Global Information Assurance Certification Paper

The Art of Steganography

GSEC Practical (v.1.4b)
By Deborah A. Whitiak

Abstract

"The goal of steganography is to hide messages inside other harmless messages in a way that does not allow any enemy to even detect that there is a second secret message present." {1}. Steganography can be used as a means to transmit and hide messages without the fact of the transmission being discovered. If suspicion is raised then the goal of steganography is defeated. A mere picture of your pet, house, or most recent vacation could conceal the plans for the next terrorist attack.

Purpose of this Paper

This paper aims to explain steganography by providing a brief history, early research, advantages, disadvantages, and also to introduce a few methods of hiding data such as: text, images, and audio. In addition, the paper will review the detection of steganography. All research for this paper has been conducted through the Internet, which is relied upon for up-to-date information.

History

Steganography literally means "covered writing" {2}. It has been a favorite method of hiding information by military and political leaders for thousands of years. Early records indicate that Greek historians used steganographic methods. When the Greek tyrant Histiaeus was held as a prisoner by king Darius in Susa during the 5th century BCE, he used a method of tattooing a secret message on the head of a trusted slave to send messages to his son-in-law Aristagoras in Miletus. The head of the trusted slave was shaved and the message was tattooed on his scalp. When the slave's hair had grown he was sent to deliver the message {3}.

Another method from ancient Greece was to etch a message on a tablet and then cover the tablet with wax. Demeratus, a Greek, needed to notify Sparta that Xeres intended to invade Greece. To avoid capture, he scraped the wax off of the tablets and wrote the message on the underlying wood. Then he covered the tablets with wax again. The tablets appeared to be blank and unused so they passed inspection.

Invisible inks have always been a popular method of steganography. Ancient Romans used to write between the lines of text using invisible inks based on

readily available substances such as fruit juices, urine, and milk. When heated, the invisible inks would darken, and become legible. Invisible inks were used as recently as World War II.

It has been noted that the Abul Nidal organization and Bin Laden's al Qa'ida organization were using computerized Internet files by methods of e-mail, steganography, and encryption to communicate to their operations. It has been reported that the alleged hijackers in the September 11th attacks had Internet e-mail accounts and were using them to communicate with each other. Mohammed Atta, one of the alleged hijackers was repeatedly seen in a Florida library downloading pictures of children and Middle Eastern scenes which authorities suspect he used as secret method of communication {4}.

Research

An early researcher in steganography was Johannes Trithemius (1462-1526), a German monk. *Steganographia*, his first work, described systems of magic and prophecy, but it also contained a complex system of cryptography (the method of encryption and decryption of messages). One of the earliest books on steganography was a four hundred-page work written by Gasperi Schotti in 1665 and called *Steganographica.* Most of his ideas came from Trithemius. Due to the wrath of powerful factions, preventing publications, authors of such books often concealed their names in their works.

Bishop John Wilkins, a master at Trinity College, Cambridge, later devised a concept of coding messages in music and string knots to invisible inks. He described the principles of cryptanalysis by letter frequencies. He applied the idea of coding messages to well-known concepts instead of magic and prophecy. He then argued against those who opposed publication in the field {5}.

Advantages

Why use steganography? The advantage of using steganography is to conceal information. The transmission of messages is transparent to any given viewer. Messages can be concealed in different formats that are undetectable and un-readable to the human eye. Steganographic technologies are very important in Internet privacy today. With the use of steganography and encryption, corporations, governments, and law enforcement agencies can communicate secretly.

Encryption protects data and can be detected; the only thing missing is the secret key for decryption. Steganography is harder to detect under traditional traffic-pattern analysis {6}. Steganography enhances the privacy of personal communication. Since encryption can be detected and some governments prohibit the use of encryption, steganography can be used to supplement encryption {7}. Additional layers of security are a benefit to secrecy. If a

steganogrphic message is detected, there still is the need for the encryption key. A hidden message need not be encrypted to qualify as steganography. The method of encrypting a message and then using steganography is most widely used by steganographers.

Disadvantages

However, there are disadvantages to mention. One of the biggest disadvantages is that quite frequently the size of a stegnated image is usually larger that the original image. There can be color changes, especially evident if well-know images are chosen as the steganographic cover. Images can be degraded when trying to analyze them.

