STEGANOGRAPHY TORN APART!

Author: Punit Ganshani
Email Address: punit@ganshani.com

Published In: DeveloperIQ, May 2004, Asia

This may seem to be an ordinary beginning to an ordinary article. It is not. There's a secret message hidden here, in this very paragraph. It's not in view, and its source is modern. But the art of hiding messages is an ancient one, known as steganography. Yes, it's happening since ages that messages within army troops are sent through this old art of hiding them behind images.

Now in an era of technology spreading its tentacles wide, steganography is just not hiding text in images. Steganography is the dark cousin of cryptography, the use of codes. While cryptography provides privacy, steganography is intended to provide secrecy.

Changing of bits of the pixels of the image, writing text within the same and making it appeal as the original one. It's of course not merely an art; it's all around what our article deals with!

The Evolution!
While discussing in terms of computer security, steganography is really nothing new, as it has been around since the times of ancient Rome where at one age, text was traditionally written on wax that was poured on top of stone tablets. And to obscure the message this wax was then nailed out. Message still remained carved on the tablets, tablets that appealed no different than the ordinary ones!

Coined from the words "stego" and "cryptography", steganography means "the art of writing in cipher, or in characters, which are not intelligible except to persons who have the key; cryptography"

What caused its blooming up?
Well it was not very famous till it was suspected that terrorists connected with the September 11 attacks might have used it for covert communications. While no such connection has been proven, the concern points out the effectiveness of steganography as a means of obscuring data. Indeed, along with encryption, steganography is one of the fundamental ways by which data can be kept confidential.

Implementing Steganography!
There are a vast number of tools that are available for steganography. An important distinction that should be made among the tools available today is the difference between tools that do steganography, and tools that do steganalysis, which is the method of detecting steganography and destroying the original message.

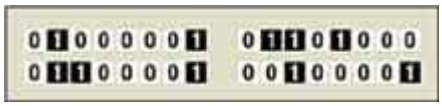
Steganalysis focuses on this aspect, as opposed to simply discovering and decrypting the message, because this can be difficult to do unless the encryption keys are known.

Steganography strips less important information from digital content and injects hidden data in its place. This is done over the spectrum of the entire image. Let's see the different ways of implementing it!

One Way of Implementing Steganography: The Pure Steganography!

The following sequence of 24 bits represents a single pixel in an image. Its 3 bytes of color information provide a total of 256 different values for each color (red, green and blue) and thus can represent a total of 16.7 million colors. This particular value displays as a dark green:

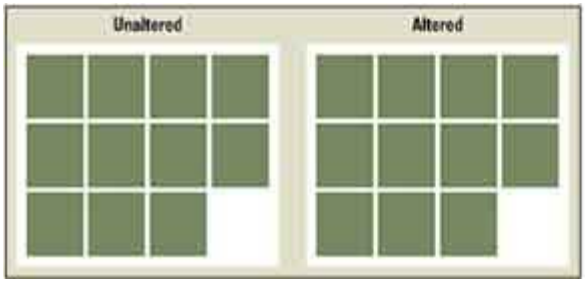
Now, let's take 11 of these pixels that represent, say, part of a **solid-color** background. In the following sequence, the least significant (rightmost) bit of each 8-bit byte has been co-opted to hide a text message—the four characters Aha!—in ASCII binary:



Here are the bits behind those 11 pixels:

The hidden message occupies 32 of those 264 bits (**about 12%**) and contains four 8-bit bytes. In the diagram, each maroon or gold box represents a bit that had to be changed to include the hidden message. Notice that only 15 of 264 bits (**less than 6%**) had to be changed and only eight of the 11 pixels were altered.

The two figures below represent the 11 colored pixels we've been manipulating. The figure on the left is the original, unaltered version. The one on the right has been modified, as shown above. And yes, there is no difference either!



If instead of 11 pixels we had a 300KB bitmap file, we could accommodate a text message of 36KB, or about 6,000 words

An Example

The most popular methods use digitized photographs, so let's explore these techniques in some depth. Digitized photographs and video also harbor plenty of white noise. A digitized photograph is stored as an array of colored dots, called pixels whose values often range from 0-255.

Each number is stored as eight bits (zeros and ones), with a one worth 128 in the most significant bit (on the left), then 64, 32, 16, 8, 4, 2, and a one in the least significant bit (on the right) worth just 1.

