Steganography
(Steganalysis)

Victor Habrahamashon
CMPT-495
Dr. Robila

Abstract. Steganography encompasses methods of transmitting secret message through innocuous cover carriers in such a manner, that the existence of the embedded message is undetectable. Carriers of such message may resemble images, audio, video or any digitally represented code of transmission. Steganalysis the science utilized to disrupt the transmission of steganographic encrypted messages, through detection, extraction, disabling or destruction of such hidden information.

Introduction

This paper will surmise the many uses for steganography and the recent science of steganalysis. Steganography differs from the better known practice of cryptography, in which a message is purposely garbled and can only be decipher by those who posses the key. There are a several steganographic methods that we have witness (especially if you read any of the Tom Clancy or James Patterson books), ranging from invisible ink and microdots to secreting a hidden message in the second letter of each word of a large body of text and spread spectrum radio communication. With computers and networks, there are many other ways of hiding information, such as:

- Hidden text within Web pages
- Hiding files in "plain sight"
- Null ciphers (e.g., using the first letter of each word to form a hidden message in an otherwise innocuous text)

On the Internet, spam is a potential carrier medium for hidden messages. Consider the following:

> **Dear Friend, This letter was specially selected to be sent to you! We will comply with all removal requests ! This mail is being sent in compliance with Senate bill 1621 ; Title 5 ; Section 303 ! Do NOT confuse us with Internet scam artists . Why work for somebody else when you can become rich within 38 days ! Have you ever noticed the baby boomers are more demanding than their parents & more people than ever are surfing the web ! Well, now is your chance to capitalize on this ! WE will help YOU sell more & SELL MORE . You can begin at absolutely no cost to you ! But don't believe us ! Ms Anderson who resides in Missouri tried us and says "My only problem now is where to park all my cars" . This offer is 100% legal . You will blame yourself forever if you don't order now ! Sign up a friend and your friend will be rich too . Cheers ! Dear Salaryman , Especially for you - this amazing news . If you are not interested in our publications and wish to be removed from our lists, simply do NOT respond and ignore this mail ! This mail is being sent in compliance with Senate bill 2116 , Title 3 ; Section 306 ! This is a ligitimate business proposal ! Why work for somebody else when you can become rich within 68 months ! Have you ever noticed more people than ever are surfing the web and nobody is getting any younger ! Well, now is your chance to capitalize on this . We will help you decrease perceived waiting time by 180% and SELL MORE . The best thing about our system is that it is absolutely risk free for you ! But don't believe us ! Mrs Ames of Alabama tried us and says "My only problem now is where to park all my cars" . We are licensed to operate in all states ! You will blame yourself forever if you don't order now ! Sign up a friend and you'll get a discount of 20% ! Thanks ! Dear Salaryman , Your email address has been submitted to us indicating your interest in our briefing ! If you no longer wish to receive our publications simply reply with a Subject: of "REMOVE" and you will immediately be removed from our mailing list . This**

mail is being sent in compliance with Senate bill 1618 , Title 6 , Section 307 . THIS IS NOT A GET RICH SCHEME . Why work for somebody else when you can become rich within 17 DAYS ! Have you ever noticed more people than ever are surfing the web and more people than ever are surfing the web ! Well, now is your chance to capitalize on this ! WE will help YOU turn your business into an E-BUSINESS and deliver goods right to the customer's doorstep ! You are guaranteed to succeed because we take all the risk ! But don't believe us . Ms Simpson of Wyoming tried us and says "Now I'm rich, Rich, RICH" ! We assure you that we operate within all applicable laws . We implore you - act now ! Sign up a friend and you'll get a discount of 50% . Thank-you for your serious consideration of our offer .

