

e-mail :

Subhash.Nigam@abbott.com

## Steganography: Methodology and Challenges

Subhash Nigam, Section Manager, Scientific Informatics and Automation, Global Pharmaceuticals Research and Development (GPRD), Abbott Labs, Abbott Park, IL 60064.

**Steganography comes from the Greek word *steganos*, which means covered or hidden, and *graphein*, which means to scribe or to write. Basically, it means writing that others cannot see. Due to a surge in computational power and capabilities of the hardware and software, steganography is now at the forefront of modern security-enabling techniques.**

### INTRODUCTION

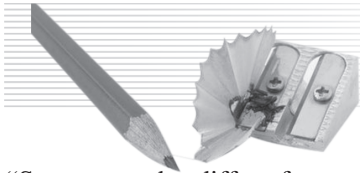
In today's highly competitive and dynamic world, it is the data and information that fuels the engine of the global economy. In order to ensure that information is secured and does not go to unintended destination, the concept of hiding information has attracted both scientists and artists to come up with creative solutions to protect a piece of information from falling into wrong hands. Securing and safeguarding data and information has always been a challenge for professionals in information technology and management. The problem is considered to be of utmost importance to computer scientists because of its applicability in securing and protecting data from unintended hands. There have been numerous incidents where the security and integrity of data, and hence confidential information, was compromised. In this day and age of distributed computing using N-tier based architecture and, distribution and sharing of information, the art and science of protecting data from wrong hands is becoming more and more complex.

Numerous advancements in cryptography have taken place to ensure that data stays concealed. One of the methods that hide information from unwanted hands is called steganography. The challenge for steganography is to veil information in such a way that it looks normal to those who are not looking for hidden information. This age-old art has been practiced since the early days of message passing, and is now used more frequently when moving data in binary digits over the modern computer networks. The word "steganography" comes from the Greek word *steganos*, which means covered or hidden, and *graphein*, which means to scribe or to write. Basically, it means writing that others cannot see. Due to a surge in computational power and capabilities of the hardware and software, steganography is now at the forefront of modern security-enabling techniques.

### METHODOLOGY

The process of steganography starts with taking the confidential message and encrypting it using any means. This gives streams of bits called ciphertext. The next step is to take a picture or a sound file to act as a coartext, and inserting the ciphertext into it. This is done by manipulating bits that represent spaces, font, font-size, font-effects, colors, etc. Manipulating the bits and inserting ciphertext into the coartext produces a file called stegotext. The job of coartext is to hide the ciphertext in such a way that it looks like a regular picture or a sound file. The recipient of the stegotext will "uncover" the stegotext, separate coartext from ciphertext and then decode the ciphertext, extracting the original message.

Stenography is considered as the dark cousin of cryptography. Cryptography codes a message into some gibberish-looking streams of characters, where as steganography hides the secret information by blending it with publicly exposed information. The main advantage of steganography over cryptography is that as the messages travel, they do not attract attention of "onlookers" or of recipients. Many times, steganography and cryptography are used simultaneously to ensure security of the stealthy message. Everything looks fine to someone who is not looking for hidden information. "Digital images are good vehicles of steganography. All images contain redundant data: information as to color, for example, that is present but unnecessary for the picture to be seen and understood. This enables the senders of a secret message to substitute digitized text for some of the redundant pixels in a photo" (Information Security for Technical Staff, pg. 369, 2003). Using steganography, messages and entire files can be embedded within image and/or audio files. The files and messages remain hidden unless you know how to detect and expose them.



“Steganography differs from encryption, though in practice they’re often combined. Unlike stego, which aims for undetectability, encryption relies on ciphers or codes to keep a message private after it has been detected” (McCullagh, 2001).

