way of transmitting a secret message to his people. He had the idea of shaving a willing slaves head and tattooing the message onto his scalp. When the slave's hair grew back, he was sent to deliver the message to Histiaeus' army [1].

The objective of steganography is to send a message through some media known as a carrier, to a receiver, while preventing anyone else from knowing that the message exists. The carrier can be one of many different digital media, but the most common is the image. The image should not attract any attention as a carrier of a message and should compare as close as possible to the original image by the human eye. When images are used as the carrier in steganography, they are generally manipulated by altering one or more bits of the byte that make up the pixels of the image. The least significant bit (LSB) may be used to encode the bits of the message. These LSB's can then be read by the recipient of the stego image and put together as bytes to reproduce the hidden message, providing they have the stego key – the password for the stego image.

Steganalysis is the art of discovering a message. Breaking a steganographic system involves detecting that steganography has been used, reading the embedded message and proving that the message has been embedded to third parties. Steganalysis methods are also used by the steganographer to determine whether the message is secure and whether the process has been successful. Detection involves observing relationships between combinations of cover, stego media and steganography tools. This can be achieved by passive observation of patterns or unusual exaggerated noise and visual corruption. The patterns visible to the human eye could broadcast the existence of the message and point to signatures of certain methods or tools used.

If numerous comparisons are made between the cover images and the stego images, patterns can begin to emerge. Some of the methods of carrying out steganography produce characteristics that act as signatures for that particular steganography method. Detection might involve looking at areas in the image where colour does not flow well from one area to the next. The attacker should obviously not be familiar with the cover image; otherwise it would make it a lot easier for comparison.

Today steganography is used for transmitting data, as well as hiding trademarks in images and music. This is known as digital watermarking. Cryptography and steganography are different in their methods of hiding information. Cryptography scrambles a message and hides it in a carrier, so that if it is intercepted it would be generally impossible to decode. Steganography hides the very existence of the message in the carrier. When the message is hidden in the carrier a stego-carrier is formed e.g. a stego-image. If successful, it would be perceived to be as close to the original carrier or cover image by the human eye. Images are the most widespread carrier medium [2]. Images are the most widespread carrier medium [2]. They are used for steganography in the following way:

The message may firstly be encrypted. The sender (or embedder) embeds the secret message to be sent into a graphic file [3]. This results in the production of what is called the stego-image. Additional secret data may be needed in the hiding process e.g. a stegokey [4]. The stegoimage is then transmitted to the recipient.

The recipient or the extractor extracts the message from the carrier image. The message can only be extracted if there is a shared secret between the sender and the recipient. This could be the algorithm for extraction or a special parameter such as a key (the stegokey).

To make the stegonographic process even more secure the message may be compressed and encrypted before it is hidden in the carrier. Figure 1 illustrates the principles of steganography where a carrier message has a message is added and put through a Stegosystem Encoder. The Stegoimage is then sent through the appropriate channels to a Stegosystem Decoder [5].

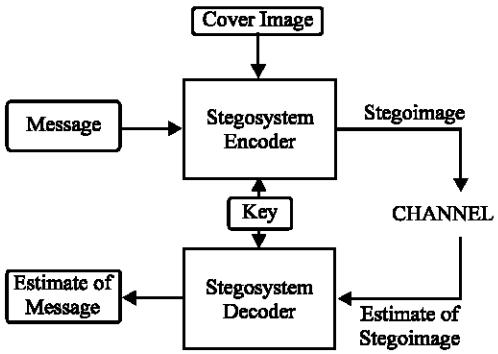


Fig. 1: Steganographic System [5]

For grayscale images each pixel has a value between 0 and 255. The image is broken down into co-ordinates and pixels. The carrier image must be either the same size or larger than the message. The least significant bit of each pixel of the carrier is changed to the least significant bit if each pixel of the message to be hidden. This has the effect of hiding the message but making it appear to be the carrier. The human eye cannot detect and message or any difference to the carrier. It then has to be passed through a stegoimage decoder for the hidden message to be extracted. A username and password is required at this stage.

This is where Cryptography and Steganography can be used together. When the message is compressed it takes up less space in the carrier and will minimize the amount of information to be sent. It also limits the chances of being seen or detected in the carrier. The random message resulting from encryption and compression would be easier to hide than a message with a high degree of regularity. Encryption and compression are recommended in conjunction with steganography, as it offers a higher degree of security and reliability.

