STEGANOGRAPHY: AN OVERVIEW

JAMMI ASHOK[1]

[1]Professor and Head , Department of Information Technology,
Geethanjali College of Engg. & Technology, Hyderabad

Y.RAJU[2]

[2] Associate Professor, Department of Information Technology,
Geethanjali College of Engg. & Technology, Hyderabad

S.MUNISHANKARAIAH[3]

[3]Associate Professor, Department of Information Technology,
Geethanjali College of Engg. & Technology, Hyderabad

K.SRINIVAS[4]

[4]Associate Professor, Department of Information Technology,
Geethanjali College of Engg. & Technology, Hyderabad

Abstract

Steganography is a technique of hiding information in digital media in such a way that no one apart from the intended recipient knows the existence of the information. Steganography can be viewed as akin to cryptography. Both have been used throughout recorded history as means to protect information. At times these two technologies seem to converge while the objectives of the two differ. Cryptographic techniques "scramble" messages so if intercepted, the messages cannot be understood. Steganography, in an essence, "camouflages" a message to hide its existence and make it seem "invisible" thus concealing the fact that a message is being sent altogether. An encrypted message may draw suspicion while an invisible message will not. Besides hiding data for confidentiality, this approach of information hiding can be extended to copyright protection for digital media like audio, video, and images.

1. Introduction

Steganography or Stego as it is often referred to in the IT community, literally means, "covered writing" which is derived from the Greek language. Steganography is defined by Markus Kahn as follows, "Steganography is the art and science of communicating in a way which hides the existence of the communication. In contrast to Cryptography, where the enemy is allowed to detect, intercept and modify messages without being able to violate certain security premises guaranteed by a cryptosystem, 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 message present".

In a digital world, Steganography and Cryptography are both intended to protect information from unwanted parties. Both Steganography and Cryptography are excellent means by which to accomplish this but neither technology alone is perfect and both can be broken. It is for this reason that most experts would suggest using both to add multiple layers of security.

Steganographic technologies are a very important part of the future of Internet security and privacy on open systems such as the Internet. Steganographic research is primarily driven by the lack of strength in the cryptographic systems on their own and the desire to have complete secrecy in an open-systems environment.

Many governments have created laws that either limit the strength of cryptosystems or prohibit them completely. This has been done primarily for fear by law enforcement not to be able to gain intelligence by wiretaps, etc. This unfortunately leaves the majority of the Internet community either with relatively weak and a lot of the times breakable encryption algorithms or none at all. Civil liberties advocates fight this with the argument that “these limitations are an assault on privacy”. This is where Steganography comes in. Steganography can be used to hide important data inside another file so that only the parties intended to get the message even knows a secret message exists.

Using Steganography, information can be hidden in carriers such as images, audio files, text files, videos and data transmission. When the message is hidden in the carrier a stego-carrier is formed for example a stego-image. Hopefully it will be perceived to be as close as possible to the original carrier or cover image by the human senses. Images are the most widespread carrier medium.

2. History of Steganography

Steganography has been with us in many forms since the time of the Greek empire. Even the word steganography comes from the Greek *steganos*, hidden or covered, plus *graphein,* to write. Herodotus, the Greek historian recorded the story of a slave used as the medium to transmit the hidden message. The slave's head was shaved and the message tattooed on the bare skull after which the hair was allowed to re-grow. The slave was sent to the message recipient who shaved the slave's head to reveal the message. Hopefully the message was not time-dependent! Lord Robert Baden-Powell, as scout for the British during the Boer War marked the positions of Boer artillery bases by embedding maps into drawings of butterflies. Appearing innocent to a casual observer, certain markings on the wings were actually the positions of the enemy military installations. Later, Axis and Allied spies used invisible inks containing fruit juice or urine to transmit messages that would reveal themselves when heated or when in the presence of ultraviolet light.

In the mid-90s a number of the older techniques of hiding messages inside other messages and even images became more popular with the advent of modern software and powerful computers.

