

Digital Steganography:

Hiding Data within Data



Donovan Artz • Los Alamos National Laboratory

To human eyes, data usually contains known forms, like images, e-mail, sounds, and text. Most Internet data naturally includes gratuitous headers, too. These are media exploited using new controversial logical encodings: steganography and marking.

Take a moment and think about my first paragraph. If your only thoughts are that I wrote a vague and awkward introduction, then I have succeeded in using a very simple form of steganography. Observe the first letter of every word in the first paragraph and then add some appropriate spaces in the resulting string of characters. You should discover a hidden message: “The duck flies at midnight. Tame uncle sam.” If you suspected that my first paragraph contained a hidden message before I told you it was there, then my attempt at steganography failed. My chances of success are fairly high, however, given that most individuals never consider the fact that there is more than one way to parse something that looks like written English in a magazine.

Internet users frequently need to store, send, or receive private information. The most common way to do this is to transform the data into a different form. The resulting data can be understood only by those who know how to return it to its original form. This method of protecting information is known as encryption. A major drawback to encryption is that the existence of data is not hidden. Data that has been encrypted, although unreadable, still exists as data. If given enough time, someone could eventually unencrypt the data. A solution to this problem is steganography.

The Data beneath the Wax

Digital steganography is the art of inconspicuously hiding data within data. Steganography’s goal in general is to hide data well enough that unintended

recipients do not suspect the steganographic medium of containing hidden data. In my opening example, I used this article as the carrier data. I inserted my hidden data using a technique I hoped the reader would never suspect. But now that I have revealed this article as a medium of steganography, my success in using a similar technique will be limited. This is true simply because you might now suspect that something other than an article on steganography (maybe another hidden message) exists within the remainder of my writing.

Steganography and data hiding are not new concepts. It is believed that steganography was first practiced during the Golden Age in Greece. An ancient Greek record describes the practice of melting wax off wax tablets used for writing messages and then inscribing a message in the underlying wood. The wax was then reapplied to the wood, giving the appearance of a new, unused tablet. The resulting tablets could be innocently transported without anyone suspecting the presence of a message beneath the wax.

The important concept from this history lesson is that communication does not have to occur over standard open channels using well-known methods. The Internet, in its massive, protocol-laden glory, is a playground for the modern steganographer. For example, think of an IP packet as the wax tablet previously mentioned. The packet’s data field is equivalent to the writing in the wax. The headers serve as the wood in this analogy – who ever looks at an IP packet’s headers, much less the data alignment padding? Most every protocol, language, and data format on the Internet has room for rent.

Uses of Steganography

Steganography is a means of storing information in a way that hides that information’s existence. Paired

randoM capitalosis is a rare disease of ten contracted by careless internet users. This sad illness causes the affected person to randomly capitalize letters in a body of text. please do not confuse this disease with a blatant attempt at steganography.

Figure 1. Difficulty hiding data. This text, encoded as 8-bit ASCII, is 254 bytes long. The hidden message, at 23 bytes, is about ten times smaller than the carrier data but remains conspicuous.

with existing communication methods, steganography can be used to carry out hidden exchanges. Governments are interested in two types of hidden communications: those that support national security and those that do not. Digital steganography provides vast potential for both types. Businesses may have similar concerns regarding trade secrets or new product information. Avoiding communication in well-known forms greatly reduces the risk of information being leaked in transit.

Businesses have increasingly taken advantage of another form of steganography, called *watermarking*. Watermarking is used primarily for identification and entails embedding a unique piece of information within a medium without noticeably altering the medium. (For details, see Katzenbeisser and Petitcolas¹ and Barán et al.²) For example, if I create a digital image, I can embed in the image a watermark that identifies me as the image's creator. I would achieve this by manipulating the image data using steganography, such that the result contains data representing my name without noticeably altering the image itself. Others who obtain my digital image cannot visibly determine that any extra information is hidden within it. If someone attempts to use my image without permission, I can prove it is mine by extracting the watermark. Watermarking is commonly used to protect copyrighted digital media, such as Web page art and audio files.

Steganography can also enhance individual privacy. Although not a substitute for encryption, digital steganography provides a means of communicating privately. Of course, this is effective only if the hidden communication is not detected. If a person simply wants to communicate without being subjected to his or her employer's monitoring systems, then digital steganography is a good solution — the most private communication is the one that never existed!