There are a variety of digital carriers or places where data can be hidden. Data may be embedded in files at imperceptible levels of noise and properties of images can be changed and used in a way useful to your aim. Features such as luminescence, contrast and colours can be changed according to which one is most useful to your particular application. This study focuses on bit values of pixel in the grayscale range which can be altered to embed hidden images inside other images, without changing the actual appearance of the carrier image.

Image Steganography: In Least Significant Bit (LSB) substitution, the least significant bit is changed because this has little effect to the appearance of the carrier message as shown in Fig. 2.

Image 1:

1	0	0	0	0	0	0	0

The grayscale pixel bit size is: 128

Image 2:

1	0	0	0	0	0	0	1

The grayscale pixel bit size here is: 129 with the LSB changed.

Image 3:

0	0	0	0	0	0	0	0

By changing the MSB here the bit size has changed from 128 to 0.

Fig. 2: LSB and MSB Substitution

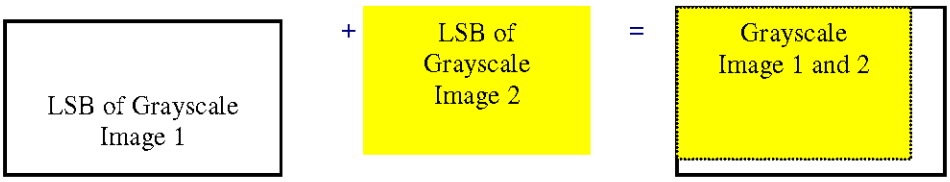


Fig. 3: Image Processing Results

This shows that the grayscale image would change significantly if there were any other bit changed than the LSB. It changes more and more the closer you get to the MSB. When the LSB is changed, the pixel bit value changes from 128 to 129, which is undetectable with the human eye. With the MSB changed, the pixel bit value changes from 128 to 0, which makes a significant change to the grayscale view. The theory is that if you take two grayscale images, and change the LSB of image one to the LSB of image two for each co-ordinate or pixel, image two will be hidden in image one as illustrated in Fig. 3.

When the second image is embedded in the first, there should be no detectable change or alteration to the appearance of the first image. A digital image is the most common type of carrier used for steganography. A digital image is produced using a scanner, camera or other digital device. The digital representation is an approximation of the original source. The system used for producing the images focuses two dimensional patterns of varying light intensity and colour onto a sensor. The pattern, in this case a grayscale pattern, has a co-ordinate system with the origin in the upper left corner of the image. The pattern can be described by a function f(x,y). The pattern can be described as an array of numbers that represent light intensities at various points. These are known as pixels. Sampling is the process of measuring the value of the image function, f(x,y) at discrete intervals. Each sample is the small square area of the image known as the pixel. The raster data of an image is the part of the image that can be seen on screen. Pixels are indexed by x and y co-ordinates (x and y are integer values). Dense sampling produces high-resolution images in which there are many pixels, each contributing to a small part of the image. Coarse sampling results in a low-resolution image in which there are fewer pixels.

Application of Steganography to Images: When an image is used as a carrier in steganography, it is generally manipulated by changing on or more bits of the byte, in our case the LSB. If it corresponds to the bit to be hidden or embedded it is left unchanged. Otherwise it is changed to correspond to the hidden bit. These LSB's can then be read by the recipient of the stego image and put together as bytes to reproduce the hidden message. In a grayscale image, each pixel is either black or white and has a level of between 0 and 255, as each pixel has eight bits. Steganography is carried out by changing the low order bit of a pixel, and using it to encode one bit of a character. There are two stages in the steganalysis system:

* Detecting that steganography has been used
* Reading the embedded message

Steganalysis is used by a steganographer in order to determine whether a message is secure and consequently whether the steganographic method has been successful. The aim of a stegoanalyst is to detect stegoimages, find and read the embedded message, and prove that the message has been embedded to third parties. Detection involves observing relationships between combinations of cover, message, stego-media and steganographic tools [6]. Active interference by the stegoanalyst involves removing the message without altering the stego image too much, or removing the image or message without consideration to the stego-image appearance or structure [4]. There are two necessary conditions to be fulfilled for a secure steganographic process. The key must remain unknown or undetectable to the attacker, and the attacker should not be familiar with the cover image [3]. If the cover image is known and it is impossible to keep it unknown from the attacker, the message could be embedded in a random way so that it is secure, as long as the key remains unknown. However, it is preferable that the cover image remains a secret to obtain maximum security.