This message looks like typical spam, which is generally ignored and discarded. This message was created at spam mimic, a Website that converts a short text message into a text block that looks like spam using a grammar-based mimicry idea first proposed by Peter Wayner (spam mimic 2003; Wayner 2002). The reader will learn nothing by looking at the word spacing or misspellings in the message. The zeros and ones are encoded by the choice of the words. The hidden message in the spam carrier above is:

Meet at Main and Willard at 8:30

Special tools or skills to hide messages in digital files using variances of a null cipher are not necessary. An image or text block can be hidden under another image in a PowerPoint file, for example. Messages can be hidden in the properties of a Word file. Messages can be hidden in comments in Web pages or in other formatting vagaries that are ignored by browsers (Artz 2001). Text can be hidden as line art in a document by putting the text in the same color as the background and placing another drawing in the foreground. The recipient could retrieve the hidden text by changing its color (Seward 2004). These are all decidedly low-tech mechanisms, but they can be very effective.

Steganography today, however, is significantly more sophisticated than the examples above suggest, allowing a user to hide large amounts of information within image and audio files. These forms of steganography often are used in conjunction with cryptography so that the information is doubly protected; first it is encrypted and then hidden so that an adversary has to first find the information (an often difficult task in and of itself) and *then* decrypt it.

Carriers of such messages may resemble innocent images, audio, video, text, or any other digitally represented code or transmission. The hidden message may be plaintext, ciphertext, or anything that can be represented as a bit stream. Commercial use of steganographic techniques has evolved into digital watermarking. Watermarking does not necessarily conceal the knowledge of the hidden information other than from the human senses.

Definition

Steganography is the art of hiding information in a manner that makes the detection of the hidden information virtually impossible. Steganography derived from Greek literally means “covered writing.” It utilizes a variety of methods to conceal

information from prying eyes. These include invisible inks, microdots, character arrangement, digital signatures, covert channels, and spread spectrum communications.

Steganalysis is the technology that attempts to defeat steganography—by detecting the hidden information and extracting or destroying it.

Terminology

A message is the information hidden and may be plaintext, ciphertext, images, or anything that can be embedded into a bit stream. Together the cover carrier and the embedded message create a stego-carrier. Hiding information may require a key which is additional secret information, such as a password, required for embedding the information. A possible formula of the process may be represented as:

Cover medium + embedded message + stegokey = stego-medium (1)

In steganography comparisons are made between the cover-media, and the stego-media. Techniques available to steganalyst to break a stegangraphy encrypted message are similar to the ones available to the cryptanalysis. These are stego-only, known cover, known message, chosen stego, and chosen message. These are defined as follows:

> A *stego-only attack* is similar to the ciphertext only attack where only the stego-medium is available for analysis. If the "original" cover-media and stego-media are both available, then a *known cover attack* is available. The steganalysis may use a *known message attack* when the hidden message is revealed at some later date, an attacker may attempt to analyze the stego-media for future attacks. Even with the message, this may be very difficult and may even be considered equivalent to the stego-only attack. The *chosen stego attack* is one where the steganography tool (algorithm) and stego-media are known. A *chosen message attack* is one where the steganalyst generates stego-media from some steganography tool or algorithm from a known message. The goal in this attack is to determine corresponding patterns in the stego-media that may point to the use of specific steganography tools or algorithms.

There are other terms introduced with the technology that need to be mentioned. These words though common take on a different role when used in speaking of steganography or steganalysis:

Payload: the data that is desirable for transport (i.e. documents, audio, video, or drawings).

Carrier: the signal, stream or data file into which the payload is to be h hidden.

Channel: the type of input, such as JPEG, BMP image or Wave files.

Package: the resulting signal, stream or data file which has the payload encoded.

Encoding density: the percentage of bytes which are modified to encode the payload, typically as a floating-point number between 0 and 1.

Applications

The current increase of internet traffic and sharing of information of public as well as private information over the internet cloud has prompted added interest in steganography and steganalysis. This is evident in the publishing and recording industries as well as in our law enforcement agencies. The publishing and recording industries utilize a watermarking technique in an attempt to protect copyrighted materials such as text, images, audio and video. The financial and legal communities utilize it to exchange sensitive material over the internet. Law enforcement agencies utilize steganalysis to detect embedded hidden messages, illegal in nature.