Regardless of the technique used, the key similarity in all cases was that messages were hidden in plain view

3. Types of Steganography

There are basically three types of steganographic protocols used. They are:

- Pure Steganography
- Secret Key Steganography
- Public Key Steganography

Pure Steganography is defined as a steganographic system that does not require the exchange of a cipher such as a stego-key. This method of Steganography is the least secure means by which to communicate secretly because the sender and receiver can rely only upon the presumption that no other parties are aware of this secret message. Using open systems such as the Internet, we know this is not the case at all.

Secret Key Steganography is defined as a steganographic system that requires the exchange of a secret key (stego-key) prior to communication. Secret key Steganography takes a cover message and embeds the secret message inside of it by using a secret key (stego-key). Only the parties who know the secret key can reverse the process and read the secret message. Unlike Pure Steganography where a perceived invisible communication channel is present, Secret Key Steganography exchanges a stego-key, which makes it more susceptible to interception. The benefit to Secret Key Steganography is even if it is intercepted; only parties who know the secret key can extract the secret message.

Public Key Steganography takes the concepts from Public Key Cryptography as explained below. Public Key Steganography is defined as a steganographic system that uses a public key and a private key to secure the communication between the parties wanting to communicate secretly. The sender will use the public key during the encoding process and only the private key, which has a direct mathematical relationship with the public key, can decipher the secret message. Public Key Steganography provides a more robust way of implementing a steganographic system because it can utilize a much more robust and researched technology in Public Key Cryptography. It also has multiple levels of security in that unwanted parties must first suspect the use of Steganography and then they would have to find a way to crack the algorithm used by the public key system before they could intercept the secret message.

4. Steganographic Mediums

- Encoding Secret Messages in Text
- Encoding Secret Messages in Images
- Encoding Secret Messages in Audio

4.1 Encoding secret messages in text

Encoding secret messages in text can be a very challenging task. This is because text files have a very small amount of redundant data to replace with a secret message. Another drawback is the ease of which text based Steganography can be altered by an unwanted parties by just changing the text itself or reformatting the text to some other form (from .TXT to .PDF, etc.). There are numerous methods by which to accomplish text based Steganography.

Line-shift encoding involves actually shifting each line of text vertically up or down by as little as 3 centimeters. Depending on whether the line was up or down from the stationary line would equate to a value that would or could be encoded into a secret message.

Word-shift encoding works in much the same way that line-shift encoding works; only we use the horizontal spaces between words to equate a value for the hidden message. This method of encoding is less visible than line-shift encoding but requires that the text format support variable spacing.

Feature specific encoding involves encoding secret messages into formatted text by changing certain text attributes such as vertical/horizontal length of letters such as b, d, T, etc. This is by far the hardest text encoding method to intercept as each type of formatted text has a large amount of features that can be used for encoding the secret message.

4.2 Encoding secret messages in images

Coding secret messages in digital images is by far the most widely used of all methods in the digital world of today. This is because it can take advantage of the limited power of the human visual system (HVS). Almost any plain text, cipher text, image and any other media that can be encoded into a bit stream can be hidden in a digital image.

Two of the more popular digital image encoding techniques used today. They are least significant bit (LSB) encoding and masking and filtering techniques.

Least significant bit (LSB) encoding is by far the most popular of the coding techniques used for digital images. By using the LSB of each byte (8 bits) in an image for a secret message, you can store 3 bits of data in each pixel for 24-bit images and 1 bit in each pixel for 8-bit images. As you can see, much more information can be stored in a 24-bit image file.
Depending on the color palette used for the cover image (i.e., all gray), it is possible to take 2
LSB's from one byte without the human visual system (HVS) being able to tell the difference.

For example, the following grid can be considered as 3 pixels of a 24 bit color image, using 9 bytes of memory:
(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 needs to be changed to insert the character successfully. On average only half of the bits in an image will need to be modified to hide a secret message using the maximal cover size. The resulting changes that are made to the least significant bits are too small to be recognized by the human eye, so the message is effectively hidden.