Steganography and Encryption

The purpose of steganography is not to keep others

from knowing the hidden information — it is to keep others from thinking that the information even exists. If a steganography method causes someone to suspect the carrier medium, then the method has failed. Steganography's success thus relies heavily on the naïveté of human beings; for example, when did you last check your e-mail headers for hidden messages?

Encryption and steganography achieve separate goals. Encryption encodes data such that an unintended recipient cannot determine its intended meaning. Steganography, in contrast, does not alter data to make it unusable to an unintended recipient. Instead, the steganographer attempts to prevent an unintended recipient from suspecting that the data is there.

Those who seek the ultimate in private communication can combine encryption and steganography. Encrypted data is more difficult to differentiate from naturally occurring phenomena than plain text is in the carrier medium. Several existing steganography tools can encrypt data before hiding it in the chosen medium (see the sidebar, "Steganography Sites", p. 79).

Limitations

Steganography is constrained by the same assumption that exists for encryption. If Alice wants to send an image with a hidden message to Bob, she must first privately agree with Bob on a method of steganography. Under the encryption model, Bob can be fairly sure when he's got some ciphertext. However, in the steganography model, it will be difficult for Bob to know when an image is just an image.

Consider the scenario in which Alice borrows Bob's digital camera and neglects to tell him to pay special attention to every 73rd byte in the images she sends him. Because Bob is ignorant of Alice's steganographic efforts, the large number of pictures he receives from her will only decrease the chance that Bob will let Alice borrow his digital camera again.

The amount of data that can be effectively hidden in a given medium tends to be restricted by the size of the medium itself (see Figure 1). The fewer constraints that exist on the integrity of the medium, the more potential it has for hiding data. For example, this paragraph is constrained by the rules of the English language and a specific topic of discussion. It would be difficult for me to hide a secret message in this paragraph due to the limited number of ways one can reasonably alter this text under those constraints. In contrast, consider a

large uncompressed image of television static, as illustrated in Figure 2. A significantly greater proportion of data could be embedded in such an image without causing suspicion, aside from the fact that pictures of television static are of questionable value in the first place.

The message from Figure 1 can be successfully hidden in Figure 2 using the following method:

- Excluding the black border, start at the pixel in the upper left corner of the image.
- Set the color value of the current pixel to the ASCII value of the corresponding character in the message you want to hide.
- Move two pixels to the right. If you are at the edge of the image, wrap around and skip a line.
- Repeat the previous two steps until the entire message is coded.

Methods of Digital Steganography

There are many techniques available to the digital steganographer. The most common technique is to exploit the lenient constraints of popular file formats. Many publicly available software packages use this technique on a variety of media.

Images as Carriers

Images are a good medium for hiding data (for details, see Pan, Chen, and Tseng³). The more detailed an image, the fewer constraints there are on how much data it can hide before it becomes suspect. The JPHide/JPSeek package (<http://linux01.gwdg.de/~alatham/stego.html>) uses the coefficients in a JPEG to hide information. A newer method (<http://www.know.comp.kyutech.ac.jp/BPCSe/BPCSe-principle.html>) embeds data in visually insignificant parts of an image. Both of these methods alter the image; however, you can explore image degradation using different images and messages of varying length. An alternative, specific to GIF images, is to manipulate an image's palette in order to hide data. Gifshuffle (<http://www.darkside.com.au/gifshuffle/>) does not alter the image itself in any visible way; rather, it permutes a GIF image's color map, leaving the original image completely intact.

Audio File Carriers

Several packages also exist for hiding data in audio files. MP3Stego (<http://www.cl.cam.ac.uk/~fapp2/steganography/mp3stego/>) not only effectively hides arbitrary information, but also claims to be a partly robust method of watermarking MP3 audio files. The Windows