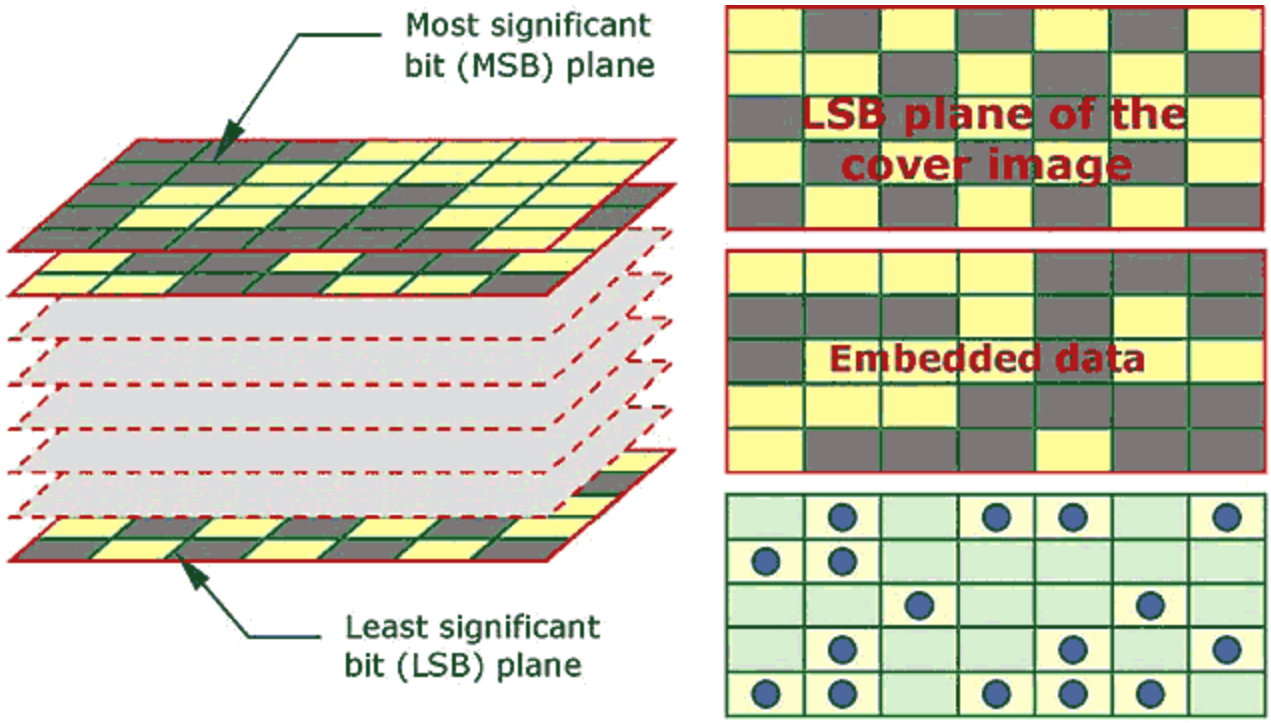
Methods

There are many ways in which messages can be hidden in digital media. Programs are written to access slack and unallocated space directly. Images are the most popular cover media for steganography and can be stored in a straightforward bitmap format (such as BMP) or in a compressed format (such as JPEG). Palette images are usually in the GIF format. Information hiding is accomplished either in the space domain or in the frequency domain. In terms of insertion schemes, several methods (such as substitution, addition, and adjustment) can be used. One adjustment approach is Quantization Index Modulation (QIM), which uses different quantizers to carry different bits of the secret data [2]. Although a simple unified method for classifying these techniques does not exist, some popular approaches are used in downloadable steganographic tools some of which are listed in the table below.

