CERIAS Tech Report 2001-139
A Review of Data Hiding in Digital Images
by E Lin, E Delp
Center for Education and Research
Information Assurance and Security
Purdue University, West Lafayette, IN 47907-2086

A Review of Data Hiding in Digital Images

Eugene T. Lin and Edward J. Delp
Video and Image Processing Laboratory (VIPER)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana

Abstract

In this paper we will overview the use of data hiding techniques in digital images. In particular we will describe how one can use Steganography to hide information in a digital image. Steganography is related to cryptography and is the basis for many of the digital watermarking techniques currently being developed. The interest in data hiding has risen with the recent activity in digital copyright protection schemes. One way to protect the ownership of a digital image is to secretly embed data in the content of the image identifying the owner. This paper will review recent developments in data hiding, specifically as it pertains to copyright protection of digital images.

Introduction

Steganography is the practice of encoding secret information in a manner such that the very existence of the information is concealed. Throughout history, many steganographic techniques have been documented, including the use of cleverly-chosen words, invisible ink written between lines, modulation of line or word spacing, and microdots [1,2,3]. Usually the secret information is concealed by the use of an innocuous cover as to not arouse suspicion if hostile agents discover the cover. As an example, the cover text:

> I'm feeling really stuffy. Emily's medicine wasn't strong enough without another febrifuge.

hides the sentence "Meet me at nine" if the reader retains the second letter of each word in sequence.

Digital Steganography

A typical digital steganographic encoder is shown on Figure 1. The *message* is the data that the sender wishes to remain confidential and can be text, images, audio, video, or any other data that can be represented by a stream of bits. The *cover* or *host* is the medium in which the message is embedded and serves to hide the presence of the message. This is also referred to as the message *wrapper*. The message embedding technique is strongly dependent on the structure of the cover, and in this paper covers are restricted to being digital images. It is not required that the cover and the message have homogeneous structure. For example, it is possible to embed a recording of Shakespeare's lines (an audio stream message) inside a digital portrait of the famous playwright (an image cover).

The image with the secretly embedded message produced by the encoder is the *stego-image*. The stego-image should resemble the cover image under casual inspection and analysis. In addition, the encoder usually employs a *stego-key* which ensures that only recipients who know the corresponding decoding key will be able to extract the message from a stego-image.

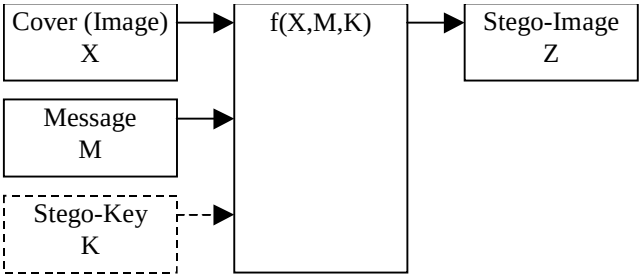

Figure 1. Steganographic Encoding

Recovering the message from a stego-image requires the stego-image itself and a corresponding decoding key if a stego-key was used during the encoding process. The original cover image may or may not be required; in most applications it is desirable that the cover image not be needed to extract the message.

Steganography is not the same as cryptography. In cryptography, the structure of a message is changed to render it meaningless and unintelligible unless the decryption key is available. Cryptography makes no attempt to disguise or hide the encoded message. Steganography does not alter the structure of the secret message, but hides it inside a cover. It is possible to combine the techniques by encrypting a message using cryptography and then hiding the encrypted message using steganography. The resulting stego-image can be transmitted without revealing that secret information is being exchanged. Furthermore, even if an attacker were to defeat the steganographic technique and detect the message from the stego-image, he would still

Address all correspondence to E. J. Delp, ace@ecn.purdue.edu, http://www.ece.purdue.edu/~ace, or +1 765 494 1740.

require the cryptographic decoding key to decipher the encrypted message.

Applications

There are many applications for digital steganography of images, including copyright protection, feature tagging, and secret communications [1,4].