Another issue to mention, text messages are limited in size for the hiding of data. They need redundant data to replace a secret message. Changing the type of the format or replacing the readable text can alter text messages.

Through the use of new technology, some Internet firewalls can detect steganographic messages. As this technology evolves detecting steganographic messages can be a drawback because an important message may be deleted or quarantined and this message may be the one that will save a country.

Hiding Data in Text

Text can be used as a method to transmit secret messages using steganography. This can be accomplished through the shifting of original text to the left or right, up or down, or changing the vertical/horizontal length of certain characters. All methods using text require that knowledge of the original text be known in order to be able to successfully extract the secret message. Only small amounts of data can be hidden when hiding data in text. Thus, this method is known to have a common low data rate. The following text methods, line-shift coding, word-shift coding, feature specific coding, semantic, syntactic are described in the next paragraphs.

As mentioned in Duncan Sellars', "An Introduction to Steganography", line-shift coding vertically shifts text lines 1/300 of an inch up or down. This method appears to be the most visible to the reader. Carefully examining a text document can indicate if the lines are up or down from the document's stationary lines. This method is most difficult to retrieve a secret message, as modifying the document can damage the original secret message.

Word-shift coding resembles line-shift coding. Horizontal locations of text are shifted to the left or right of the text lines. This method is less detectable than line-shift coding, but a natural spacing appearance needs to be maintained to accomplish this method such as text justification.

Feature specific coding encodes a secret message into formatted text by altering the size of certain characters in a horizontal/vertical format. Bits can be encoded into text by altering the size of characters such as b, d, t, T, etc. This is accomplished by extending or shortening the upward vertical lines of these characters.

Syntactic coding utilizes punctuation and contractions. Error in punctuation can be easily visible and changing the order of words in a text message can definitely change its' meaning.

Lastly, semantic coding manipulates words by replacing them with synonyms. A problem arises here because there are times when words cannot be exchanged for equivalent ones. The words "sick" and "ill" are synonyms. Both appear as adjectives most of the time. Only the word "sick" is used as a noun, and then it is usually preceded by the word "the". The word "ill" has some meanings that are not the same as "sick" (synonymous to "bad").

Hiding Data in Images

Images are arrays of numbers that represent light intensities at various points, or pixels. Digital images are stored in either 24-bit or 8-bit per pixel. Taking advantage of these pixels, messages can be encoded into an image {8}. Three methods of hiding data in images are least significant bit (LSB) insertion, masking and filtering techniques, and algorithms and transformations.

Least significant bit (LSB) insertion is most widely used, since it is the most simple to accomplish. Data is inserted in the least and the second to least bits of pixels. This method can be detected because of the data changes to pixels, which may result in change of color. Selection of the cover image needs to be considered so that any color degradation would not be noticeable. The use of famous images should be avoided for this reason.

Masking and filtering utilize digital watermarking (a pattern of bits entered into an image file that identifies the file's copyright information, author, rights, etc.). The use of digital watermarking comes from the method of watermarks on stationary. Watermarking is not steganography. Steganography conceals data in images, whereas in watermarking, the actual bits are scattered throughout the image in a way that it cannot be manipulated or identified.

Lastly, the use of algorithms and transformations are used to transmit JPEG images across the Internet. One example of transformation is spread-spectrum. Spread-spectrum communication transmits a narrowband signal over a larger bandwidth so that the spectral density in the channel looks like noise. Other algorithms and transformations can change luminance, hide information by dividing available bandwidth into multiple channels, and altering the brightness of pixel blocks.

Hiding Data in Audio

Hiding data in audio, exploits how the human auditory system (HAS) interprets sounds. This method becomes especially challenging, since the HAS is extremely sensitive. The HAS drowns quiet sounds and emphasizes larger sounds. The goal of steganography in audio is to exploit this weakness. Bits that encode sound outside the range of human hearing can be encoded with covert data {9}. The following paragraphs discuss steganographic methods using audio.

Low-bit encoding encodes a binary string in the least significant bit (LSB) of an audio file. Although large amounts of data can be encoded in an audio signal, this method can be easily destroyed by channel noise. Data is easily destroyed in transmissions other than digital to digital {10}.