Watermarking: Watermarking is the process of hiding information in a carrier in order to protect the ownership of text, music, films and art [3]. Watermarking can be used to hide or embed visible or hidden copyright information [7]. Steganographic techniques can be used for the purposes of digital watermarking. Often information is hidden about the carrier itself providing further information about the carrier, which is not explicitly, displayed [8]. Watermarks in images are hidden mainly so that the do not disturb or distort the image rather than to avoid detection. They are generally hidden in more significant areas of the image and are not lost by compression. The main aim of watermarking is to prevent unlawful

reproduction of a product. The ID of the author can be hidden so that if the image is circulated, the ID of the author still remains embedded in the image. In some cases watermarks are clearly visible. In these cases they are not a type of steganography, but are part of the actual image.

Watermarking does not impair the image. This is a main concern with visible watermarking. Even though the watermark can be seen, it must be inserted in such a way that it does not interfere with the original image. The underlying image must still be legible. If the watermark blocks large portions of the original image or the entire image, it is not an effective watermark. There is also no point in a watermark that can be easily removed. The typical litmus test with the watermark is that if the watermark is removed then the image should be impaired or destroyed. Even though only a small amount of data is to be embedded, it should be inserted in more than one place so that it is more difficult to remove. Someone who is trying to remove the watermark would be unlikely to detect all of the watermarks from the original image. There are two types of digital watermarking, visible and invisible. As can be seen from Fig. 4 and 5 [1], the two contain watermarks; the first figure contains a visible watermark, whereas the second watermark is invisible. Unlike steganography, it is irrelevant if a digital watermark is detected. Some companies prefer their watermarks to be visible to deter possible thieves. It is essential though, that the watermark cannot be removed or tampered with. Figure 6 illustrates how an image undergoes the watermarking process and appears at the other side apparently unaltered to the user. In an image such as this, the least significant four bits could be changed without any perceptual change in the resultant image allowing enough space to hide a secret message [9].

Invisible Watermarks: In visible watermarking, a pattern is applied to a file or image so that it is undetectable by the human eye. With an invisible watermark you can change certain pixels in an image so the human eye cannot tell the difference from the original image. A computer program can however detect these discrepancies. Another factor

Fig. 4: Visible Watermarking

Fig. 5: Invisible Watermarking

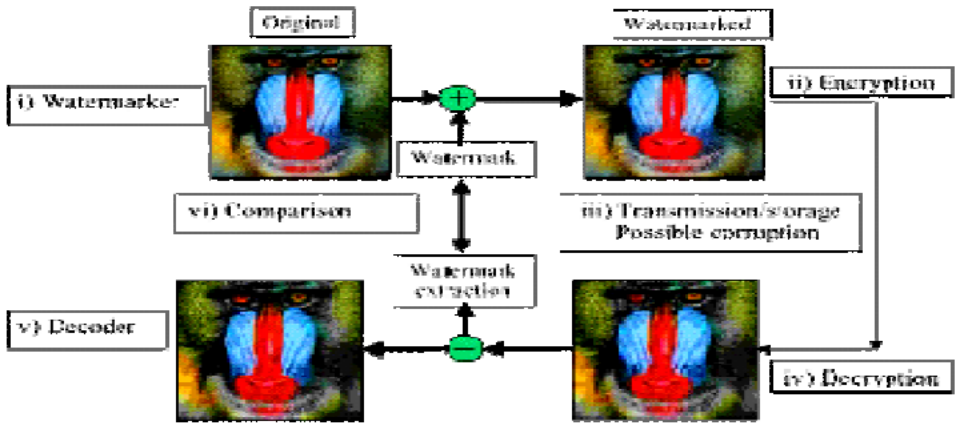


Fig. 6: Colored Watermarking Process [10]

Table 1: Steganography vs. Watermarking

Characteristic	Steganography	Digital Watermarking
Amount of Data	As much as Possible	Small Amount
Ease of Detection	Very Difficult to Detect	Not Critical with Visible Watermarks
Ease of Removal	Very Important it cannot be Removed	Important it cannot be Removed
Goal Of Attacker	To Detect the Data	To Remove the Data
Goal of User	To Hide Information so it cannot be Removed	To Embed a Signature to Prove Ownership
Current Uses	Covert Communications	Protecting Rights of Owners

