Proceedings of Student/Faculty Research Day, CSIS, Pace University, May 6th, 2005

The Rise of Steganography

Alan Siper, Roger Farley and Craig Lombardo

Introduction

Remember the last time you went shopping online? Remember all the pictures of clothes and electronics you viewed? What if those images weren't really for you? What if those pants you were looking at were really detailed blueprints of military installations? You would never know. This is the nature of steganography. Revelations about this technique will be discussed in this paper. Topics will include its history, why it is used, how it works, techniques, counter-measures, risks, legal and ethical issues, and the future.

History of Steganography

To understand steganography, we must first understand its predecessor: cryptography. Cryptography is the art of protecting information by transforming it into an unreadable format, called cipher text. To decipher this unreadable format, a secret key is required.

Cryptography has followed man through many stages of evolution. Cryptography can be found as far back as 1900 B.C. in ancient Egyptian scribe using non-standard hieroglyphics in an inscription. From 500 – 600 B.C. Hebrew scribes used ATBASH, a reversed alphabet simple solution cipher. From 50 - 60 B.C. Julius Caesar used a simple substitution with the normal alphabet in government communications. Cryptography continued through history with may variations. Today cryptography has reached a new level, quantum cryptography. Quantum cryptography combines physics and cryptography to produce a new cryptosystem that cannot be defeated without the sender and receiver having the knowledge of the attempted and failed intrusion. Through the long history of cryptography, steganography was developed and flourished on its own.

Steganography comes from the Greek steganos (covered or secret) and -graphy (writing or drawing). Steganography can be defined as the hiding of information by embedding messages within other, seemingly harmless messages, graphics or sounds. The first steganographic technique was developed in ancient Greece around 440 B.C. The Greek ruler Histaeus employed an early version of steganography which involved: shaving the head of a slave, tattooing the message on the slaves scalp, waiting for the growth of hair to disclose the secret message, and sending the slave on his way to deliver the message. The recipient would have the slave's head to uncover the message. The recipient would reply in the same form of steganography. In the same time period, another early form of steganography was employed. This method involved Demerstus, who wrote a message to the Spartans warning of eminent invasions from Xerxes. The message was carved on the wood of wax tablet, and then covered with a fresh layer of wax. This seemingly blank tablet was delivered with its hidden message successfully. Steganography continued development in the early 1600s as Sir Francis Bacon used a variation in type face to carry each bit of the encoding.

Steganography continued over time to develop into new levels. During times of war steganography is used extensively. During the American Revolutionary War both the British and American forces used various forms of Invisible Inks. Invisible Ink involved common sources, this included milk, vinegar, fruit juice, and urine, for the hidden text. To decipher these hidden messages required light or heat. During World War II the Germans introduced microdots. The microdots were complete documents, pictures, and plans reduced in size to the size of a period and attached to common paperwork. Null ciphers were also used to pass secret messages. Null ciphers are unencrypted messages with real messages embedded in the current text. Hidden messages were hard to interpret within the innocent messages. An example of an innocent message containing a null cipher is:

> Fishing freshwater bends and saltwater coasts rewards anyone
> feeling stressed. Resourceful anglers usually find masterful
> leapers fun and admit swordfish rank overwhelming any day.

By taking the third letter in each word the following message emerges:

> Send Lawyers, Guns, and Money.[1]

Stenography Techniques in a Digital Age

Steganography and cryptology are guided by the same principal, to hide messages in a specific medium. However, they have one distinct difference, cryptology is dependant on hiding the meaning of the message, where as steganography is dependant on hiding the presence of a message altogether. Oddly enough, steganography accomplishes this through little of its own traits. The sheer size of the Internet and its vast amounts of data is what accomplishes this fete, and for this reason, it can be a very effective method of securing data transfer.

The most common steganography technique, using mostly image and sound carrier files, is called Least Significant Bit Substitution (LSBS) or overwriting. As the name implies, LSBS involves overwriting the bit with the lowest arithmetic value. The result of this process alters the original output very slightly. This is done slightly enough to be unlikely to be detected from human senses.

This is a simple example of LSBS. Take this straightforward bit sequence as a piece of a carrier file:

```
10010101  00001101  11001001   10010110
00001111  11001011  10011111   00010000
```

Underlined are the Least Significant Bits in each byte group. The significance of these bits is so minor when compared to the whole, that altering these bits could produce close to the same result.