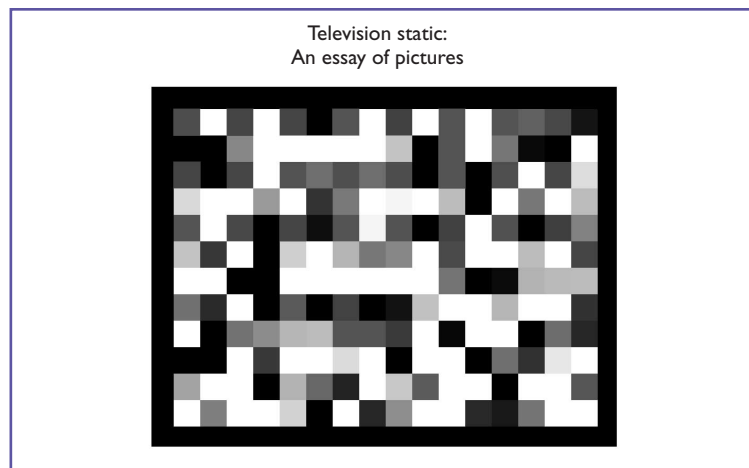


Figure 2. A medium conducive to success. The message from Figure 1 can be hidden using a picture of television static. The static in the black box can be encoded in 192 bytes, which is only eight times larger than the hidden message.

Wave format lets users hide data using StegoWav (<http://www.radiusnet.net/crypto/steganography/Java/stegowav.zip>) or Steghide (<http://steghide.sourceforge.net/>). Steghide alters the least significant bits of data in the carrier medium. Although nearly equal in data-hiding potential, the large size of meaningful audio files makes them less popular than image files as a steganographic medium.

Data Ordering

Ordering data that does not have an ordering constraint is often an effective method of steganography. Each permutation of a set of objects can be mapped to a positive integer. This mapping can then be used to encode hidden data by altering the order of objects that are not considered ordered by the carrier medium. While this technique generally does not change the information quality, hidden data can easily be lost if the medium is encoded again. For example, if I have a GIF whose color map contains hidden data, I can open the GIF in my favorite graphics-editing package and save it again. Visually, the resulting GIF should be identical to the original, but the ordering of the color map may have been lost.

Figure 3 (next page) shows examples of steganography using data ordering. In the HTML example, the ALT and NAME tags can be swapped to represent one bit of data. The first line, if used, could represent a 0; if the second line is used, it could represent a 1. The ordering of NAME and ALT in additional IMG tags can represent additional bits. In the simple example using Perl code, each line assigns the same sum

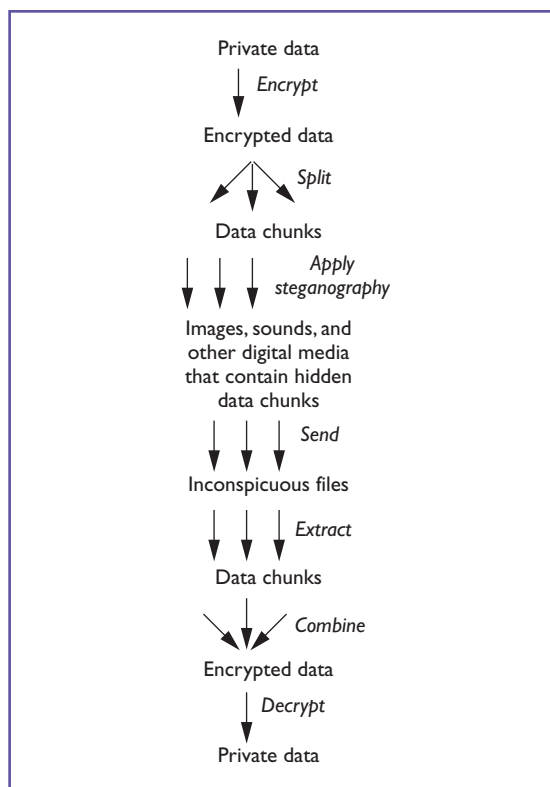


Figure 3. Process flow of secretly transmitting data. Using steganography, private data remains undetectable until it reaches its intended audience.

```

HTML
<IMG SRC="i1.jpg" ALT="image 1" NAME="i1">
<IMG SRC="i1.jpg" NAME="i1" ALT="image 1">

Perl
$d = $a + $b + $c;
$d = $a + $c + $b;
$d = $b + $a + $c;
. . .

E-mail
to: dono@drexel.edu,
    dono@lanl.gov
to: dono@lanl.gov,
    dono@drexel.edu
  
```

Figure 4. Examples of steganography using permutations. HTML and PERL offer steganographic potential, but steganography using data ordering can be limited by the number of permutations possible.

to variable \$c. The ordering of the variables to be summed is not constrained, thus two bits in this example can be mapped to a permutation of the variables. It is not possible to represent three bits, because three bits represent eight numbers and there are only six permutations.