EzStego	online.securityfocus.com/tools/586/scoreit/
F5	wwwrn.inf.tu-dresden.de/~westfeld/f5.html
Hide and Seek v4.1	ftp://ftp.csua.berkeley.edu/pub/cypherpunks/steganography/
Hide and Seek for Win95	ftp.hacktic.nl/pub/crypto/incoming/
Hide4PGP	www.heinz-repp.onlinehome.de/Hide4PGP.htm
Jpeg-Jsteg	ftp://ftp.funet.fi/pub/crypt/steganography/
Mandelsteg	ftp://idea.sec.dsi.unimi.it/pub/security/crypt/code/
MP3Stego	www.cl.cam.ac.uk/~fapp2/steganography/mp3stego/
OutGuess	www.outguess.org/download.php
Steganos	www.steganos.com/en/
S-Tools v4	members.tripod.com/steganography/stego/s-tools4.html
White Noice Storm	ftp://ftp.esua.berkeley.edu/pub/cypherpunks/steganography/

For a steganography tool table, see
www.jjtc.com/Steganography/toolmatrix.htm

LSB modification *is* based on modifying the least significant bits (LSBs), of the pixel values in the space domain. In a basic implementation, these pixels replace the entire LSB-plane with the stego-data; on average, 50% of the LSBs are flipped (as shown in the figure below). With more sophisticated schemes in which embedding locations are adaptively selected, depending on human vision characteristics. Popular tools include EzStego, S-Tools, and Hide and Seek. In general, simple LSB embedding is susceptible to image processing.



A basic LSB approach. Bit-planes of a grayscale image are sketched on the left with MSB on top. Dark and light boxes represent binary values 0s and 1s, respectively, of the pixels on different bit-planes. The LSB-plane of the cover image on the top right is replaced with the hidden data in the middle, which becomes the LSB-plane of the stego-image. The bottom-right map indicates differences between LSB planes of the cover- and stego-images. Circles represent the flipped bits; with an average of 50% bits in the LSB plane changed, the stego-image is visually identical to the cover.

The following simple example of a 3 pixels grid of a 24 bit color image, using 9 bytes of memory illustrate the LSB method:

(00100111 11101001 11001000)
(00100111 11001000 11101001)
(11001000 00100111 11101001)

When the character **A**, which binary value equals **10000001**, is inserted, the following grid results:

(00100111 11101000 11001000)
(00100110 11001000 11101000)
(11001000 00100111 11101001)

In this case, only three bits needed to be changed to insert the character successfully.

Masking approaches are similar to visible watermarking in which pixel values in masked areas are raised or lowered by some percentage. Reducing the increment to a certain degree makes the mark invisible. In the patchwork method, pairs of patches are selected pseudo-randomly; pixel values in each pair are raised by a small constant value in one patch and lowered by the same amount in the other.

Transform domain techniques is widely used for robust watermarking. Similar techniques can also realize large-capacity embedding for steganography. Candidate transforms include discrete cosine transform (DCT), discrete wavelet transforms (DWT), and discrete Fourier transforms (DFT). By being embedded in the transform domain, the hidden data resides in more robust areas, spread across the entire image, and provides better resistance against signal processing.

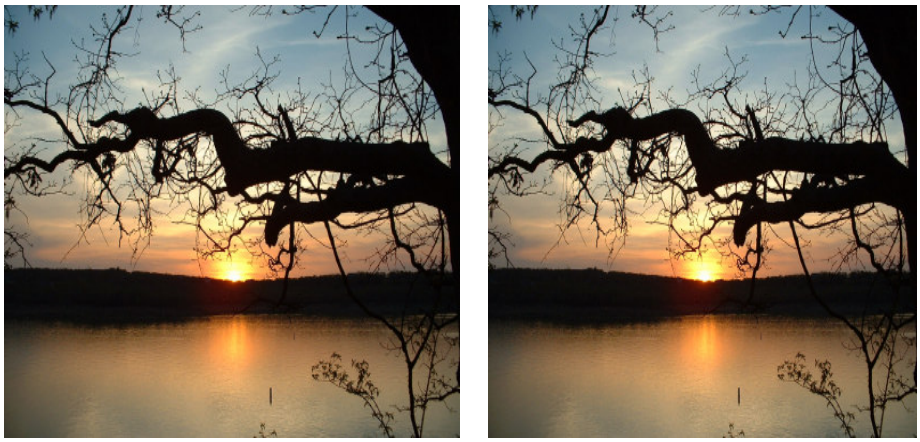
Techniques incorporated in compression algorithms such as to integrate the embedded data with an image-compression algorithm like a JPEG. For example, the steganographic tool Jpeg-Jester takes a lossless cover-image and the message to be hidden to generate a JPEG stego-image. In the coding process, DCT coefficients are rounded up or down according to individual bits to be embedded. Such techniques are attractive because JPEG images are popular on the Internet.