1001010***0*** 0000110***1*** 1100100***0***   1001011***0***
0000111***0*** 1100101***1*** 1001111***1***   0001000***1***

Only half of the Least Significant Bits were changed in the virgin sample, and yet the character G has been discretely imbedded into the sequence. Judging from the amount of bits needed to make even the simplest of files, it is easy to imagine just how much hidden data can be secretly embedded using Least Significant Bit Substitution.[2]

Although the example above would be a perfectly viable way to use LSBS, it is too basic to be practical. The main goal of steganography is to shield the presence of a hidden message from human senses. However, modern steganography detection applications and techniques (steganalysis) has altered those goals to include securing the hidden message from both human senses, and digital applications alike. Due to this reason, almost all steganography applications use some kind of randomization technique in which the altered Least Significant Bits are spread out randomly across the carrier file. This practice creates the biggest obstacle for steganalysis

[1] Steganography - SEC202. http://www.jjtc.com/stegdoc/sec202.html.

[2] An Overview of Steganography for the Computer Forensics Examiner. http://www.garykessler.net/library/fsc_stego.html

applications. Now, not only does the application have to check if a file might carry a hidden message, but it has to compare that file to the vast amount of randomization techniques known.

Stenography Detection (Steganalysis)

Steganalysis could be simply defined as the detection of steganography by a third party. This is done in a variety of ways and is usually based around how much prior information is available. Whether the analyst has access to all the medium and algorithms will greatly affect the method he/she uses. Below is a list and description of some common steganalysis methods.

Carrier Comparison:

With this method the Steganalysis application visually inspects and compares the carrier images. If enough inconsistencies are detected between the two, it might be enough to label the file suspect.

Although this technique can be effective, many steganography tools reserve overwriting only bits that would cause the least amount of distortion to the carrier file. Some of these bits include areas of brighter color in images, and louder sounds in audio files. This clever technique is an obstacle for the steganalysis method. Luckily, not too many steganalysis tools take advantage of this technique yet.

Structural Inspection:

Similar to carrier comparison, structural inspection studies the carrier file for inconsistencies. However, this method would be used in the instance where only the suspect file was available. Mostly all steganography algorithms will cause some sort of structural oddity that would suggest manipulation of the original file. These oddities often include the redundancy of color palettes in image files.

Statistical Analysis:

This technique is usually used when the analyst is working in the blind. Steganographic techniques commonly alter the natural statistics of the carrier file. As a result of this phenomenon, it is possible to study and inspect possible carrier images, and determine whether the type of file studied deviates from the expected norm.

This method is one of the fastest growing steganalysis techniques. Many experts feel that the future of detecting and stopping the harmful uses of steganography lies in searching and inspecting vast amounts of information looking for suspect data files. However, because most steganography algorithms take pains to preserve the properties of the virgin carrier file, this technique is not as currently as effective as many analysts would hope. The method is further complicated when the hidden message is encrypted.

Risks & National Security

Corporations are increasingly becoming more aware of the risks related to steganography. With firewalls, intrusion detection systems, and other related security tools not yet able to detect messages hidden within a carrier file, hackers are able to do such tasks as plant the blueprints for penetrating a company’s computer systems within audio files stored on a firm’s own Web site. Globally, governments seek to ensure that they can obtain encryption keys to read messages when they suspect that such culprits as drug smugglers, money launderers, and terrorists are using encrypted messages. As a result, those with malicious intent will not use a method of communicating that can be easily intercepted and read.[3]

A search on the Internet will unveil hundreds of thousands of links to pages on steganography. The links include links to free downloadable software and the mathematical formulas behind how the technique works. Criminals of any type try to conceal data daily and steganography gives them another, and online, option of doing it.[4] When used for the wrong purposes, the technique becomes a threat to the security of the worldwide information infrastructure.[5]

Steganography poses risks that reach as far as affecting national security. While steganography seems to be a good method of exchanging sensitive information in a secure manner, it can also be misused. Speculation exists that terrorists use these techniques to communicate via seemingly innocent Web sites.[6] The theory is that terrorist groups are allegedly hiding maps and photographs of terrorist targets and posting instructions for terrorist activities on sports chat rooms, pornographic bulleting boards, and other web sites.[7] In fact, soon after September 11, 2001, an investigation took place of the scanning of over two million images from eBay's Web site for hidden messages contained in them. No information was supposedly found, but the government, at the least, certainly became aware of the risk.[8]

The threat raised by steganography is very real. Its use is not easy to detect or intercept, as the information does not need to be broadcast across the Internet. The hidden message can reside unsuspectingly on a Web site for example and can be viewed from around the world. Although the main threat at the moment is to national security, the technology is undoubtedly being used for other immoral purposes. Therefore, although steganography is not yet the sort of threat that IT auditors are battling against on a regular basis, it is one that needs to be considered and understood for possible future occurrences.