Some common approaches to hide information in digital images include least significant bit insertion, masking and filtering, and algorithms and transformation. Joint Photographic Experts Group (JPEG) files are suitable to hide messages because they use markers. One of the many interesting approaches in creating a steganographic image is replacing a comment marker in a copyright image by the information that we want to hide. This approach replaces a comment in a commercially available image by the hidden information. The approach keeps the size of the file and the image same. Another popular approach is to randomly select a marker in the image and then inserting a comment marker along with the message to be hidden. This approach changes the size of the file, however, the size of the image does not change. Another common technique for hiding secret images is called Substitution-Based. Using this technique “the secret bits are embedded in the image while following a pseudo-random path through the image...this technique embeds the secret bits by comparing the current secret bit to be hidden with the least significant bit (LSB) of the RGB color in the color palette that is pointed at by the current pixel index value.” If the bits match then it is considered to be already embedded, if they do not match then the “algorithm needs to search the RGB color palette for a similar, replacement color that does have a LSB matching the secret bit” (Ryder, May, 2004). Compression algorithms play an important role in steganography. GIF and 8-bit BMP files utilizes lossless compression and JPEG employs lossy compression (Kessler, September, 2001). “Lossless compression lets us reconstruct the original message exactly; therefore, it is preferred when the original information must remain intact. Lossy compression may save space but may not maintain the original image’s integrity” (Jajodia & Johnson, February 1998).

Creating and implementing novel ways for embedding secrets into digital objects such as pictures and sound-bytes also requires new ways to create countermeasures to detect the hidden secrets. This is called steganalysis. Using statistical analysis, to some extent, it is possible to find if an image has been modified. Probing an image’s statistical properties to find any deviations from the norm, gives researchers ability to detect if a file has been modified by steganography. Embedding encrypted data into a GIF image changes the histogram of its color frequencies. The changes in color frequencies can determine if an image has been modified

(Honeyman & Provos, August 31, 2001). A steganalyst has to use many techniques on suspected images. This makes the job of steganalyst difficult since sometimes decisions produced by different steganalysis techniques are in contradiction. This is where information fusion methods called composite steganalyzers that aggregate the outputs of multiple steganalysis techniques come in to assist (Kharrazi, Sencar, & Memon, 2006).

### **Challenges**

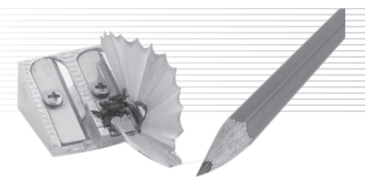
Steganography continues to be a successful technology with its own set of challenges. With all the good that this technique can do, it also carries the baggage of negative aspects of information hiding. The challenge is to ensure that “information should not be modified by unauthorized subjects (integrity); information should be available to authorized subjects when required (availability); improper disclosure of information should be detected and prevented (confidentiality)”. (Siponen & Oinas-Kukkonen, February, 2007).

As a double edged sword, one of the key problems in steganography is its inability to detect and track data coming and leaving the internal network of a secured corporate environment. The network traffic can also carry malicious software and code-snippets that are hidden in the incoming data streams. In June 2000, a Seattle-based forensics-consulting firm Electronic Evidence Discovery (EED) found a couple of emails with simple images attached that were sent by an engineer. These harmless-looking images contained two hidden and stolen engineering specifications that were critical assets of the company that hired the forensics company (Radcliff, 2002).

Steganography can also be used as a tool in the hands of terrorists to share messages and their covert plans of creating havoc. Terrorists may have communicated to plan the events of September 11, 2001 (Kellen, 2001). U.S. officials have little doubt that al-Qaeda and other militant groups are using the Web to set up terrorist attacks against the United States. They found some sites that “...contained encrypted information that directed al-Qaeda members to a more secure site where instructions for attacks were given” (Kelley, July 10, 2002). According to officials, “Steganography has, in fact, cropped up as a concern. The remnants of al-Qaeda might be able to communicate and continue to coordinate using this method. Intelligence agencies checked out a Website for so-called Azzam Publications, thinking it held stego messages for jihad fighters” (Maney, 2001). Using steganography, terrorists need not be technically savvy to shroud their conversations. They could simply hide essential elements of their conversation into a simple bit map and send it over a data network.

Another challenge to Steganography is the ability of damaging

# Articles



## *Steganography: Methodology and Challenges*

the tools that make the forensic tools reliable. Damaging these forensic tools has devastating effect on automated forensic analysis. Tools such as Timestamp, Slacker, Transmogrify spoof the forensic evidence and create problems for law enforcement in detecting crimes (Berghel, April 2007). Another problem is that “[since] steganography is used to combine explanatory information with an image (like doctor’s notes accompanying an X-ray)... [it is possible that] it could accidentally degrade or render an image resulting in misleading results” (Radcliff, 2002). There are also trade-offs involved in embedding larger message size (Chandramouli & Memon, 2003).