Spread-spectrum techniques spread data throughout the cover-image (such as frequency hopping). A stego-key is used for encryption to randomly select the frequency channels. White Noise Storm is a popular tool using this technique. With embedded data as the object to be transmitted, the cover-image is viewed as interference in a covert communication framework. The embedded data is first modulated with pseudo-noise so the energy is spread over a wide frequency band, achieving only a very low level of embedding strength. This is valuable in achieving imperceptibility.

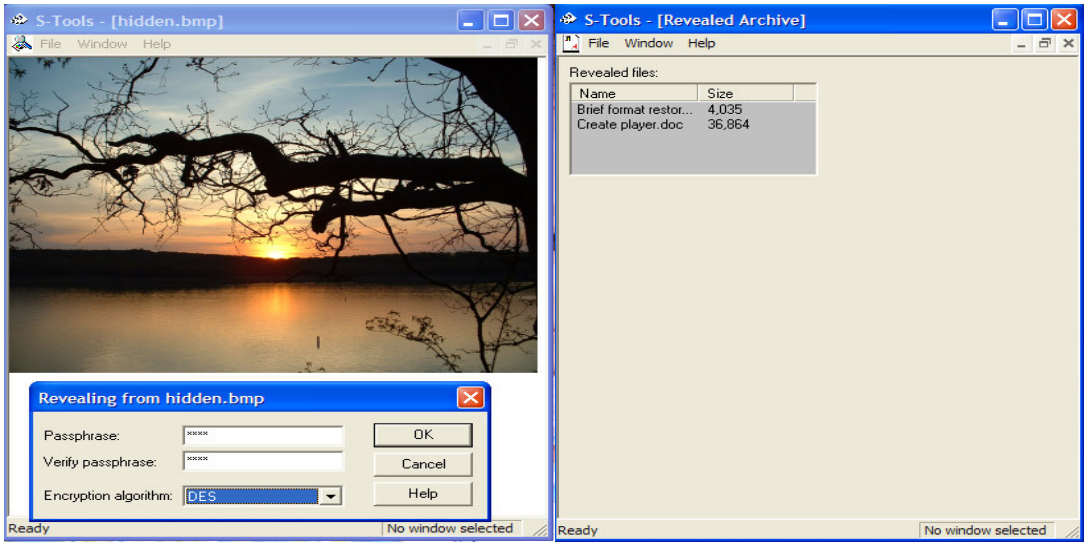
Detection

Despite the fact steganography tools alter only the least-significant image component, they leave detectable traces in the stego image, allowing for the possible detection of hidden information. Steganalysis breakdown into two major types of techniques; visual analysis and statistical analysis. Visual analysis tries to reveal the presence of secret communication through inspection. This is accomplished by the naked eye or assistance of a computer of the suspect file. The comparison of original files with the suspect file if possible would give some merit to suspicions.