Phase coding substitutes the phase of an audio segment with a reference phase that represents the data. This technique is based on the HAS sensitivity for phase variation. A sound file is divided into blocks and each block initial phase is modified using an embedded message.

Spread-spectrum encodes streams of information by spreading data across as much of the frequency spectrum as possible. This allows the signal reception, even if there is interference on some frequencies. These usually add noise to sound. Embedded signals can be filtered through a perceptual mask, so that the most audible components of the added noise are reduced in power.

Echo data hiding embeds data into a host audio signal by introducing an echo. By varying three parameters of the echo, the initial amplitude, the delay rate, and the offset, the data are hidden. The human ear cannot distinguish between two signals as the offset between the original and the echo decrease and the two signals blend. The echo is perceived as resonance. This technique has a high success rate.

Embedding Data

The goal of steganography is to conceal data. There are a few features and restrictions to successfully hide data. "The goal is for the data to remain "*hidden.*" {11}. The word "*hidden*" has two meanings here, (1) the data can be "*hidden*" and not visible to the human eye (2) the data can be visible and still not visible to the human eye. If the focus is deterred from the data, the data will not be seen, which means that it is *"hidden"*. The following guidelines represent a few features and restrictions when embedding data.

- The cover media should not be degraded by the embedded data. It should appear that the cover media does not look distorted. It should not have a noticeable change in color composition, or that of the luminance. Frequently a cover media's size becomes enlarged; this can look very suspicious.

- Embedded data should be directly embedded into the cover media not in the header or wrapper. The embedded data needs to remain intact across different file formats.

- The embedded data should be immune to any manipulation. This ranges from any intentional modification or any modification through transmission.

- Error correcting codes should be inserted in the cover media to ensure integrity of data when or if the cover media is modified or tampered with.

- The embedded data should be recoverable and intact if only fragments of the cover media remain.

Detecting Steganography

Attempts to detect the presence of steganographic messages are referred to as "attacks". Usually attacks are either passive or active. In a passive attack, the interceptor is able to intercept the data. In an active attack the interceptor is able to manipulate the data. Whether the attack is passive or active the steganographer must use caution when choosing certain data hiding techniques so that unusual patterns do not stand out to expose the possibility of hidden information.

Shifts in word and line spacing may be difficult to detect hidden information in text. Opening the text with a word processor can easily reveal appended spaces and "invisible" characters.

Images can appear to have distortions. The steganographer needs to ensure that the "picture taken" is the "picture seen". The original image and the stego-image should not have detectable variations in color composition, luminance, and pixel relationships. Steganographers need to avoid well-known images for this reason. A checksum can be embedded into a stego-image to be used as a tampering detector. The drawback here is that there is no way to tell how much the information has been tampered with.

Audio can be detected with "visible noise". Echoes and shadow signals reduce the chance of audible noise but can be detected with little processing.

Today Internet firewall filters have become more sophisticated to detect packets that contain hidden information or invalid information. Software to detect steganography is starting to become more readily available as freeware on the Internet.

Digital Watermarking

Another use of steganography is digital watermarking. Watermarking is a special technique of creating invisible digital marks in images and audio files that carry copyright information. These marks can be detected by special software that can derive a lot of useful information from the watermark, when the file was created, who holds the copyright, how to contact the author, etc.

A large amount of copyrighted material is reproduced today, that is, acquired without purchase. Visual artists today on the Internet risk losing their work to anyone who knows how to right click on a file or image in a browser window. Without the digital watermark to protect their artistry, artists lose out on considerable income in royalties. Artists insert identifying markers into their images to identify and protect their work.

Margaret Thatcher discovered the source of a cabinet leak to the press in the early 1980's using digitally watermarked files distributed to cabinet members. Apparently press leaks of several private cabinet documents prompted her to have code programmed into the word processors to encode their identity in the word spacing. Disloyal ministers could be traced through this.

File sharing sites have increased the need of content producers to defend their material through the use of the digital watermark.

Conclusion

This paper has provided an overview of steganography, which by definition literally means, "covered writing". Some methods discussed were hiding data in text, images, and audio. We also discussed a few detection methods, which are a fraction of the methods currently used to detect steganography.

