Steganographically Encoded Data

Adrian VASILESCU, Bucharest, Romania

Steganography is the art of hiding information in ways that prevent its detection. Though steganography is an ancient craft, the onset of computer technology has given it new life. Computer-based steganographic techniques introduce changes to digital covers to embed information foreign to the native covers. Such information may be communicated in the form of text, binary files, or provide additional information about the cover and its owner such as digital watermarks or fingerprints. This paper explains steganography, provides a brief history and describes how steganography is applied in hiding information in images.
***Keywords**: Steganography, information hiding, digital image, digital watermarking*

Steganography is the art and science of writing hidden messages in such a way that no one apart from the intended recipient knows of the existence of the message; this is in contrast to cryptography, where the existence of the message itself is not disguised, but the content is obscured.

The word "*Steganography*" is of Greek origin and means *"covered, or hidden writing"*. Its ancient origins can be traced back to 440 BC. Herodotus mentions two examples of Steganography in *The Histories of Herodotus*. Demeratus sent a warning about a forthcoming attack to Greece by writing it on a wooden panel and covering it in wax. Wax tablets were in common use then as re-usable writing surface, sometimes used for shorthand. Another ancient example is that of Histiaeus, who shaved the head of his most trusted slave and tattooed a message on it. After his hair had grown the message was hidden. The purpose was to instigate a revolt against the Persians. Later, Johannes Trithemius's book *Steganographia* is a treatise on cryptography and steganography disguised as a book on black magic. Generally, a steganographic message will appear to be something else: a picture, an article, a shopping list, or some other message. This apparent message is the *covertext*. For instance, a message may be hidden by using invisible ink between the visible lines of innocuous documents.

The advantage of steganography over cryptography alone is that messages do not attract attention to themselves, to messengers, or to recipients. An unhidden coded message, no matter how unbreakable it is, will arouse suspicion and may in itself be incriminating, as in some countries encryption is illegal

Steganographic techniques

Modern steganographic techniques

- 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. This technique works most effectively where the decrypted version of data being overwritten has no special meaning or use: some cryptosystems, especially those designed for filesystems, add random looking padding bytes at the end of a ciphertext so that its size can't be used to know what was the plaintext size. Examples of software that use this technique include FreeOTFE and TrueCrypt.
- 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 keypresses in some applications (telnet or remote desktop) can mean a delay in packets, and the delays in the packets can be used to encode data. There is no extra processor or network activity, so the steganographic technique is "invisible" to the

user. This kind of steganography could be included in the firmware of keyboards, thus making it invisible to the system. The firmware could then be included in all keyboards, allowing someone to distribute a keylogger program to thousands without their knowledge.

- Content-Aware Steganography hides information in the semantics a human user assigns a datagram; these systems offer security against a non-human adversary/warden.

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, 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, if the story is true, carried a warning to Greece about Persian invasion plans .
- 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 codetext giving information about ship movements, etc. Her case became somewhat famous and she became known as the Doll Woman.Counter-propaganda: During the Pueblo Incident, US crew members of the USS Pueblo (AGER-2) research ship held as prisoners by North Korea communicated in sign language during staged photo ops to inform the United States that they had not defected, but were instead captured by North Korea. In other photos presented to the US, the crew members gave "the finger" to the unsuspecting North Koreans, in an attempt to discredit the pictures that showed them smiling and comfortable.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 crypto (in countries where strong crypto is forbidden) can safely hide OTP messages in their session keys.

Additional terminology

In general, terminology analogous to (and consistent with) more conventional radio and communications technology is used; however, a brief description of some terms which show up in software specifically, and are easily confused, is appropriate. These are most relevant to digital steganographic systems.

The *payload* is the data it is desirable to transport (and, therefore, to hide). The *carrier* is the signal, stream, or data file into which the payload is hidden; contrast "*channel*" (typically used to refer to the type of in-

put, such as "a JPEG image"). The resulting signal, stream, or data file which has the payload encoded into it is sometimes referred to as the *package*. The percentage of bytes, samples, or other signal elements which are modified to encode the payload is referred to as the *encoding density* and is typically expressed as a floating-point number between 0 and 1. In a set of files, those files considered likely to contain a payload are called *suspects*. If the *suspect* was identified through some type of statistical analysis, it may be referred to as a *candidate*.