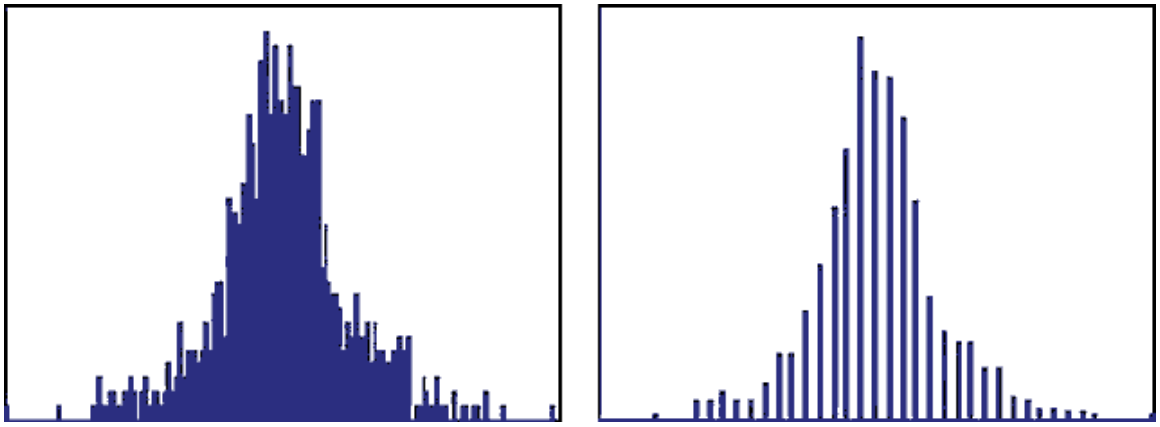
The following are two images the one on the right is the original and the on the left holds documents hidden within it by Andy Brown's S-Tool program.



The two are identical in every way inspection by the naked eye reveals nothing. But as you can see from the following screen shot of S-Tools there are file hidden in the image. Some embedding tools cause subtle changes in the set of possible values that may be taken by the pixel gray levels and/or transform coefficients.



A histogram analysis may be used to reveal such a signature (see Figure below); the left histogram represents the distribution of a particular group of coefficients taken from a cover-image generated by the double transform scheme. In a stego-image, the discreteness of the histogram on the right is a clear sign of QIM embedding.