Copyright Protection: A secret copyright notice or watermark can be embedded inside an image to identify it as intellectual property [5, 6]. This is the watermarking scenario where the message is the watermark [5, 6]. The "watermark" can be a relatively complicated structure. In addition, when an image is sold or distributed an identification of the recipient and time stamp can be embedded to identify potential pirates. A watermark can also serve to detect whether the image has been subsequently modified [7]. Detection of an embedded watermark is performed by a statistical, correlation, or similarity test, or by measuring other quantity characteristic to the watermark in a stego-image. The insertion and analysis of watermarks to protect copyrighted material is responsible for the recent surge of interest in digital steganography and data embedding.

Feature Tagging: Captions, annotations, time stamps, and other descriptive elements can be embedded inside an image, such as the names of individuals in a photo or locations in a map. Copying the stego-image also copies all of the embedded features and only parties who possess the decoding stego-key will be able to extract and view the features. In an image database, keywords can be embedded to facilitate search engines. If the image is a frame of a video sequence, timing markers can be embedded in the image for synchronization with audio. The number of times an image has been viewed can be embedded for "pay-per-view" applications.

Secret Communications: In many situations, transmitting a cryptographic message draws unwanted attention. The use of cryptographic technology may be restricted or forbidden by law. However, the use steganography does not advertise covert communication and therefore avoids scrutiny of the sender, message, and recipient. A trade secret, blueprint, or other sensitive information can be transmitted without alerting potential attackers or eavesdroppers.

Characterizing Data Hiding Techniques

Steganographic techniques embed a message inside a cover, various features characterize the strengths and weaknesses of the methods. The relative importance of each feature depends on the application [4].

Hiding Capacity: Hiding capacity is the size of information that can be hidden relative to the size of the cover. A larger hiding capacity allows the use of a smaller cover for a message of fixed size, and thus decreases the bandwidth required to transmit the stego-image.

Perceptual Transparency: The act of hiding the message in the cover necessitates some noise modulation or distortion of the cover image. It is important that the embedding occur without significant degradation or loss of perceptual quality of the cover. In a secret communications application, if an attacker notices some distortion that arouses suspicion of the presence of hidden data in a stego-image, the steganographic encoding has failed even if the attacker is unable to extract the message. Preserving perceptual transparency in an embedded watermark for copyright protection is also of paramount importance because the integrity of the original work must be maintained [6].

For applications where the perceptual transparency of embedded data is not critical, allowing more distortion in the stego-image can increase hiding capacity, robustness, or both.

Robustness: Robustness refers to the ability of embedded data to remain intact if the stego-image undergoes transformations, such as linear and non-linear filtering, addition of random noise, sharpening or blurring, scaling and rotations, cropping or decimation, lossy compression, and conversion from digital to analog form and then re-conversion back to digital form (such as in the case when a hard copy of a stego-image is printed and then a digital image is formed by subsequently scanning the hardcopy.) Robustness is critical in copyright protection watermarks because pirates will attempt to filter and destroy any watermarks embedded in images [5, 6]. Anti-watermarking software is already available on the Internet and have been shown effective in removing some watermarks [8,9]. These techniques can also be used to destroy the message in a stego-image.

Tamper Resistance: Beyond robustness to destruction, tamper-resistance refers to the difficulty for an attacker to alter or forge a message once it has been embedded in a stego-image, such as a pirate replacing a copyright mark with one claiming legal ownership. Applications that demand high robustness usually also demand a strong degree of tamper resistance. In a copyright protection application, achieving good tamper resistance can be difficult because a copyright is effective for many years and a watermark must remain resistant to tampering even when a pirate attempts to modify it using computing technology decades in the future.

Other Characteristics: Computational complexity of encoding and decoding is another consideration and individual applications may have additional requirements. For example, for a copyright protection application, a watermark should be resistant to collusion attacks where many pirates work together to identify and destroy the mark.

Data Embedding

Current methods for the embedding of messages into image covers fall into three categories: Least-Significant Bit embedding (or simple embedding), transform techniques, and methods that employ perceptual masking.