While using a 24 bit image gives a relatively large amount of space to hide messages, it is also possible to use a 8-bit image as a cover source. Because of the smaller space and different properties, 8 bit images require a more careful approach. Where 24 bit images use 3 bytes to represent a pixel, an 8 bit uses only one. Changing the LSB of that byte will result in a visible change of color, as another color in the available palette will be displayed. Therefore the cover image needs to be selected more carefully and preferably be in grayscale, as the human eye will not detect the difference between different gray values as easy as with different colors.

The only problem with this technique is that it is very vulnerable to attacks such as image changes and formatting (i.e., changing from .GIF to .JPEG).

Masking and filtering techniques for digital image encoding such as Digital Watermarking (i.e. -integrating a companies logo on there web content) are more popular with lossy compression techniques such as (.JPEG). This technique actually extends an image data by masking the secret data over the original data as opposed to hiding information inside of the data. Some experts argue that this is definitely a form of Information Hiding, but not technically Steganography. The beauty of Masking and filtering techniques are that they are immune to image manipulation which makes there possible uses very robust.

4.3 Encoding secret messages in audio

Encoding secret messages in audio is the most challenging technique to use when dealing with Steganography. This is because the human auditory system (HAS) has such a dynamic range that it can listen over. The only weakness in the (HAS) comes at trying to differentiate sounds (loud sounds drown out quiet sounds) and this is what must be exploited to encode secret messages in audio without being detected.

Three of the more popular encoding methods for hiding data inside of audio are:

- Low-bit encoding
- Phase-coding
- Spread spectrum

Low-bit encoding embeds secret data into the least significant bit (LSB) of the audio file. The channel capacity is 1KB per second per kilohertz (44 kbps for a 44 KHz sampled sequence). This method is easy to incorporate but is very susceptible to data loss due to channel noise and resampling.

Phase coding substitutes the phase of an initial audio segment with a reference phase that represents the hidden data. This can be thought of, as sort of an encryption for the audio signal by using what is known as Discrete Fourier Transform (DFT), which is nothing more than a transformation algorithm for the audio signal.

Spread spectrum encodes the audio over almost the entire frequency spectrum. It then transmits the audio over different frequencies which will vary depending on what spread spectrum method is used. Direct Sequence Spread Spectrum (DSSS) is one such method that spreads the signal by multiplying the source signal by some pseudo random sequence known as a (CHIP). The sampling rate is then used as the chip rate for the audio signal communication.

5. Steganography Techniques

5.1 Historical Steganographic Techniques

Steganography has been widely used in historical times, especially before cryptographic systems were developed. Examples of historical usage include:

Hidden messages in wax tablets: in ancient Greece, people wrote messages on the wood, and then covered it with wax so that it looked like an ordinary, unused tablet.

Hidden messages on messenger's body: also in ancient Greece. Herodotus tells the story of a message tattooed on a slave's shaved head, hidden by the growth of his hair, and exposed by shaving his head again. The message allegedly carried a warning to Greece about Persian invasion plans. This method has obvious drawbacks:

It is impossible to send a message as quickly as the slave can travel, because it takes months to grow hair. A slave can only be used once for this purpose. (This is why slaves were used: they were considered expendable.)

Hidden messages on paper written in secret inks under other messages or on the blank parts of other messages.

During and after World War II, espionage agents used photographically produced microdots to send information back and forth. Since the dots were typically extremely small—the size of a period produced by a typewriter or even smaller—the stegotext was whatever the dot was hidden within. If a letter or an address, it was some alphabetic characters. If under a postage stamp, it was the presence of the stamp. The problem with the WWII microdots was that they needed to be embedded in the paper, and covered with an adhesive (such as collodion), which could be detected by holding a suspected paper up to a light and viewing it almost edge on. The embedded microdot would reflect light differently than the paper.

More obscurely, during World War II, a spy for the Japanese in New York City, Velvalee Dickinson, sent information to accommodation addresses in neutral South America. She was a dealer in dolls, and her letters discussed how many of this or that doll to ship. The stegotext in this case was the doll orders; the 'plaintext' being concealed was itself a code text giving information about ship movements, etc. Her case became somewhat famous and she became known as the Doll Woman.