Table 1 shows the differences between steganography and digital watermarking. It depends on the use, which one is chosen; each has its advantages and disadvantages. The user will know best which one to choose depending on the application and situation.

to consider when applying an invisible watermark is the actual size of the pixels. The smaller the pixel, the less chance there is of detecting the change in colour. The strength of invisible watermarks is that the image quality is not degraded or changed according to the user or consumer. When looking at the image, there is no way of telling there is a watermark, yet the digital image is still protected. Invisible watermarks are effective, though, only while the image is in digital form. If a digital image that has an invisible watermark is printed out, and then rescanned, the watermark is effectively removed.

Visible Watermarking: A visible watermark makes slight modifications to an image. The transformation is such that the image can still be seen, but the watermark is effectively laid over the top of it. One of the advantages of visible watermarks is that even if an image is printed and scanned the watermark is still visible. A visible watermark image will usually make use of a light grayscale tone or simply make the pixels that contain the watermark slightly lighter or darker than the surrounding area. When applying a visible watermark, it is essential that the watermark is applied to enough of the image that it cannot be removed, and the original image can still be seen and is still legible. If you apply too much of the watermark, all you will see is the watermark and little of the actual image. Complex mathematical formula can be used to make watermarks more robust, but at a general level a watermarking program finds a group of pixels and adjusts the pixels in a way that the watermark can be seen but the image is not destroyed. The usual way of performing this operation is to make the colour of specific pixels darker.

Attacks on Watermarks: This will involve trying to remove or distort the watermark. The hidden information should be made such an integral part of the image so that it is impossible to remove it without destroying the image. If the watermark is hidden in the LSB, all the individual has to do is flip one LSB and the information cannot be recovered. Various image processing techniques may be used to attack a watermark. It is particularly successful when using the same algorithm as was used to produce the existing watermark. One common problem with watermarks is that their existence is often advertised so that potential users know that the image has copyright information embedded.

Requirements of a Steganographic System: There are a number of ways of hiding information in image pixels. In some methods the objective is to store the message in a random way so as to make it more difficult to detect. These methods typically involve the use of a key – the stego key. The best types of images to use are black and white grayscale or natural photographs. The redundancy of the data helps to hide the existence of a secret message. A cover image should contain some randomness. It should contain some natural uncertainty or noise, as hiding information may introduce enough noise to raise suspicion. Therefore the carrier or cover image must be carefully selected. Once it has been used, the image should not be used again and should be destroyed. A familiar image should not be used, it is better for steganographers to create their own images.

Perceptual Transparency: The Watermark should not affect the quality of the original image. Where possible, the watermark should go undetectable, as this increases security. Attackers find detectible watermarks much easier to remove or manipulate.

Robustness: This is a measure of how well the watermark withstands various methods of image processing. The image may be subjected to filtering, rotation, translation, cropping, scaling etc. as part of image processing. The more robust the watermark is the better it will perform when these methods are applied. If the watermark algorithm is embedded using the spatial or frequency domain, it will withstand the image processing much better. There is also a watermark type

called 'Fragile', which is intentionally made non-robust as they are used for authentication of original material rather than tracing it back to the source after processing.

Security: To improve security, it is important that third parties cannot alter the watermark even if they know the algorithm for embedding and recovering.

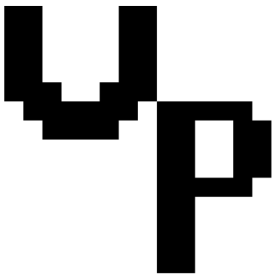
Payload of Watermark: The amount of data that can be stored in a watermark depends on the application. For example, in copyright a payload of one bit is sufficient, but for intellectual copyrights such as ISBN a payload of 60-70 bits is required. Watermarking Granularity is a term used to represent the number of bits that are actually needed to represent a watermark within an image.

Oblivious vs. Non-Oblivious: in copyright and data protection applications, the extraction algorithm can use the original unwatermarked data or image to find the watermark. This is non-oblivious watermarking. In other applications such as indexing or copy protection the watermark extraction algorithm cannot access the original image and therefore makes detection and extraction very difficult for possible attackers. This type of watermarking is known as oblivious.