Least-Significant Bit Encoding

A digital image consists of a matrix of color and intensity values. In a typical gray scale image, 8 bits/pixel are used. In a typical full-color image, there are 24 bits/pixel, 8 bits assigned to each color components.

The simplest steganographic techniques embed the bits of the message directly into the least-significant bit plane of the cover image in a deterministic sequence. Modulating the least-significant bit does not result in a human-perceptible difference because the amplitude of the change is small. Other techniques "process" the message with a pseudo-random noise sequence before or during insertion into the cover image.

The advantage of LSB embedding is its simplicity and many techniques use these methods [10]. LSB embedding also allows high perceptual transparency. However, there are many weaknesses when robustness, tamper resistance, and other security issues are considered. LSB encoding is extremely sensitive to any kind of filtering or manipulation of the stego-image. Scaling, rotation, cropping, addition of noise, or lossy compression to the stego-image is very likely to destroy the message. Furthermore an attacker can easily remove the message by removing (zeroing) the entire LSB plane with very little change in the perceptual quality of the modified stego-image.

Steganos

"Steganos" is a LSB embedding system developed in Germany that can embed data inside a variety of image, audio, and text covers [11]. The latest version of the software (version 1.5) was used below to illustrate LSB embedding.

The cover image Lena is shown on Figure 2, which is a 256x256 pixel 8-bit grayscale image. The message is a text file containing a single line: "Digital Image Steganography: Data Sneaking Between the Pixels." Using Steganos, the stego-image shown on Figure 3 is produced. (The encryption facility of Steganos was disabled.) The difference image is shown on Figure 4, where "white" pixels indicate the spatial locations where the images differ.

Steganos was able to recover the message when the stego-image was made available for decoding. However to evaluate the fragile nature of the embedding, Gaussian additive noise (with zero-mean and unit variance) was added to each pixel intensity value in the stego-image to produce the altered stego-image shown on Figure 5. Steganos was not able to extract the message. The software erroneously believed that the modified stego-image contained some encrypted data and asked for a decryption key.

Figure 2: Cover Image

Figure 3: Steganos Stego-Image

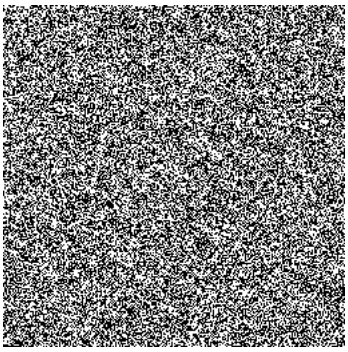

Figure 4: Difference Image

Figure 5: Modified Stego-Image

Transform Embedding Techniques

Another class of techniques is embedding the message by modulating coefficients in a transform domain, such as the Discrete-Cosine Transform (DCT) (used in JPEG compression), Discrete Fourier Transform, or Wavelet Transform. Transform techniques can offer superior robustness against lossy compression because they are designed to resist or exploit the methods of popular lossy compression algorithms. An example of a transform-based steganographic system is the "Jpeg-Jsteg" software [10], which embeds the message by modulating DCT coefficients of the stego-image based upon bits of the message and the round-off error during quantization. Transform-based steganography also typically offer increased robustness to scaling and rotations or cropping, depending on the invariant properties of a particular transform.

Spread-spectrum techniques and redundant encoding of the message can be employed in situations where robustness is critical [5, 6, 12]. The watermark or message can be thought of as a narrowband signal encoded in a larger frequency band (the cover). By spreading the energy of the embedded message across many frequency bands (such as by frequency hopping) the energy at any particular frequency band is reduced. Therefore the message becomes more difficult to detect or modify without damaging the cover. Error correcting coding can be applied to the message during embedding to allow recovery even when some areas of the stego-image may be damaged or altered.

Perceptual Masking Systems

Recently, a great deal of research has been reported in expanding the hiding capacity and robustness of steganographic techniques by exploiting the properties of the human visual system [5, 6, 13]. The development of accurate human vision models facilitates the design and development of perceptual masking hiding systems [6].