The one-time pad is a theoretically unbreakable cipher that produces ciphertexts indistinguishable from random texts: only those who have the private key can distinguish these ciphertexts from any other perfectly random texts. Thus, any perfectly random data can be used as a covertext for a theoretically unbreakable steganography. A modern example of OTP: in most cryptosystems, private symmetric session keys are supposed to be perfectly random (that is, generated by a good Random Number Generator), even very weak ones (for example, shorter than 128 bits). This means that users of weak cryptography (in countries where strong cryptography is forbidden) can safely hide OTP messages in their session keys.

5.2 Modern Steganographic Techniques

Modern steganography entered the world in 1985 with the advent of the Personal Computer applied to classical steganography problems. [3] Development following that was slow, but has since taken off, based upon the number of 'stego' programs available.

- Concealing messages within the lowest bits of noisy images or sound files.
- Concealing data within encrypted data. The data to be concealed is first encrypted before being used to overwrite part of a much larger block of encrypted data.

- Chaffing and winnowing
- Invisible ink
- Null ciphers
- Concealed messages in tampered executable files, exploiting redundancy in the i386 instruction set
- Embedded pictures in video material (optionally played at slower or faster speed).
- A new steganographic technique involves injecting imperceptible delays to packets sent over the network from the keyboard. Delays in key presses in some applications (telnet or remote desktop software) can mean a delay in packets, and the delays in the packets can be used to encode data.
- Content-Aware Steganography hides information in the semantics a human user assigns a datagram; these systems offer security against a non-human adversary/warden.
- BPCS-Steganography - a very large embedding capacity steganography.
- Blog-Steganography. Messages are fractionalized and the (encrypted) pieces are added as comments of orphaned web-logs (or pin boards on social network platforms). In this case the selection of blogs is the symmetric key that sender and recipient are using. The carrier of the hidden message is the whole blogosphere.

5.3 Major Steganography Techniques

- Injection
- Substitution
- Generation of new files

5.3.1 Injection

Injection refers to the insertion of a message into an existing medium. The simplest example is the use of the hidden attribute in Microsoft Word, which allows for hiding text with a special, hidden font. This very simple technique was used to store notes and references during the creation of this document. A casual observer can view the report and not be aware of rough notes that are easily revealed by going to Word's tools/options and clicking on "hidden text". The HTML language allows for the hidden attribute that works in a similar fashion by hiding text from a web browser. Moving up the technology scale and into the security arena it could be argued that the Unicode vulnerability, a technique that has corrupted many web servers by hiding commands in unprintable pieces of web addresses is also a form of stego.

5.3.2 Substitution

This technique replaces data in the original file with a coded representation of the original message. The colors of "pixels", tiny elements of digital images are often represented by the value of a number contained in an eight-bit byte of data. For example, three increasingly redder shades of red might be represented as follows:

"00001100" or decimal 12 might represent basic red in a particular 8-bit color palette. Each of the following numbers would then represent a minor increase in the redness.

"00001101" or decimal 13

"00001110" or decimal 14.

The likelihood of a casual observer noticing the difference in the shades in the middle of a picture is very slight. The result is that steganographers are able to use the 2 least significant bits to encode messages and while the image does degrade slightly, it is not apparent to the naked eye.

Fig (a): Original picture

Fig (b): Picture with hidden message

The above two figures show a picture of a family pet. Figure (a) was taken with a digital camera at a 320 X 240 resolution. Figure (b) shows the result of using S-Tools4 to insert a small text file in to the image using the least significant bit (LSB) technique.

5.3.3 Generation of a new file

Both insertion and substitution require a host file, sometimes called a container, in reference to images, and a host signal in reference to audio signals. Host files, like the pictures above, contain embedded message but may also exhibit characteristics that reveal a pattern that can be used by steganalysis tools to detect the presence of the message.