Not all uses of steganography are bad. Watermarks can be inserted in identification cards making it harder for a counterfeiter to duplicate the card. Another positive is that perfectly legal, confidential information can be carried more securely. Furthermore, companies are capitalizing on the technique to make everyday business more secure. For example, Digimarc Corporation, a leading supplier of secure media solutions, provides secure watermarking identification solutions to governments across the globe.[9]

Legal & Ethical Ramifications

Using steganography for illegal and/or immoral purposes challenges one's ethics. The very idea of steganography, hiding a message so that only the creator and intended audience

[3] Oliphant, Alan. "Steganography: When Security Becomes A Threat." January 1, 2003. The Institute of Internal Auditors. http://www.theiia.org/itaudit/index.cfm?fuseaction=forum&fid=523.

[4] Bartlett, John. "The Ease of Steganography and Camouflage." March 17, 2002. The SANS Institute. http://www.sans.org/rr/whitepapers/vpns/762.php.

[5] Wang, Huaiqing and Shuozhong Wang. "Cyber warfare: steganography vs. steganalysis." Queue. December 2004/January 2005. http://acmqueue.org/modules.php?name=Content&pa=showpage&pid=241.

[6] Oliphant, Alan. "Steganography: When Security Becomes a Threat." January 1, 2003. The Institute of Internal Auditors. http://www.theiia.org/itaudit/index.cfm?fuseaction=forum&fid=523.

[7] Bartlett, John. "The Ease of Steganography and Camouflage." March 17, 2002. The SANS Institute. http://www.sans.org/rr/whitepapers/vpns/762.php.

[8] Loney, Matt. "Secret codes 'not hidden in Web images'." September 26, 2001. ZDNet UK. http://news.zdnet.co.uk/internet/0,39020369,2096060,00.htm.

[9] Digimarc Corporation. 2005. http://www.digimarc.com/watermark/idmarc/.

know about it and can read it, opens a wide array of unethical possibilities. With more emphasis on detecting steganography the past few years, the stricter the clamp down will be on those abusing steganography for malicious reasons.

The recent crackdown on the free downloading of multimedia online could have been aided by the use of steganography. Steganography, or digital watermarking in this case, could have helped identify copies of illegally distributed music.[10] Concern now is attributed towards child pornography. The wide use of the Internet makes it easy for unethical individuals to hide an image of child pornography within a seemingly harmless looking image, text, or sound file. These culprits can transmit and receive images, or post images on Web sites that possess the damaging material without raising alarm to anyone but the intended audience.[11]

In late 1998, the government passed the Digital Millennium Copyright Act (DMCA). The bill was implemented from treaties signed at the World Intellectual Property Organization (WIPO) conference a couple years prior. The DMCA makes it a crime to distribute software that cracks copy protection schemes. The DMCA includes a list of provisions including limiting online service providers from copyright infringement for merely transferring information over the Internet. As related to steganography, service providers would not be held accountable for altered multimedia (e.g. images with hidden messages) passing through their network. On the other hand, service providers are required to remove material from their customers' Web sites that appears to constitute copyright infringement.[12]

The DMCA states "no person shall circumvent a technological measure that effectively controls access to a work protected under this title." Countries are required to penalize citizens who remove digital watermarks from media. The Motion Picture Association of America (MPAA) and the Recording Industry Association of America (RIAA) use the bill to protect their content from illicit copying and also to bring to justice those who committed activities that were legal prior to the law's passage.[13]

In 2003 "Super-DMCA," as it was coined, bills were under review in several states in the U.S. after being passed in a few others. The broad language used in some versions of the bill angered technologists, including steganography expert Neil Provos. For example, Michigan residents cannot intentionally "assemble, develop, manufacture, possess, deliver, offer to deliver, or advertise" any device or software that conceals "the existence or place of origin or destination of any telecommunication service." Providing written instructions on creating such a device or program was also made illegal and violators of the bill can be penalized with a maximum of four years in prison. Provos added at the time, "It's very difficult, reading the law, it makes basically everything that I do illegal."[14]

Legal action taken against abusers of the steganography technique will not be easy. Detection tools and steganalysis measures will need to improve before the proper culprits can be

[10] Tierney, Emma. "Steganography: NOW and THEN." Elements. http://www.ul.ie/elements/Issue6/Steganography.htm.

[11] Astrowsky, Brad H. "STEGANOGRAPHY: Hidden Images, A New Challenge in the Fight Against Child Porn." Anti-Child Porn Organization. http://www.antichildporn.org/steganog.html.

[12] "The Digital Millennium Copyright Act." The UCLA Online Institute for Cyberspace Law and Policy. Feb 8, 2001. http://www.gseis.ucla.edu/iclp/dmca1.htm.

[13] Glass, Brett. "Hide In Plain Sight." October 2002. PC Magazine. http://www.findarticles.com/p/articles/mi_zdpcm/is_200210/ai_ziff31242.