Countermeasures

The detection of steganographically encoded packages is called steganalysis. The simplest method to detect modified files, however, is to compare them to the originals. To detect information being moved through the graphics on a website, for example, an analyst can maintain known-clean copies of these materials and compare them against the current contents of the site. The differences (assuming the carrier is the same) will compose the payload. In general, using an extremely high compression rate makes steganography difficult, but not impossible; while compression errors provide a good place to hide data, high compression reduces the amount of data available to hide the payload in, raising the encoding density and facilitating easier detection (in the extreme case, even by casual observation).

Applications

An example from modern practice

Image of an artic hare. By removing all but the last 2 bits of each color component, an almost completely black image results. Making the resulting image 85 times brighter results in the image below.

Image extracted from above image.

The larger the cover message is (in data content terms — number of bits) relative to the hidden message, the easier it is to hide the latter. For this reason, digital pictures (which contain large amounts of data) are used to hide messages on the Internet and on other communication media. It is not clear how commonly this is actually done. For example: a 24-bit bitmap will have 8 bits representing each of the three color values (red, green, and blue) at each pixel. If we consider just the blue there will be 2^8 different values of blue. The difference between say 11111111 and 11111110 in the value for blue intensity is likely to be undetectable by the human eye. Therefore, the least significant bit can be used (more or less undetectably) for something else other than color information. If we do it with the green and the red as well we can get one letter of ASCII text for every three pixels.

Stated somewhat more formally, the objective for making steganographic encoding difficult to detect is to ensure that the changes to the carrier (the original signal) due to the injection of the payload (the signal to covertly embed) are visually (and ideally, statistically) negligible; that is to say, the changes are indistinguishable from the noise floor of the carrier.

From an information theoretical point of view, this means that the channel must have more capacity than the 'surface' signal requires, that is, there must be redundancy. For a digital image, this may be noise from the imaging element; for digital audio, it may be noise from recording techniques or amplification equipment. In general, electronics that digitize an analog signal suffer from several noise sources such as thermal noise, flicker noise, and shot noise. This noise provides enough variation in the captured digital information that it can be exploited as a noise cover for hidden data. In addition, lossy compression schemes (such as JPEG) always introduce some error into the decompressed data; it is possible to exploit this for steganographic use as well.

Steganography can be used for digital watermarking, where a message (being simply an identifier) is hidden in an image so that its source can be tracked or verified.

In the era of Digital video recorder and de-

vices like TiVo, TV commercials authors have figured out how to make use of such devices as well - by putting a hidden message which becomes visible when played at frame-by-frame speed

Usage in modern printer

Steganography is used by some modern printers, including HP and Xerox brand color laser printers. Tiny yellow dots are added to each page. The dots are barely visible and contain encoded printer serial numbers, as well as date and time stamps. This is another way of watermarking.

Rumored usage in terrorism

The rumors about terrorists using steganography started first in the daily newspaper USA Today on February 5, 2001. The articles are still available online, and were titled "*Terrorist instructions hidden online*", and the same day, "*Terror groups hide behind Web encryption*". In July of the same year, the information looked even more precise: "*Militants wire Web with links to jihad*".

A citation from the *USA Today* article: "*Lately, al-Qaeda operatives have been sending hundreds of encrypted messages that have been hidden in files on digital photographs on the auction site eBay.com*". These rumors were cited many times - without ever showing any actual proof - by other media worldwide, especially after the terrorist attack of 9/11.

References

[1] **Peter Wayner**, *Disappearing Criptography – Information Hiding, Steganography and Watermarking*, Morgan Kaufmann, New York, 2004.

[2] **Andreas Westfeld**, *Steganographie: Grundlagen, Analyse, Verfahrensentwicklung,* Springer, March 2006.

[3] **Ross Anderson**, *Information Hiding*, Cambridge 1996

[4] **Peter Wayner**, *Translucent Databases*, Barnes&Noble, NY 2005

[5] **S. Katzenbeisser** and **F. Petitcolas,** *Information Hiding Techniques for Steganography and digital Watermarking,* Artech House, Boston, 2000**.**

[6] **N.F. Johnson, Z. Duric,** Information Hiding: Steganography and Watermarking - Attacks and Countermeasures**,** Kluwer Academic Publishers, 2000.