To eliminate this potential weakness, a coded message can be generated as part of an original computer generated text, audio or image file. One example of an authoring program demonstrating stego techniques is called “Spam Mimic” found at http://www.spammimic.com/index.shtml. This web site allows the viewer to encode a message in a message that looks like spam email.

6. Steganalysis

Steganalysis is the art and science of stopping or detecting the use of all steganographic techniques mentioned earlier. In Steganalysis, the goal is to be able to compare the cover-object (cover message), the stego-object (the cover message with the hidden data embedded in it) and any possible portions of the stego-key (encryption method) in an effort to intercept, analyze and/or destroy the secret communication.

There are six general protocols used to attack the use of Steganography

1) Stego only attack - only the stego object is available for analysis.

2) Known cover attack - the original cover object and the stego object are available for analysis.

3) Known message attack - the hidden message is available to compare with the stego-object.

4) Chosen stego attack - the stego tool (algorithm) and stego-object are available for analysis.

5) Chosen message attack - takes a chosen message and generates a stego object for future analysis.

6) Known stego attack - the stego tool (algorithm), the cover message and the stego-objects are available for analysis.

Generally, bitmap images (.BMP) have known and predictable characteristics. One such characteristic is the probability of near duplicate colors. Bitmap images get their color from a central color table, which by its nature have little, or no near duplicate colors. When hidden data is embedded into the (LSB) of a bitmap image, it increases the number of near duplicate colors dramatically. Generally speaking, any bitmap image with more than fifty near duplicate colors should raise the suspicion of embedded data being present.

7. Applications

The three most popular research uses for Steganography in an open systems environment are covert channels, embedded data and digital water marking.

Covert channels in TCP/IP involve masking identification information in the TCP/IP hesders to hide the true identity of one or more systems. This can be very useful for any secure communication needs over open systems such as the Internet when absolute secrecy is needed for an entire communication process and not just one document.

Using containers (cover messages) to embed secret messages into is by far the most popular use of Steganography today. This method of Steganography is very useful when a party must send a top secret, private or highly sensitive document over a open systems environment such as the Internet. By embedding the hidden data into the cover message and sending it, you can gain a sense of security by the fact that no one knows you have sent more than a harmless message other then the intended recipients.

Although not a pure steganographic technique,Digital watermarking is very common in today’s world and does use steganographic techniques to embed information into documents. Digital water marking is used for copyright reasons by companies or entities that wish to protect their property by either embedding their trademark into their property or by concealing serial numbers/license information in software etc.,

Digital watermarking is very important in the detection and prosecution of software pirates and digital thieves.

8. Cryptography v/s Steganography

Cryptography and steganography are often lumped together even though they are very unalike, yet complementary, technologies with different purposes. Cryptography attempts to change the contents of a file or message in such a way that it is not readable by someone who is not the intended recipient. The intended recipient would have a key that would allow the encrypted file to be unlocked and viewed as planned by sender one problem with encryption is that, it does nothing to hide the fact that a message is being transferred and in fact may even make it more obvious. Its strength is in the difficulty of figuring out the means of encryption and the key to decrypt the message.

Steganography, on the other hand attempts to hide the message in such a way that the observer may not even realize that the message is being exchanged. Combining the two technologies provides a method of communication that is not only difficult to find but also to decipher. The ability to make intercepting stego-files is difficult and to disseminate them is so easy, that has lead to the rise in their use by not only privacy fanatics but also possibly by the terrorist community.

9. The Future

Given the nascent nature of this technology, the scope of the business issues that are affected and the continuing impact of Moore's law on the power of computing it is difficult to predict with any certainty where we will be in 5 years. Nevertheless, the following predictions are presented as a reasonable set of possibilities

- Steganographic techniques will become more common and increasingly sophisticated.
- Steganalysis tools will also become more complex but will typically be behind their steganographic counterparts.
- . A stego process will be developed to embed Trojans, worms and viruses in media such as images or audio files and have they become active by viewing or listening to the files. In 2001, the Nimda worm demonstrated that it was possible to get a virus just by visiting an infected web site. In January of 2002, viruses were being delivered by Macromedia flash images. One day, merely viewing a bitmap image might cause a virus attack on your PC.
- Intrusion Detection Systems (IDS) will include images as part of their attack signatures.
- Anti-virus software will be developed with steganalytical capabilities to detect viruses in audio and image files.
- A strong tamper-resistant, economically viable digital watermark will be developed.

