Discovering Covert Digital Evidence

By

Chet Hosmer, Christopher Hyde

From the proceedings of

The Digital Forensic Research Conference

DFRWS 2003 USA

Cleveland, OH (Aug 6th - 8th)

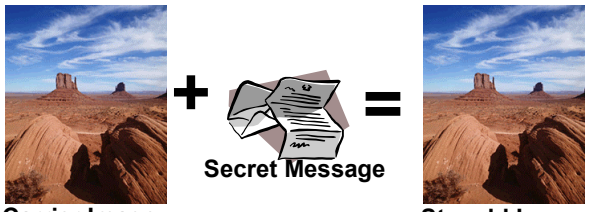
Discovering Covert Digital Evidence

Chet Hosmer & Christopher Hyde
WetStone Technologies, Inc.
DFRWS - 2003

Steganography Then & Now

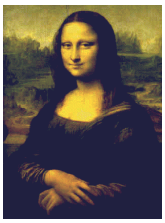
Steganography is a word originating with the Greeks meaning 'covered writing'. Its use can be traced back at least 2,500 years, when Demaratus smuggled a secret message under the cover of wax in order to warn Sparta of an impending attack on Greece[i]. Other examples used throughout history include invisible ink, null-ciphers, and microdots.

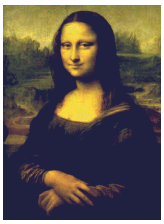
In our modern times, digital images, audio files and streaming video have become carriers for hidden information, while our networks are the high-speed delivery channels. Unfortunately, these modern delivery channels are still guarded by unsuspecting sentries, in the form of firewalls and sensors that are unable to detect the messages that may be hidden inside of digital images and audio files.



Why would criminals or terrorist use an ancient method to hide data? Surely modern cryptography provides a more proven method for keeping information private. Actually, the distinction between steganography and encryption is the answer. Encryption is used to keep the contents of information private or confidential, and only those holding the proper keys can extract the secret contents. The sole purpose for the use of steganography, on the other hand, is to hide the fact that the secret message, (possibly containing incriminating evidence) even exists. The military calls this 'covert communication,' and the path for this communication is called a 'covert channel'.

To illustrate an example of what steganography does, presented here are two images whose differences are undetectable to the naked eye. The original 680 KB image on the left appears identical to the image on the right. The image on the right, however, contains a secret message in Microsoft® Word format that is 88KB in size, or about 13% of the original cover image.


Original


Original + Secret Message

Increasing Use

During the last decade, technology for digitally manipulating image, video, and audio data has advanced tremendously, resulting in the rapid increase of information of hiding in binary data files. The sophistication of the steganography programs has also evolved from simple homemade solutions to commercial grade offerings. And, the proliferation of steganography software is widespread. During a recent research effort, WetStone Technologies examined over 100 programs that hide messages in image files, audio video files, and in text documents. Many of these programs are freely downloadable from the Internet, while some charge a nominal fee of between $49 and $89. Our research indicates that over one million copies of steganography software have been downloaded or purchased over the Internet during the last 24 months.

Methodologies

There are many techniques used for hiding secret messages in images. The common denominator among them is that they combine a carrier file with a secret message to produce a resulting binary file containing the hidden message. The process hides the data in such a way that the changes are indiscernible to the human eye,(or ear in the case of audio). The methodologies vary widely from simple least significant bit (LSB) modification to sophisticated JPEG and MP3 transform modifications.

In LSB substitution, modifications are made to the least significant bits of the cover file's individual

pixels or sounds, thereby encoding hidden data. Generally, these changes make imperceptible modifications. In the more sophisticated methodologies, like transform modifications, the actual compression algorithms are modified in a way that data can be hidden during the encoding process. Still other methodologies of digital steganography exist including word substitution, file appending, covert channels, and file source modification.

Dangers of Steganography

The increase in availability, sophistication, and popularity of steganography programs increases the potential opportunities for industrial espionage, trade secret theft, cyber weapon exchange, and criminal coordination and communication. There is a growing concern in law enforcement and private investigative processes that steganography is an unknown, and until now, an undetected component in various criminal activities.

Obviously, the explosion of Internet traffic provides the perfect environment for data hiding via steganography. With the vast amount of potential cover files that are posted, e-mailed, and viewed on web pages every day, it is impossible to manually scan even a fraction of them. Tremendous potential exists for a significant amount of illicit traffic to go undetected by law enforcement. Furthermore, as the bandwidth of the Internet increases to afford easier transmission of large files such a real time video, there will be an increased risk that illegal communications will go undetected.