Evaluation: An effective way to implement steganography is to use images as the carrier and hidden message and use the pixel values as the method by changing the LSB. We restricted our investigation to grayscale images and the effect of transferring the data through embedding images in carriers. Image processing was done through Matlab 6.0. As mentioned earlier, LSB Substitution involves embedding a watermark by replacing the least significant bit of the image data with a bit of the watermark data. Detection can be done visually or by correlation methods. One of the drawbacks of this method is that if the algorithm is discovered, it is relatively easy for someone to alter it and defeat the purpose. This is why the watermark is often placed in more than one place in the original image.

512x512 pixels

Fig. 7: The Image



16x16 pixels

Fig. 8: The Watermark

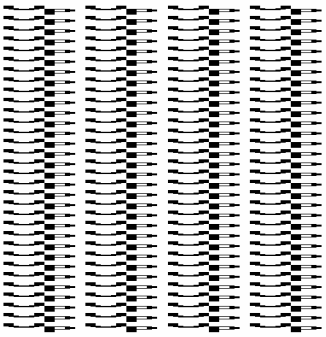


Fig. 9: The Recovered Watermark

Figure 7 shows the actual image used for the tests and Fig. 8 shows the watermark used. Figure 9 shows the watermark recovered when the code is implemented with the LSB method. This shows the repetitive nature of this method, making it more difficult for attackers to manipulate.

There however a number of drawbacks with LSB Substitution. LSB Substitution can survive simple operations such as cropping, as it is placed in numerous locations, but any addition of noise or compression of the image is going to overcome the effects quickly. Also, if the watermark is detected, an attacker would only have to replace all LSB bits with a '1' fully defeating the effects.

One solution to improve the robustness of the watermark is to use a pseudo random number generator to determine the pixels to be used for the embedding. Security is increased, but it does require effectively a password or key to be sent with the image, or shared among users. Provided the attacker does not receive the password it would be very difficult to manipulate the watermark. We are currently investigating the use of pseudo random number generation algorithms to work alongside LSB Substitution.

CONCLUSION

Digital watermarking is used by those who wish to prevent others from stealing their material. LSB

substitution is not a very good candidate for digital watermarking, but it is very useful in the art of steganography, due to its lack of robustness. LSB embedded watermarks can easily be removed using techniques that do not affect the image visually to the point of being noticeable. Furthermore if one of the other embedding algorithms is used, the encoded message can be easily recovered and even altered by an attacker. It would appear that LSB will remain in the domain of steganography due to its useful nature and its overall capacity of information.

REFERENCES

1. Cole, E., 2003. Hiding in Plain Sight. John W. Wiley, ISBN: 0-471-44449-9.
2. Westfield, A. and A. Pfitzmann, 1999. Attacks on Steganographic Systems. Third International Workshop, IH'99 Dresden Germany, October Proceedings, Coputer Science, 1768: 61- 76.
3. Zollner J., H. Federrath, H. Klimant, A. Pfitzmann, R. Piotraschke, A. Westfeld, G. Wicke and G. Wolf, 1998. Modelling the Security of Steganographic Systems, Information Hiding, 2nd International Workshop, IH'98 Portland, Oregon, USA, Computer Science, 1525: 344-354.
4. Pfitzmann, B., 1996. Information Hiding Terminology collected by Birgit Pfitzmann. Information Hiding First international Workshop, Cambridge.
5. Marvel, L.M., C.G. Boncelet and C.T. Retter, 1998. Reliable Blind Information Hiding for Images. Proc. of Information Hiding Workshop, pp: 48-62.
6. Johnson, N.F., Zoran Duric and Sushil Jajodia, 2001. Information Hiding, and Watermarking-Attacks & Countermeasures, Kluwer.
7. Wayner, P., 2002. Disappearing Cryptography, Information Hiding: Steganography and Watermarking. 2nd Edition, Morgan Kaufmann.
8. Johnson, N.F. and Sushil Jajodia, 1998. Steganalysis of images created using current Steganography Software, Centre for secure Information Systems, George Mason University, Fairfax, Virginia, information Hiding Second Workshop, IH'98 Portland, Oregon USA, proceedings Computer Science 1525, pp:273-289.
9. Strange, 2000. http://www.strangehorizons.com/2001/20011008/steganography.shtml
10. Vallabha, V.H., 2001. Multiresolution watermark based on wavelet transform for digital images. MATHLAB Central Document Repository, http://www.mathworks.com/matlabcentral/