The person who develops the last item will undoubtedly become very rich. There is probably no better motivating factor for learning about Steganography.

10. Conclusion

Steganography has its place in security. It is not intended to replace cryptography but supplement it. Hiding a message with steganography methods reduces the chance of a message being detected. However, if that message is also encrypted, if discovered, it must also be cracked (yet another layer of protection).

There are an infinite number of steganography applications. This paper explores a tiny fraction of the art of steganography. It goes well beyond simply embedding text in an image. Steganography does not only pertain to digital images but also to other media (files such as voice, other text and binaries; other media such as communication channels, the list can go on and on).

Consider the following example: A person has a cassette tape of Pink Floyd's "The Wall." The plans of a Top Secret project (e.g., device, aircraft, covert operation) are embedded, using some steganographic method, on that tape. Since the alterations of the "expected contents" cannot be detected, (especially by human ears and probably not easily so by digital means) these plans can cross borders and trade hands undetected. How do you detect which recording has the message?

This is a trivial (and incomplete) example, but it goes far beyond simple image encoding in an image with homogeneous regions. Part of secrecy is selecting the proper mechanisms. Consider encoding using a Mandelbrot image [Hastur].

In and of itself, steganography is not a good solution to secrecy, but neither is simple substitution and short block permutation for encryption. But if these methods are combined, you have much stronger encryption routines (methods). For example (again over simplified): If a message is encrypted using substitution (substituting one alphabet h another), permute the message (shuffle the text) and apply a substitution again, then the encrypted ciphertext is more secure than using only substitution or only permutation. NOW, if the ciphertext is embedded in an [image, video, voice, etc.] it is even more secure. If an encrypted message is intercepted, the interceptor knows the text is an encrypted message. With steganography, the interceptor may not know the object contains a message.

11. References

[1] http://en.wikipedia.org/wiki/Steganography
[2] http://www.jjtc.com/Steganalysis/
[3] http://www.stegoarchive.com

12. Biography

Prof J.Ashok is currently working as Professor and Head of Information Technology at Geethanjali College of Engg. & Technology, Hyderabad, A.P, INDIA. He has received his B.E. Degree from Electronics and Communication Engineering from Osmania University and M.E. with specialization in Computer Technology from SRTMU, Nanded, INDIA. His main research interest includes neural networks, Bioinformatics and Artificial Intelligence. He has been involved in the organization of a number of conferences and workshops. He has been published more than 30 papers in International journals and conferences. He is currently doing his Ph.D from Anna University and going to submit in the month of Jan 2011.

Mr.Y.Raju is currently working as Associate Professor in the department of IT at Geethanjali College of Engg. & Technology, Hyderabad, A.P, INDIA. He received his M.Tech.(CSE) from JNTU,Hyderabad INDIA. His main research interest includes Information Security, data mining and data ware housing and Bio Informatics.

Mr.S.Munishankaraiah is working as Associate Professor at Geethanjali College of Engineering and Technology, Hyderabad, A.P, INDIA. He has received his B.E. in Computer Science and Engineering from Kakatiya University and Master of Technology in Computer Science and Engineering from Jawaharlal Nehru Technological University. His main research interests include Data Mining and Information Retrieval.

Mr.K.Srinivas is working as Associate Professor in Geethanjali College of Engineering and Technology, Hyderabad, A.P, INDIA. He has received his B.E. in Computer Science and Engineering from S.R.T.M.University and Master of Technology in Computer Science from Jawaharlal Nehru Technological University. He is pursuing Ph.D. from Jawaharlal Nehru Technological University, Hyderabad. His main research interests include Data Mining and Information Retrieval.