Steganographic techniques designed to be robust to lossy image compression must insert the message into the cover in a manner that is perceptually significant. Techniques that attempt to embed information only in a perceptually insignificant manner, such as LSB embedding techniques, are vulnerable to having the embedded data distorted or quantized by lossy image compression.

The masking properties of the human visual system allow perceptually significant embedding to be unnoticed by an observer under normal viewing conditions [6]. "Masking" refers to the phenomenon where a signal can be imperceptible to an observer in the presence of another signal (referred to as the masker.) The masking properties are the reason why it is difficult for one to find a randomly placed needle in a haystack; the needle can be in plain view to an observer (not obscured by any object) yet the observer will have great difficulty locating the needle.

Masking (sometimes referred to as image-adaptive [6]) systems perform analysis of the image and use the information to determine appropriate regions to place the message data. Masking systems can also use the analysis to vary the strength (amplitude) of the embedded data based upon local image characteristics to maximize robustness. These systems can embed in either the spatial or a transform domain.

Steganalysis

Steganalysis is the practice of attacking steganographic methods by detection, destruction, extraction, or modification of embedded data. Understanding the means by which attackers can defeat steganographic systems is necessary for the design and development of superior, more robust systems. The meaning of a successful attack is dependent on the application; for a secret communication application the mere detection and proof that some kind of data is hidden within the stego-image is a successful attack. For a steganalyst pirate attempting to defeat a copyright mark, a successful attack requires that he not only detects the mark but also destroys or modifies the mark without significant degradation of the perceptual quality of the stego-image.

There are parallels between the techniques of cryptanalysis and steganalysis [10]. In both fields, it is assumed that the attacker understands the method used to encrypt or hide the data. That is, the entire secrecy of a particular method must lie with the selection of the encryption or stego-key and not in the intricate workings or proprietary nature of the method. (This is known as Kerkhoff's Principle in cryptography.) Steganalysis techniques can be divided into five categories: *stego-only*, *known cover*, *known message*, *chosen stego*, and *chosen-message*. In a stego-only attack, only the stego-image is available for steganalysis. This is similar to the ciphertext-only attack in cryptanalysis and is the weakest form of attack. In a known cover attack, both the original cover and a corresponding stego-image is available. The known message attack is when the steganalyst knows the secret message embedded in a stego-image. A chosen-stego attack (similar to a chosen ciphertext attack in cryptanalysis) is when access to the message extraction tool is available so the attacker does not have to deduce the decoding algorithm. The most powerful attack is the chosen message attack, where the steganalyst has access to the steganography encoding tool itself and can embed and analyze messages of his own choosing.

Detecting the presence of a watermark or embedded data in covers can be performed by examination of the stego-image for excessive noise or distortions. In some cases, the distortions can be visible under human observation with an experienced observer. The known cover attack simplifies distortion analysis because the stego-image can then be compared with the cover to determine the exact distribution of alteration or modulation. The chosen-message attack also allows the steganalyst to generate many cover – stego-image pairs and then use analysis to determine if there are any "signatures" or recognizable features of a particular steganographic method.

Destroying the presence of embedded data without destroying the perceptual quality of the stego-image can be a trivial or a very difficult task depending on the steganographic method employed to embed the data. (Destroying the embedded data and the stego-image at the

same time is a trivial problem; simply erase the whole image.) For any LSB embedding or simple bit-wise modulation schemes, destruction of the message can be performed by zeroing the entire LSB plane. For attacking non-robust steganographic methods, anti-watermarking software such as Stirmark [9] or UnZign [8] has been shown to be effective in destroying an embedded message. Destroying a robust embedded message without appreciably damaging the stego-image may be a challenge because the goal of the design of robust watermarking techniques is to ensure that the watermark may be removed only by significant damage to the stego-image. Often a series of transformations are used hoping that while a technique may be robust to each transformation applied independently, the combination of transformations will overwhelm the robustness of the method and destroy the message or mark while leaving the quality of the stego-image acceptible. Known-message and known-cover attacks increase the ability of the steganalyst to remove the message without damaging the stego-image because the modulation performed by the steganographic technique can be characterized and then removed. A chosen-message attack allows the steganalyst to characterize the distortions applied by the method under study.