What is the threat, then, to our corporate infrastructures from this new form of invisible ink? The answer comes in three basic areas:

First, the use of steganography provides the ability to smuggle sensitive information out of an organization by a disgruntled or angry employee. This can be accomplished by hiding the information via steganographic techniques inside an innocuous looking attachment to an email message.

Second, the use of steganography allows someone to hide criminal or unethical information on corporate resources. A quick scan of a departing user's desktop might reveal digital photos from the company picnic, family outings, or a collection of clipart used for building slide presentations. These all will appear completely harmless and innocuous, when in fact they may contain potentially incriminating information that had been hidden from view.

Finally, the use of steganography allows for the utilization of corporate resources to communicate criminal or terrorist information. Most companies are diligent in their awareness regarding the content of their corporate web sites and public-facing information. However, every image on that corporate web site is a potential carrier of hidden information if exploited. Many of these images float around the organization and originate from many sources, providing the opportunity for insertion of secret or covert content. Furthermore, web sites that allow public postings of information (i.e., auction, dating and real estate sites) offer outsiders a way to use the Internet to communicate virtually anything without detection.

Since the terrorist attacks of 9/11, it has come to light that steganography was probably used as an element of communication during the planning of the operation. According to reports in USA Today, Wired News, and on CNN, Al-Qaeda was interested in steganography and its uses[ii].

Steganalysis and Forensic Investigation

As the use of steganography becomes more prevalent in the criminal world in both the traditional and the new and growing world of cyber crime, inspecting digital files to discover or detect steganography takes on increased importance. The emerging discipline of steganography investigation, or *steganalysis*, is a specialized subset of digital forensics, which should be practiced hand-in-hand with other investigative activities.

Postmortem Investigation Strategies

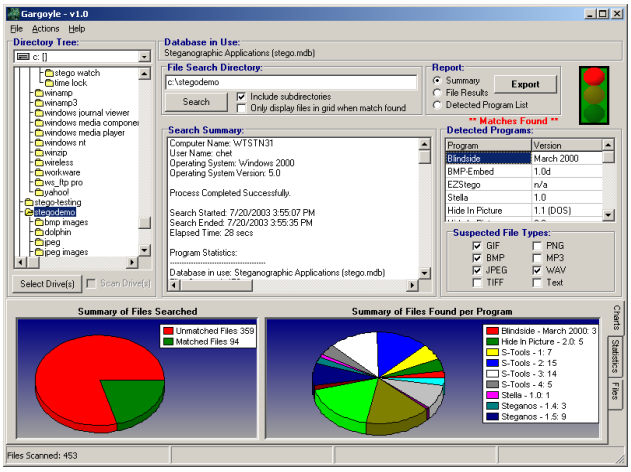
Investigation into the possible use of steganography can be a complex and time-consuming process. Examining evidence from a hard-drive that contains 10,000 images, 2,500 audio files, and 500 video clips could easily take weeks. If multiple computers, backup media or Internet analysis is required the process could, of

course, take much longer. However, applying a process similar to the one shown below, and using sophisticated technologies can significantly shorten the process.

Determining Suspicion – Why do you suspect the use of steganography? Some critical questions you might consider:

a. Does the suspect have computer savvy?
b. Does the type of crime warrant hiding incriminating information (child pornography, drug trafficking, terrorism)
c. Is the amount of digital evidence found small, but there are a large number of images or audio files on the suspect's hard drive? Did you locate duplicate images, or deleted duplicate images on the hard-drive that contain different hash values.

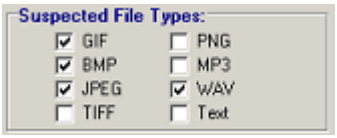
Identifying Known Stego Programs – Once you are suspicious that the suspect may have employed steganography, a search of the hard-drive for steganography programs, collateral material, or their remnants is in order. There are several methods for doing this. In the latest release of the National Software Reference Library (NSRL), Version 2.1 by the National Institute of Standards and Technology, (NIST), and the National Institute



of Justice, (NIJ) a hash set exists that was extracted from the StegoArchive CD. Performing a search using the NSRL dataset with Guidance Software's EnCase or WetStone's Gargoyle™ may provide important clues.

Expanded hash sets that provide additional results data can help to narrow down the exact steganography program used. By identifying the specific steganography tool(s) two distinct benefits can be gained. First, the types of carrier files that specific steganography programs support can be obtained. For example, if the only steganography program found is GIFITUP, then you can focus your search on GIF image files. The second benefit provided by this approach is the ability to utilize the offending program to attempt to extract the hidden data, provided you can obtain the password used by the suspect.