A difference of one or two in the intensities is imperceptible, and, in fact, a digitized picture can still look good if the least significant four bits of intensity are altered -- a change of up to 16 in the color's value. This gives plenty of space to hide a secret message. Text is usually stored with 8 bits per letter, so we could hide 1.5 letters in each pixel of the cover photo. A 640x480 pixel image, the size of a small computer monitor, can hold over 400,000 characters. That's a whole novel hidden in one modest photo!

Hiding a secret photo in a cover picture is even easier. Line them up, pixel by pixel. Take the important four bits of each color value for each pixel in the secret photo (the left ones). Replace the unimportant four bits in the cover photo (the right ones). The cover photo won't change much, you won't lose much of the secret photo, but to an untrained eye you're sending a completely innocuous picture.

With these new techniques, a hidden message is indistinguishable from white noise. Even if the message is suspected, there is no proof of its existence. To actually prove there was a message, and not just randomness, the code needs to be cracked or the random number seed guessed. This feature of modern steganography is called "plausible deniability."

Unfortunately, anyone who cares to find your hidden image probably has a trained eye. The intensity values in the original cover image were white noise, i.e. random. The new values are strongly patterned, because they represent significant information of the secret image. This is the sort of change which is easily detectable by statistics. So the final trick to good steganography is make the message look random before hiding it. And Embedded Cryptography is one such way to do it!

Yet another Way of Implementing Steganography: The Embedded Cryptography!

As mentioned before, steganalysis softwares are freely available on Internet Sites such as *http://www.jjtc.com/Steganalysis* and many more in the list. But it becomes hard for these softwares to detect the hidden text messages if they are first encrypted because then these softwares require

a Private Key to decipher the algorithm of the text hidden behind

That is where embedded cryptography jumps in!

The monk Johannes Trithemius, one of the founders of modern cryptography, has given us a scheme to conceal messages in long invocations of the names of angels, with the secret message appearing as a pattern of letters within the words. For example, as every other letter in every other word:

padiel aporsy mesarpon omeuas peludyn malpreaxo

which reveals "prymus apex."

Another clever invention was the "*Ave Maria*" cipher. There are a series of tables, each of which has a list of words, one per letter. To code a message, the message letters are replaced by the corresponding words. If the tables are used in order, one table per letter, then the coded message will appear to be an innocent prayer.

These simple systems hide a short text message in a letter that looks exactly like spam, which is as ubiquitous on the Internet today as innocent prayers were in the 16th century.

One of the softwares SpamMimic *(available on www.spammimic.com)* uses "grammar" to make the messages. For example, a simple sentence in English is constructed with a subject, verb, and object, in that order. Given lists of 26 subjects, 26 verbs, and 26 objects, we could construct a three word sentence that encodes a three letter message. If you carefully prescribe a set of rules, you can make a grammar that describes spam.

And one can’t imagine a simple text as:

"Dear Son, I will not be coming today"

will be constructed as,

Today we have launched a new company, a company that will provide benefit to all its members. A company that aims at giving discounts to its members on all the food items, books and crockery items they buy from us. And what else! The more members you make under you, the more will you earn! Its not just chain marketing, it more than that. Its about making your own house, a home. Just pay 1700 INR and get yourself registered. Make just 6 members under you and you get back 50% of your amount but if you make just 3 more that is 9 members you get back 75% of your amount. Don't think, just act now! Offer closes on 31st March 2004

This text is now stored in a picture file by changing the bits and what so ever may be the software, who cares!

A software can **detect that there is some text in the picture** *(steganalysis software)* but its very tough to find out what **algorithm** has been applied to encipher the text as above. And the best way to encrypt a text is by using a pseudo-random number generator.

And yes, a good steganographic technique should provide secrecy even if everyone knows it's being used!

Steganography and Security
Steganography is an effective means of hiding data, thereby protecting the data from unauthorized or unwanted viewing. But stego is simply one of many ways to protect the confidentiality of data. It is probably best used in conjunction with another data-hiding method. When used in combination, these methods can all be a part of a layered security approach. Some good complementary methods include:

- Encryption - Encryption is the process of passing data or plaintext through a series of mathematical operations that generate an alternate form of the original data known as ciphertext. The encrypted data can only be read by parties who have been given the necessary key to decrypt the ciphertext back into its original plaintext form. Encryption doesn't hide data, but it does make it hard to read!