[14] Poulsen, Kevin. SecurityFocus. "'Super-DMCA' fears suppress security research." April 14, 2003. http://www.theregister.co.uk/2003/04/14/superdmca_fears_suppress_security_research/.

brought to justice. Once convicted, a precedent will also need to be set by the courts by handing out harsh punishments, discouraging others to follow the same path. As Hany Farid, Assistant Professor of Computer Science at Dartmouth University, puts it, "The courts are terribly unprepared to handle the new breed of digital criminals that has emerged, along with the rapid increase in low-cost and sophisticated digital technology. As the criminals get smarter, so must we."[15]

The Future of Steganography

Steganography continues to improve. With every discovery of a novel steganography format, new applications must be devised. These advancements in steganography have taken us to today's methods of inserting data to images, documents and sound.

For every step steganography has taken to hide the data over the past 1500 years, mankind has worked hard to find the hidden messages. With today's computer steganographics, finding and decoding the hidden messages have become more complicated. Currently, steganalysts are working hard to detect the hidden messages within images, documents, and sound. Steganalysis starts with suspected data files. The steganalyst uses forensic statistician information to help reduce the number of files. The analyst then compares the questionable data files to similar data files. The similarity is based on the same digital camera or digital audio device. [16] The analyst is looking at visual detection (jpeg, bmp, gif, etc.), audible detection (wav, mpeg, etc.), statistical detection (changes in patterns of pixels or Least Significant Bit) or histogram analysis, and structural detection (view file properties/content, size difference, date/time difference, contents – modifications, checksum).[17] Once steganography is detected, and the information is extracted, it may still be encoded. At this point, cryptanalysis techniques may be applied. Steganalysts have just started their battle against the hidden data. Much more must be done to detect the dangerous data hidden behind the innocent looking pictures.

In the future, steganography will continue to grow, as our needs for security expand. The security issues for "The Good" and "The Bad" are the same: "we must hide our data." At the same time, both will work on decoding the new forms of steganography to obtain the other's data. Either way you look at it, new forms of steganography and steganalysis will be developed.

Stegtunnel is one of the new tools written to hide data within TCP/IP header fields. This data is hidden in the sequence number and the IPID fields of packets used for TCP connection.[18] With this new steganographic method, data will be completely undetectable unless the key is known. Stegtunnel is currently available, but one must consider current weaknesses that are being addressed. Stegtunnel contains weaknesses such as drop packet protocol, out of order packets, and random ISN's and IPID's are not noticeable. Stegtunnel is working on future versions of the program to resolve these problems.

DNA based steganography is designed to take the old microdot steganography one step further. The new technology in this proposal is based on a genomic steganography procedure that was developed and published by Clelland et al in 1999.[19] This new procedure

[15] "Secret messages in digital images pose a challenge one Dartmouth researcher can't resist." August 10, 2001. http://www.dartmouth.edu/~news/releases/2001/aug01/embedded.html.

[16] Steganalysis - Wikipedia, the free encyclopedia, http://en.wikipedia.org/wiki/Steganalysis

[17] http://www.blackhat.com/presentations/bh-usa-04/bh-us-04-raggo/bh-us-04-raggo-up.pdf#search='steganalysis%20forensics'

[18] Considerations in the design of stegtunnel, http://www.synacklabs.net/projects/stegtunnel/

[19] DNA Based Steganography for Security Marking, http://www.polestarltd.com/ttg/isspeeches/051403/

takes a DNA encoded message that is camouflaged with genomic DNA. [20] The data is hidden amongst millions of other similar looking DNA molecules. This is taken one step further and concealed in a microdot. To recover the data, the recipient must first find the microdot. The recipient must also have the primer sequence and the encryption key to read the data. With the key, the molecule is detected and then read by a DNA sequence analysis.

Conclusion

Government officials and law makers all over the world recognize that the Internet should remain a single medium that is affordable to the public. Affordable, accessible Internet access, however, comes at a cost. The Internet's use can be abused and used to create as much harm as it can good. The possible risks involved with the unmonitored and uncontrolled exchange of information can be described as staggering. What is even more disturbing is how little statistical information we have on how widespread the uses of techniques like steganography are. As the use of steganography grows in both frequency and complexity, the current truth might no longer matter. This is a risk that no government or computer forensics examiner should take. Ignoring the significance of steganography because of the lack of statistics is "security through denial" and not a good strategy.[21]

[20] Hiding messages in DNA microdots,
www.cs.memphis.edu/~garzonm/pub_old/datsec/dnastegano.pdf#search='genomic%20steganography'

[21] An Overview of Steganography for the Computer Forensics Examiner.
http://www.garykessler.net/library/fsc_stego.html