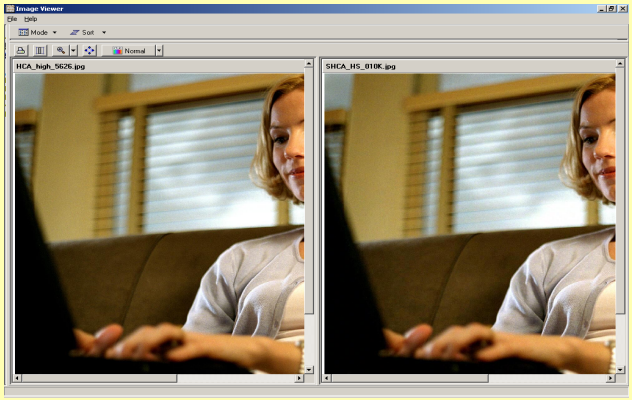
Identifying Suspect Files – Once it has been determined that steganography programs were in use by the suspect, identification of the suspect carrier files (carrier files: text, digital images, audio or video files that can provide cover for hidden messages or information) will help to narrow the search. Gargoyle identifies the suspect file types for you.



Identifying Suspicious Carriers – Once you have narrowed the search to known file types, you can perform analysis of the suspect files by examining them. Assuming that the suspect files are images, several tools are available for the detection of steganography. The simplest method is to view images or examine audio files for unusual or suspicious characteristics. The basic fundamental concept behind steganography is to hide information such that the "normal" view of the image is not distorted by adding the payload, or hidden material. The normal view of the image to most of us is the projection of the Red-Green-Blue (RGB) values on a computer display. Therefore commercial grade stego attempts to hold the RGB view of image data constant. However in doing so, subtle changes can be seen by viewing other notational forms of the image, such as Hue or Saturation. Using a Stego Image Viewer, Normal and Saturation views are shown on the left in the figure on the following page.

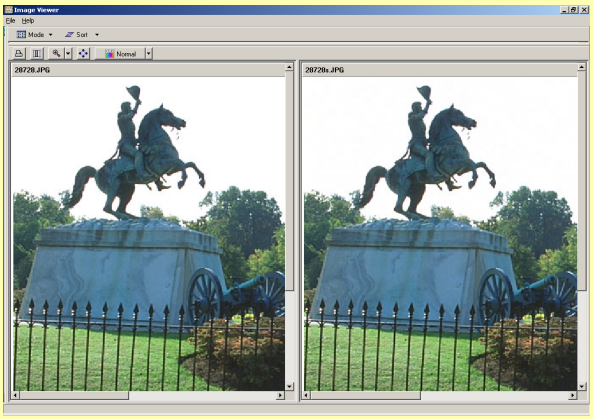
Again on the following page, in the example on the right, we started with an 111K JPEG carrier file, and used the JSTEG steganography program to imbed a 5K byte file. In this example we contrast the depiction of the Normal image vs. that of the notational view of Hue.

Normal View



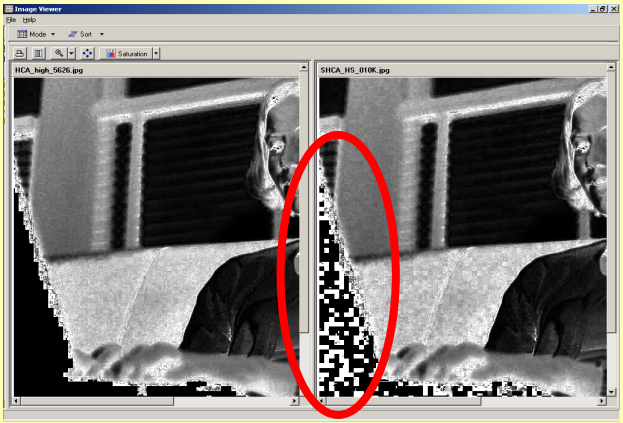
Both images look normal, no visual clues exist

Normal View



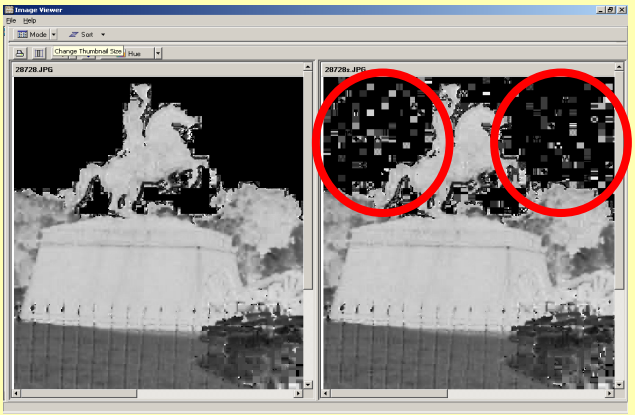
Again, both images look normal, no visual clues exist

Saturation View



In this view the image on the right clearly is distorted with a recognizable pattern indicating steganography.