- Hidden directories (Windows) - Windows offers this feature, which allows users to hide files. Using this feature is as easy as changing the properties of a directory to "hidden", and hoping that no one displays all types of files in their explorer.

- Hiding directories (Unix) - in existing directories that have a lot of files, such as in the /dev directory on a Unix implementation, or making a directory that starts with three dots (...) versus the normal single or double dot.

- Covert channels - Some tools can be used to transmit valuable data in seemingly normal network traffic. One such tool is Loki. Loki is a tool that hides data in ICMP traffic (like ping).

Protecting Against Malicious Steganography

Unfortunately, all of the methods mentioned above can also be used to hide illicit, unauthorized or unwanted activity. What can you do to prevent or detect issues with stego? There is no easy answer. If someone has decided to hide their data, they will probably be able to do so fairly easily. The only way to detect steganography is to be actively looking for in specific files, or to get very lucky. Sometimes an actively enforced security policy can provide the answer: this would require the implementation of company-wide acceptable use policies that restrict the installation of unauthorized programs on company computers.

Using the tools that you already have to detect movement and behavior of traffic on your network may also be helpful. Network intrusion detection systems can help administrators to gain an understanding of normal traffic in and around your network and can thus assist in detecting any type of anomaly, especially with any changes in the behavior of increased movement of large images around your network. If the administrator is aware of this sort of anomalous activity, it may warrant further investigation. Host-based intrusion detection systems deployed on computers may also help to identify anomalous storage of image and/or video files.

Attacks for Detection

There are two types of attacks to detect steganography. They are the **visual attack** (actually seeing the differences in the files that are encoded) and the **statistical attack**: "The idea of the statistical attack is to compare the frequency distribution of the colors of a potential stego file with the theoretically expected frequency distribution for a stego file." It might not be the quickest method of protection, but if you suspect this type of activity, it might be the most effective.

Applications!

Most of the newer applications use steganography like a watermark, to protect a copyright on information. Photo collections, sold on CD, often have hidden messages in the photos which allow detection of unauthorized use. The same technique applied to DVDs is even more effective, since the industry builds DVD recorders to detect and disallow copying of protected DVDs.

Even biological data, stored on DNA, may be a candidate for hidden messages, as biotech companies seek to prevent unauthorized use of their genetically engineered material. The technology is already in place for this: three New York researchers successfully hid a secret message in a DNA sequence and sent it across the country.

The Uses of this technology can be listed as:

- Used to combine explanatory information with an image (like doctor's notes accompanying an X-ray)
- Embedding corrective audio or image data in case corrosion occurs from a poor connection or transmission
- Peer-to-peer private communications
- Posting secret communications on the Web to avoid transmission
- Copyright protection
- Maintaining anonymity
- Hiding data on the network in case of a breach

The Pros and Cons!
However advanced may the technology be, it has some pros and some cons which actually enlighten the path to the discovery of a new vision, new technology. And the new technology is splendid! Let's see to the drawbacks of this widely spread technology.

- Could accidentally degrade or render an image misleading
- Could counteract and be counterproductive with the original image
- Doesn't hide the fact that an e-mail was sent, negating the purpose of secret communications
- Someone else with a steganography detection and cracking tool could expose the message
- A form of this already exists, called digital watermarking, but requires use of separate hardware tools because steganographic software can't use separate hardware tools. Steganographic software also can't protect the watermark.
- Easier to open free Web-based e-mail or use cloaked e-mail
- Better to understand and effectively use standardized encryption

Conclusion
Steganography is a fascinating and effective method of hiding data behind mediums like images that has been used throughout history. Methods that can be employed to uncover such devious tactics, but the first step are awareness that such methods even exist. There are many good reasons as well to use this type of data hiding, including watermarking or a more secure central storage method for such things as passwords, or key processes. Regardless, the technology is easy to use and difficult to detect. The more that you know about its features and functionality, the more ahead you will be in the game.

References
ONLINE REFERENCES

[1] Steganography, by Neil F. Johnson, George Mason University, http://www.jjtc.com/stegdoc/sec202.html

[2] http://dictionary.reference.com/search?q=steganography

[3] The Free On-line Dictionary of Computing, © 1993-2001 Denis Howe
http://www.nightflight.com/foldoc/index.html

BOOK REFERED

[1] Cryptography and Network Security, William Stallings (Prentice Hall, Third Edition)