Additional Approaches

Several other approaches to steganography have also been implemented for public consumption. StegParty (<http://www.fasterlight.com/hugg/>

projects/stegparty.html) is an entertaining and unique package that hides data inside a text file. By using a set of rules defining various flexible components within the constraints of the English language, StegParty can hide an admittedly small amount of data by matching the text against its rules and replacing it with small typos, grammatical errors, or equivalent expressions. The package could likely be extended to permit hiding information in HTML and other similar, loosely constrained markup languages.

StegFS (<http://ban.joh.cam.ac.uk/~adm36/StegFS/>) is a steganographic file system for Linux. It claims it can deny the existence of files (not an intentional feature of other file systems). Also worth mentioning is Psionic's Covert TCP (<http://www.psionic.com/papers/covert/>), which hides data in TCP packets. Data hidden inside a protocol not meant for human consumption can be very effective if the data remains intact throughout transmission. A lot of new networking hardware will rewrite packets in transmission, reducing the chance that hidden data will survive in a packet header.

Dono the Superspy

To illustrate a few of these methods, I will briefly describe an entirely fictional example scenario (see Figure 4).

Dono the dashing super spy is on assignment in a foreign country. His mission is to collect top-secret information from a local government source and report his findings to headquarters quickly without raising suspicion among the local authorities. Disguised as a common tourist, Dono manages to collect the information without being detected. He now faces the challenge of transmitting it to headquarters without being discovered by technically advanced local authorities. Fortunately, Dono is an expert in steganography.

Dono's first step is to encrypt all the data using headquarters' public key. No responsible super spy would ever skip this step as there is, of course, no substitute for encryption. Realizing that the local authorities would quickly notice an e-mail to headquarters, Dono takes another approach. Keeping his disguise as a tourist, he posts some pictures of the local attractions to a Web site back home. He also records a street performance and posts it as an MP3 to a public newsgroup. Later that evening in his hotel, Dono fires up his laptop to surf the Web and proceeds to visit his favorite site. Following the routine of any normal high-tech tourist, Dono cleverly maintains his secret identi-

ty. But what about the data that needed to get back to headquarters?

Dono's innocent-looking pictures of his touring adventures were, of course, not so innocent after all. Before posting them to the Web site, Dono hid part of the encrypted data within the coefficients of the JPEG images. Since this portion of the data was divided among several highly detailed pictures, the effects were virtually undetectable. The high-quality MP3 audio file that Dono posted to a public newsgroup also contained more than the sounds of the local culture. Part of the encrypted data was encoded in randomly chosen parts of the MP3, although you could never tell from listening to it!

Finally, our hero completed his mission by surfing the Web. At least that is how it appeared to those monitoring Dono's network moves. Using a special program that simulates natural Web page surfing at a predetermined site, Dono was able to transfer the remaining data to headquarters.

When a user surfs the Web, the user's browser sends an HTTP request to the specified Web server. By manipulating the unconstrained ordering of elements in the HTTP request, Dono's software could encode a few bytes of data per page surfed. On the Web server, which was operating in cooperation with headquarters, the specific ordering of elements in the HTTP request was decoded and the elements were translated back into the original encrypted data. Unfortunately, no software implementing this steganographic method is available to the public at this time.

Using these methods, Dono was able to transfer all the original encrypted data without communicating directly with headquarters and without having to explicitly transfer anything encrypted. Headquarters reassembled the data and decrypted the original data using their private key. Thanks to steganography, Dono was able to complete his mission successfully.

Detecting Hidden Code

Steganalysis, the official countermeasure to steganography, is the art of detecting and often decoding hidden data within a given medium.