Hue View



As with the Saturation View, the image on the right clearly is distorted with a recognizable pattern indicating steganography.

Detecting Steganography – Steganalysis techniques have been developed by a variety of researchers most notably including: J. Fridrich, N. Johnson, P. Peticolas, R.J. Anderson, and others. Applying these theoretical approaches has proved at best to be difficult. Most methods provide reasonable identification of suspect images, however they suffer from high false positive rates. This is especially true when applied to a wide variety of images (image size fluctuations, image color density, varying color intensities, very small or very large images, image that have been converted between formats etc.) The difficulty can be attributed to the wide variety of images, audio and video files that exist coupled with a wide array of sources of these files. For example over 20 (viable) different manufactures of digital cameras exist today each offering dozens of models. Each model creates slightly different image formats; using different compression algorithms, with some now including digital watermarks. The ability to build a statistical model with a combination of highly accurate detection and low false positive results is complicated and problematic. If you add to this consideration of the manipulation of these images with image processing programs, editing applications, conversion and compression programs the difficulty of modeling "normal digital images" increases at a tremendous rate.

In 1999, WetStone Technologies performed the Steganography Detection and Recovery Toolkit (S-DART) research project under contract with U.S Air Force Research Laboratory. [iii]. The project was to study the possibilities of developing computer

algorithms to blindly detect steganography without knowing the embedding program used. Since that time, WetStone has built upon that research, and developed Stego Watch™, a steganography detection toolkit. Stego Watch was designed and developed not only as a commercial product benefiting law enforcement agencies, but as a framework usable by the research community to test the most promising stego detection algorithms and quickly move them from the laboratory to the frontlines. In this vein, the strength of Stego Watch is in providing a platform for new algorithms and approaches to be rapidly deployed for the detection and identification of steganography, both in postmortem and online investigations. Stego Watch exposes an API for researchers and developers that allows for new research and steganography detectors. In this way, researchers and developers can independently add new algorithms and compare/correlate their results. The API and software library module are available from WetStone.

Conclusion

Research and field testing needs to be done before there will be a reliable amount of data regarding the successes and limitations of varying investigative techniques being utilized in the rapidly growing cyber war. It is clear however that steganography will continue to be a problem for law enforcement as more users realize the value of hiding certain data. Unless reliable field methodologies for investigating covert information are developed to compliment traditional modern detection techniques, law enforcement will continue to be at a disadvantage to steganography savvy criminals. We propose three basic approaches to this challenge. 1) Increase awareness and provide training for law enforcement investigators and forensic specialists in the detection and analysis of convert information. 2) Perform pro-active cyber investigations that uncover the use of steganography and other advanced data hiding or covert techniques. 3) Intensify efforts and step-up the prosecution of those using these technologies for either criminal or terrorist purposes.

Resources

Steganography Tool Archive: www.stegoarchive.com
Information Hiding: Steganography and Watermarking - Attacks and Countermeasures, Neil Johnson, et al.

Stego Watch™ Postmortem and online steganography detection toolkit.

Gargoyle™ Hostile program detector with Steganography dataset.

Stego Image Viewer™r for the display of normal and notational forms of suspect images.

Chet Hosmer *is a co-founder, and the President and CEO of WetStone Technologies, Inc. His specialty areas include steganography, secure time, intrusion detection and response, and cyber forensics. If you would like to contact Mr. Hosmer, you may reach him at chet@wetstonetech.com.*

Christopher Hyde *is a Digital Forensic Analyst with WetStone Technologies, Inc. He has been researching steganography and digital forensics since 1999. If you would like to contact Mr. Hyde, you may reach him at chris@wetstonetech.com.*

WetStone Technologies, Inc
www.wetstonetech.com

---

[i] Kahn, The Codebreakers, 1967, page 81
[ii] http://www.usatoday.com/life/cyber/tech/2001-02-05-binladen.htm
http://www.wired.com/news/ politics/0,1283,41658,00.html
http://www.cnn.com/2001/US/09/20/inv.terrorist.search/
[iii] Steganography Detection and Recovery Toolkit (S-DART) Final Technical Report, 23 Feb 00, USAF/AFMC Contract F30602-99-C-0210.