### **Conclusion**

The current analysis of steganography shows that the challenge it has is coming up with various dynamically changing algorithm to hide information. “For every clever method and tool being developed to hide information in multimedia data, an equal number of clever methods and tools are being developed to detect and reveal its secrets” (Wang & Wang, Oct 2004). Steganography creates dark data – data that is hidden, and buries it with light data, data that is known. As computing power and deciphering techniques continue to advance, algorithms that hide and un-hide the data need to match and take advantage of the computational speeds of securing and exposing the data. Steganography also has potential to act as a double-edged sword that can potentially damage the images and degrade the quality of them, especially in the areas where images are of paramount importance, and having poor images could inadvertently lead to incorrect results. It is only through proper awareness, training, and lossless algorithms we expect steganography to gain momentum as the preferred technology that ensures that information does not fall into the wrong hands. However, “the more information is placed in the public’s reach on the Internet, the more owners of such information need to protect themselves from theft and false representation” (Jajodia & Johnson, 1998). With increasing power of search engines like Google, European Union and United States have put Google’s privacy policies under scrutiny (Kirk, September, 2007). The challenge for regulatory agencies stays the same. No matter what happens with privacy policies, detecting steganographic text will always be a difficult task to tackle. As the art and science of hiding and detecting data gets technically complex, more work is warranted to study this technology. This is necessary to ensure that decision makers of today and tomorrow use it as a modern management technique to collect and disseminate accurate information that would eventually be the basis of decisions they make. □

### **References**

- (2003). Information Security of Technical Staff, page 369, Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA 15213-3890
- Berghel, H. (2007, April). Hiding Data, Forensics, and Anti-Forensics, *Communications of the ACM*. Vol. 50 No. 4, 15-20.
- Chandramouli, R., Memon, N. (2003). Steganography Capacity: A Steganalysis Perspective. *Proceedings of SPIE-IS&T*. Vol. 5020, 173-177.
- Honeyman, P., Provos, N. (2001, August 01). Detecting Steganographic Content on the Internet. CITI Technical Report 01-11, CITI, University of Michigan, 535 W. William St., Ann Arbor, MI 48103-4943.
- Jajodia, S., Johnson, N. (1998, February). Exploring Steganography: Seeing the Unseen. *IEEE Computer*, 26-34.
- Kellen, T. (2001). Hiding in Plain View: Could Steganography be a Terrorist Tool? Information Security Reading Room. Retrieved October 14, 2007 from [http://www.sans.org/reading\\_room/whitepapers/steganography/551.php](http://www.sans.org/reading_room/whitepapers/steganography/551.php)
- Kelley, J. (2002, July 10). Militants wire web with links to jihad. *USA Today, News*, pg A01.
- Kessler, G. (2002, April). Steganography: Hiding Data in Data. *Windows & .NET Magazine*.
- Kharrazi, M., Sencar, H., Memon, N. (2006). Improving Steganalysis by Fusion Technologies: A Case Study with Image Steganography. *Transactions on Data Hiding and Multimedia – Lecture Notes in Computer Science*. Vol 4300, 123-137. New York: Springer-Verlag.
- Kirk, J. (2007, September 14). Google calls for global online privacy standard. *NetworkWorld*. Retrieved October 21, 2007 from <http://www.networkworld.com/news/2007/091407-google-calls-for-global-online.html>
- Maney, K. (2001, December 19). Bin Laden’s message could be hiding in plain sight. *USA Today, Money* B06.
- McCullagh, D. (2001). Secret Messages Come in. *Wavs*. Retrieved October 7, 2007 from <http://www.wired.com/politics/law/news/2001/02/41861>
- Radcliff, D. (2002). QuickStudy: Steganography: Hidden Data. *Computer World*. Retrieved October 9, 2007 from <http://www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=71726>
- Ryder, J. (2004, May). Steganography may increase learning everywhere. *Consortium of Computing Sciences in Colleges, CCSC: Northeastern Conference*, 154-162
- Siponen, M., Oinas-Kukkonen, H., (2007, February). Cyber Warfare: A Review of Information Security Issues and Respective Research Contributions. *The DATABASE for Advances in Information System*. Vol. 38, No.1, 60-80.
- Wang, H., Wang S., (2004, October). Cyber Warfare: Steganography vs. Steganalysis. *Communications of the ACM*. Vol.47, No. 10, 776-782.