Two major tools in steganalysis, information theory and statistical analysis, reveal in clear terms the tremendous potential for hidden information in Internet data — as long as a set of data can be compressed to a smaller size, there is room for hidden data within the medium. Accordingly, the path of seeking hidden data is treacherous and uncertain. Unless hidden data is encoded in a

Steganography Sites

Data Hiding Homepage

<http://nif.www.media.mit.edu/DataHiding/>

Information Hiding homepage

<http://www.cl.cam.ac.uk/~fapp2/steganography/>

Steganalysis

<http://www.jjtc.com/Steganalysis/>

Steganography and Digital Watermarking

<http://www.jjtc.com/Steganography/>

StegoArchive.com

<http://steganography.tripod.com/stego.html>

Watermarking Mailing List

<http://www.watermarkingworld.org/ml.html>

common, well-defined format, it may be virtually impossible to detect in the carrier data. In other words, a set bit can represent the United States' national archives or this article, depending on how I choose to define my encoding. To a steganalysis expert unable to determine the chosen encoding, a bit is just a bit. If you don't believe me, try to find the hidden message in this sentence. I imagine you are not going to have much success, unless I tell you that I encoded the secret message "I own striped pajamas" as the text of the sentence.

Steganalysis, though of great interest to businesses and governments alike, has not received the attention it deserves. Whether to benefit national security or to keep a competitive advantage in the market, the ability to control sensitive information is a critical part of maintaining a large institution. Steganalysis research has been stimulated by the increasing number of tools available for digital steganography (see the sidebar, "Steganography Sites"), yet steganalysis remains difficult to perform with great accuracy on some media. The growing field of cyberforensics — detective work in the digital domain — should create greater demand for steganalysis tools in the near future.

For further discussion on steganalysis, see Johnson and Jajodia⁴ and Johnson et al.⁵

Conclusion

The software and links mentioned in this article are just a sample of the steganography tools currently available. Steganography and steganalysis are beginning to receive increased attention as their applications become more relevant to the needs of governments, businesses, and individuals. As privacy concerns continue to develop

along with the digital communication domain, steganography will undoubtedly play a growing role in society. For this reason, it is important that we are aware of digital steganography technology and its implications.

Equally important are the ethical concerns of using steganography and steganalysis. Using steganographic techniques, software can easily transmit private user information without the user's permission or knowledge. Watermarks – already an issue in the hotly disputed domain of digital rights management – could be compromised by advanced steganalysis tools. Similar abuses of steganography and steganalysis can easily be enumerated.

Despite the lack of press given to steganography and steganalysis, these fields present interesting problems whose solutions will have profound effects on the Internet and Internet communication. Steganography, as mentioned, enhances rather than replaces encryption. Messages are not secure simply by virtue of being hidden. Likewise, steganography is not about keeping your message from being known – it's about keeping its existence from being known. With these points in mind, I ask

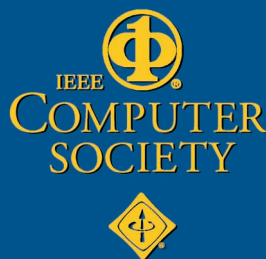
a final question: Is there another hidden message in this article? □

References

1. S. Katzenbeisser and F.A.P. Petitcolas, *Information Hiding Techniques for Steganography and Digital Watermarking*, Artech House, Boston, 2000.
2. B. Barán, S. Gómez, and V. Bogarín, "Steganographic Watermarking for Documents," *Proc. 34th Ann. Hawaii Int'l Conf. System Sciences*, IEEE CS Press, Los Alamitos, Calif., 2001.
3. H.K. Pan, Y.Y. Chen, and Y.C. Tseng, "A Secure Data Hiding Scheme for Two-Color Images," *Proc. Fifth IEEE Symp. Computers and Comm.*, IEEE Press, Piscataway, N.J., 2000.
4. N.F. Johnson and S. Jajodia, "Exploring Steganography: Seeing the Unseen," *Computer*, vol. 31, no. 2, Feb. 1998, pp. 26-34.
5. N.F. Johnson, Z. Duric, and S. Jajodia, *Information Hiding: Steganography and Watermarking – Attacks and Countermeasures*, Kluwer Academic Publishers, 2000. Available at <http://www.jjtc.com/Steganography/>.

Donovan Artz is a student at Drexel University and a research assistant at Los Alamos National Laboratory. Artz performs work for the national lab in Drexel University's Geometric and Intelligent Computing Laboratory, which is directed by William Regli.

Readers can contact the author at dono@drexel.edu.



Career Service Center

- Certification
- Educational Activities
- Career Information
- Career Resources
- Student Activities
- Activities Board

computer.org

Introducing the IEEE Computer Society Career Service Center

- Advance your career
- Search for jobs
- Post a resume
- List a job opportunity
- Post your company's profile
- Link to career services

computer.org/careers/