Steganography tools are becoming abundant and very easy to use. Many Internet sites offer free downloadable software. Often, although it is not necessary, the hidden messages will be encrypted. Encryption paired with steganography creates an extra layer of privacy. The security of the system depends upon the assumption that the enemy has full knowledge of the design and details of the steganography. The missing information is the secret key. The success of steganography is dependent upon selecting the proper cover mechanism.

Law enforcement agents are becoming very concerned about the use of steganography in criminal activity today. The placing of images on Internet web pages, and the sending of audio and text through email transmission over the Internet makes steganography transmission fairly simple {12}.

Also mentioned, is the use of steganography to apply digital watermarking. Visual artists today need to protect their work. File sharing on the Internet is at an all time high. Any type of image, video, music, text, and software can be downloaded form the Internet today. Artists are being robbed of their royalties because of this.

In the future, steganography will be more widely used and become a threat to corporations, law enforcement, and governments. Terrorists in the world will continue to communicate through the use of steganography. These messages will be not be easily detected. Unpopular web pages may hold images that contain hidden messages of the next terrorist attack.

As the economy continues to slump and the job markets decline, fearful employees of job loss may pass confidential corporate information to competitors for payoffs or a new job proposal. Insider trading and stock tips can be passed securely without interception from the SEC. Drug and gambling activity has moved off the streets and onto the Internet, making it easier to communicate worldwide. Hiding information through the use of steganography keeps these deals secret from the DEA, IRS, and ATF.

We know that technology ideas and methods have room for improvement. Steganography methods will be examined and new ideas of "how to" will emerge. Transmission methods will be perfected to ensure data integrity and transfer. New media will be discovered to hide information. Better detection methods will be discovered. Steganography today is just the "ice cap" of what steganography will be in the future.

List of References

1. Sellars, Duncan. "An Introduction to Steganography."
   URL: http://www.cs.uct.ac.za/courses/CS400W/papers99/stego.html (4 February 2003).

2. URL: http://www.webopedia.com/TERM/s/steganography.html (4 February 2003).

3. "Privacy Guide: Steganography."
   URL: http://all-nettools.com/privacy/stegano.htm (10 February 2003).

4. Ross, Brian. "A Secret Language".4 October 2001
   URL:http://abcnews.go.com/sections/primetime/DailyNews/PRIMETIME_011004_steganography (3 March 2003).

5. Sellars, Duncan. "An Introduction to Steganography."
   URL: http://www.cs.uct.ac.za/courses/CS400W/papers99/stego.html (4 February 2003).

6. Radcliff, Deborah. "Steganography: Hidden Data". (10 June 2002)
   URL: http://www.computerworld.com (3 March 2003).

7. Levitt, Jason. "Getting Ahead Of The Privacy Curve With Steganography". (8 December 1997)
   URL: http://www.iweek.com/author/internet9.htm (3 March 2003).

8. Johnson, Neil F., Jajodia, Sushil, "Exploring Steganography: Seeing the Unseen" IEEE Computer, pages 26-34 (February 1998)
   URL: http://www.jjtc.com/pub/r2026.pdf (4 March 2003)

9. Levitt, Jason. "Getting Ahead Of The Privacy Curve With Steganography". (8 December 1997)
   URL: http://www.iweek.com/author/internet9.htm (3 March 2003).

10. Bender, W., Gruhl, D., Morimoto, N., Lu, A. "Techniques for data hiding." Vol 35, Nos 3-4 pages 313-336 (February 1996).
    URL: http://www.almaden.ibm.com/cs/people/dgruhl/313.pdf (25 February 2003).

11. Bender, W., Gruhl, D., Morimoto, N., Lu, A. "Techniques for data hiding." Vol 35, Nos 3-4 pages 313-336 (February 1996).
    URL: http://www.almaden.ibm.com/cs/people/dgruhl/313.pdf (25 February 2003).

12. Bender, W., Gruhl, D., Morimoto, N., Lu, A. "Techniques for data hiding." Vol 35, Nos 3-4 pages 313-336 (February 1996).
URL: http://www.almaden.ibm.com/cs/people/dgruhl/313.pdf (25 February 2003).