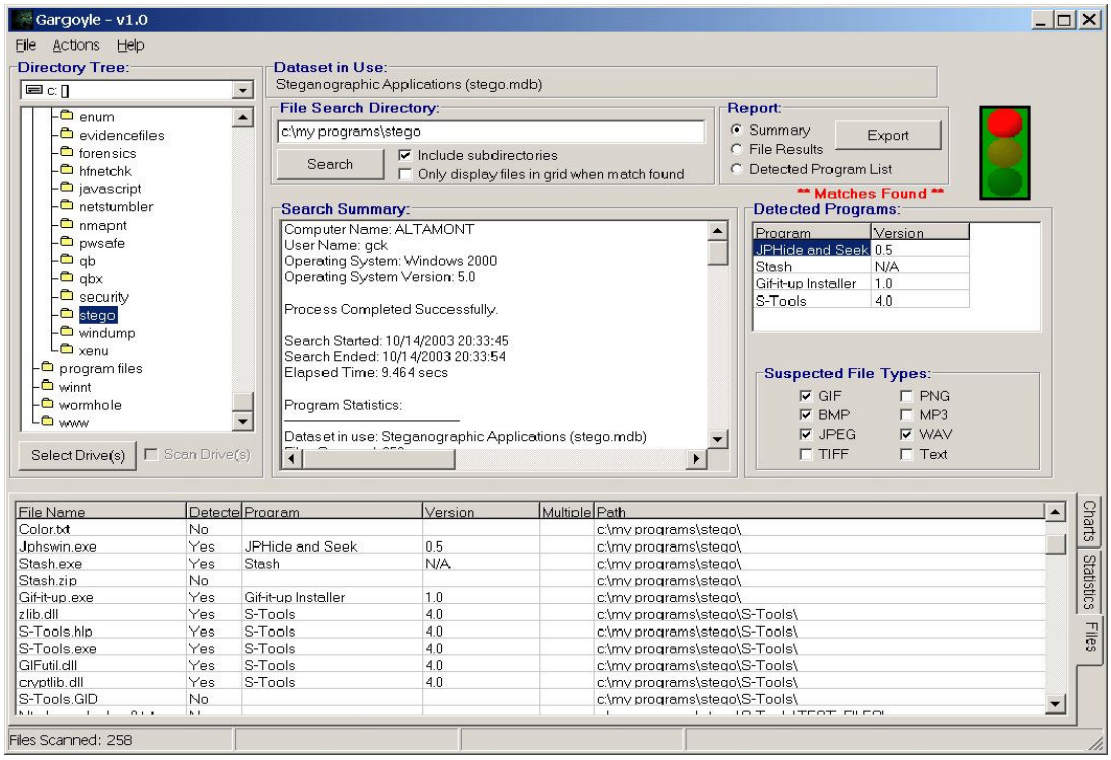


Statistical analysis is more powerful since it reveals tiny alterations in an image's statistical behavior caused by steganographic embedding. As there is a range of approaches to embedding, each modifying the image in a different way, unified techniques for detecting hidden information in all types of stego-images are difficult to find. The nominally universal methods developed to detect embedded stego-data are generally less effective than the steganalytic methods aimed at specific types of embedding.

Because detecting stego content is performed only with current steganalytic approaches, any system considered secure today may be broken tomorrow using new techniques. Some of the most popular steganalytic tools are Stego Suite by Wetstone, Stegdetect by Neils Provos, and other methods are outlined in the table below.

Steganalytic Methods	Description	Targeted Steganographic Techniques
RS steganalysis	Sensitivity of dual statistics based on spatial correlation of pixels to LSB randomization due to steganographic embedding is used in analysis.	Various LSB modification techniques
PoV-based Chi-square test	A Chi-square test checks whether the occurrence of each pair of values tends to become equal, indicating some data is embedded.	Steganography based on swapping pairs of values of pixel gray levels, colors, or DCT coefficients
Palette checking	Peculiarity in palette ordering is a clear sign of systematic modification.	Steganography in palette images
RQP method	Method based on analyzing the increased number of close-color pairs caused by embedding.	LSB embedding in true-color images
Check JPEG compatibility	Method detects unusual departure from the JPEG signature inherent in images initially stored in JPEG format.	Space-domain steganography using images initially stored in the JPEG format
Histogram analysis	Method reveals discreteness or periodicity in particular coefficients due to quantization-related modification.	QIM or other quantization-related embedding methods
Universal blind detection	Statistical quantities constructed using high-order statistics, and a detection model established with the threshold obtained in a training process.	Various steganographic techniques

WetStone Technologies' Gargoyle (formerly StegoDetect) software (WetStone Technologies 2004A) can be used to detect the presence of steganography software. Gargoyle employs a proprietary data set (or hash set) of all of the files in the known steganography software distributions, comparing them to the hashes of the files subject to search. The image below shows the output when Gargoyle was aimed at a directory where steganography programs are stored.



The following is a screen shot of the output produced by Stegdetect which utilizes a web crawler to capture images from the internet for detection of hidden communication.