The most difficult attacks are modifying the embedded data in a stego-image and deducing the stego-key used to embed the data. The latter should be treated as a complete failure of the steganographic method because the attacker becomes capable of generating messages that appear to originate from the original sender. The chosen-stego and chosen-message attacks are often employed for this kind of steganalysis. Many steganographic techniques employ methods developed in cryptography so these attacks can resemble attacks against cryptographic systems.

Conclusion

An overview of steganography was presented along with applications that can benefit from the technology. Features of stenographic systems were also discussed, followed by general descriptions of how current systems work. Finally, an overview of steganalysis was presented. Immense research in steganography continues to expand the perceptual transparency, robustness and capacity of information hiding systems.

Since ancient times, man has found a desire in the ability to communicate covertly. The recent explosion of research in watermarking to protect intellectual property is evidence that steganography is not just limited to military or espionage applications. Steganography, like cryptography, will play an increasing role in the future of secure communication in the "digital world."

References

1. N. Johnson and S. Jajodia, "Exploring steganography: seeing the unseen," *IEEE Computer*, pp. 26-34, February 1998.
2. D. Kahn, *The Codebreakers*, Macmillian, New York, 1967.
3. B. Norman, *Secret Warfare*, Acropolis Books, Washington D.C., 1973.
4. W Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Systems Journal*, Vol. 35, No. 3 and 4, pp. 313-336, 1996.
5. M. Swanson, M. Kobayashi, and A. Tewfik, "Multimedia data embedding and watermarking technologies," *Proceedings of the IEEE*, Vol. 86, No. 6, pp. 1064-1087, June 1998.
6. R. Wolfgang, C. Podilchuk and E. Delp, "Perceptual watermarks for images and video," to appear in the *Proceedings of the IEEE*, May, 1999. (A copy of this paper is availavle at: http://www.ece.purdue.edu/~ace).
7. R. B. Wolfgang and E. J. Delp, "Fragile watermarking using the VW2D watermark," *Proceedings of the SPIE/IS&T Conference on Security and Watermarking of Multimedia Contents*, SPIE Vol. 3657, San Jose, CA, January 1999.
8. UnZign software: http://altern.org/watermark, 1997.
9. Stirmark software: http://www.cl.cam.ac.uk/~fapp2/watermarking/stirmark, 1997.
10. N. Johnson and S. Jajodia, "Steganalysis of images created using current steganography software," *Lecture Notes in Computer Science*, Vol. 1525, pp. 273-289, 1998.
11. Steganos Software: http://www.demcom.com/english/steganos/index.htm
12. I. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions On Image Processing*, Vol. 6, No. 12, pp. 1673-1687, December 1997.
13. I. Cox and M. Miller, "A review of watermarking and the importance of perceptual modeling," *Proceedings of the SPIE/IST&T Conference on Human Vision and Electronic Imaging II* , SPIE Vol. 3016, San Jose, CA, pp. 92-99, February 1997.
14. M. Swanson, B. Zhu, and A. Tewfik, "Robust data hiding for mages," *Proceedings of the IEEE DSP Workshop*, Leon, Norway, pp. 37-40, Loen, Norway, September 1996.
15. R. Anderson and F. Petitcolas, "On the limits of steganography," *IEEE Journal on Special Areas in Communications*, Vol. 16, No. 4, pp. 463-473, May 1998.
16. J. Smith and B. Comiskey, "Modulation and information hiding in images," *Lecture Notes in Computer Science*, Vol. 1174, pp. 207-226, 1996.
17. F. Petitcolas and R. Anderson, "Weaknesses of copyright marking systems," *Proceedings of the ACM Multimedia and Security Workshop* (at ACM Multimedia '98*)*, pp. 55-62, Bristol, United Kingdom, September 1998.