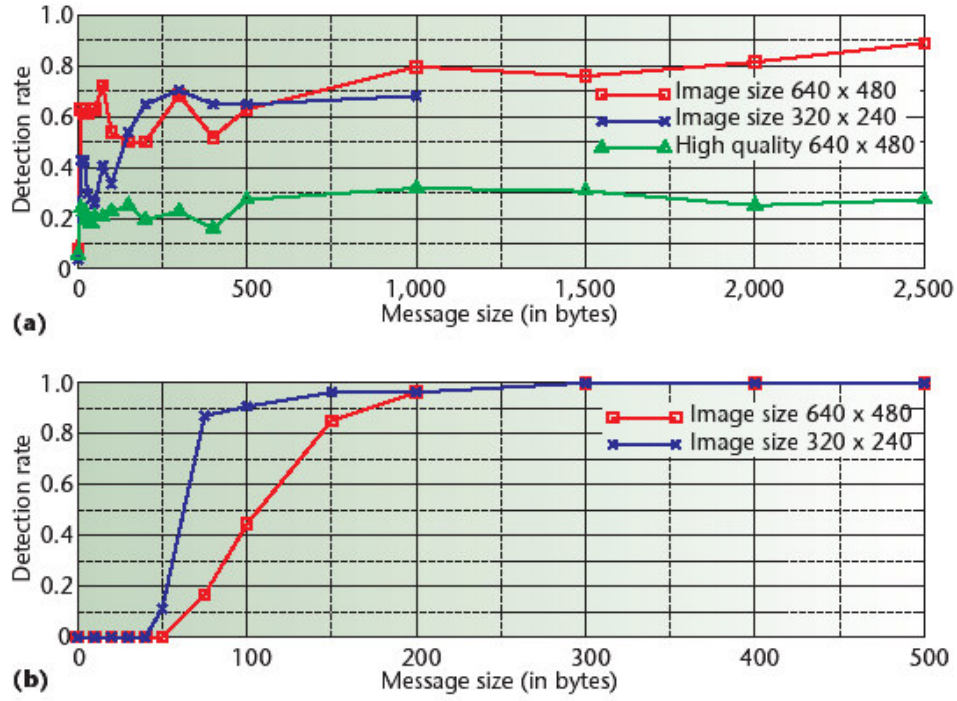


Figure 11. Using Stegdetect over the Internet. (a) JPHide and (b) JSteg produce different detection results for different test images and message sizes.

Blind detection of hidden information in apparently innocuous digital media is generally more challenging than data embedding, especially when the embedding rate is low, as steganalysts always work in passive mode. Another important consideration in steganalysis is to keep the computational complexity sufficiently low, allowing the screening of thousands (even millions) of suspected images in a reasonably short amount of time. The computation limitation may be less stringent for steganography, since in practical applications the embedding algorithm is executed on only a few images taken from a large database.

Legal Aspects

The most effective movement of individuals against the ban on strong encryption products began with a lawsuit in 1995. Bernstein, a previous graduate student of Berkeley, and Gilmore, a supporter of the Electronic Frontier Foundation, filed a case based on the grounds that computer source code is speech. Therefore, the government's limitations on the public's use of encryption products are a violation of the Free Speech Amendment. The case was settled in 1999, with Bernstein winning, and the government now allows public access to steganography applications.

Conclusion

There is a relatively high use of stegonagraphy on the Internet, and the creation of stegonagraphy monitoring and detection systems is important. Apart from their law enforcement/intelligence and anti-terrorist significance, steganographic techniques also have peaceful applications, including: in-band captioning; integration of multiple media for convenient and reliable storage, management, and transmission; embedding executables for function control; error correction; and version upgrading. Computer specialists, signal-processing researchers, and information security professionals should expect to devote much more attention to the challenging area of information hiding and detection. Most recently a University of Delaware research team received National Science Foundation funding to combat terrorism by developing techniques to detect the use of steganography, this reinforces the importance of the science of steganalysis.

Reference:

1. Cyber Warfare: Steganography vs. Steganalysis by Huaiqing Wang and Shuozhong Wang October 2004
2. Neil F. Johnson and Sushil Jajodia Center for Secure Information Systems, George Mason University, MS:4A4, Fairfax, Virginia 22030-4444 Steganalysis: The Investigation of Hidden Information
3. *R. Chandramouli and K.P. Subbalakshmi Department of ECE Stevens Institute of Technology* CURRENT TRENDS IN STEGANALYSIS: A CRITICAL SURVEY
4. Hiding in Plain Sight, Steganography and the Art of Covert Communications by Eric Cole
5. http://en.wikipedia.org/wiki/
6. www.stegoarchive.com
7. www.FBI.gov