

# Steganography and Steganalysis

J.R. Krenn

January 2004

---

## What is steganography?

Steganography, coming from the Greek words *stegos*, meaning *roof* or *covered* and *graphia* which means *writing*, is the art and science of hiding the fact that communication is taking place. Using steganography, you can embed a secret message inside a piece of unsuspecting information and send it without anyone knowing of the existence of the secret message.

Steganography and cryptography are closely related. Cryptography scrambles messages so they cannot be understood. Steganography on the other hand, will hide the message so there is no knowledge of the existence of the message in the first place. In some situations, sending an encrypted message will arouse suspicion while an "invisible" message will not do so. Both sciences can be combined to produce better protection of the message. In this case, when the steganography fails and the message can be detected, it is still of no use as it is encrypted using cryptography techniques.

Therefore, the principle defined once by Kerckhoffs for cryptography, also stands for steganography: the quality of a cryptographic system should only depend on a small part of information, namely the secret key. The same is valid for good steganographic systems: knowledge of the system that is used, should not give any information about the existence of hidden messages. Finding a message should only be possible with knowledge of the key that is required to uncover it.

---

## New technology?

Steganographic techniques have been used for centuries. The first known application dates back to the ancient Greek times, when messengers tattooed messages on their shaved heads and then let their hair grow so the message remained unseen. A different method from that time used wax tablets as a cover source. Text was written on the underlying wood and the message was covered with a new wax layer. The tablets appeared to be blank so they passed inspection without question.

In the 20th century, invisible inks were a widely used technique. In the second world war, people used milk, vinegar, fruit juices and urine to write secret messages. When heated, these fluids become darker and the message could be read.

Even later, the Germans developed a technique called the *microdot*. Microdots are photographs with the size of a printed period but have the clarity of a standard type-written page. The microdots were then printed in a letter or on an envelope and being so small, they could be sent unnoticed.

Recently, the United States government claimed that Osama Bin Laden and the al-Qaeda organization use steganography to send messages through websites and newsgroups. However, until now, no substantial evidence supporting this claim has been found, so either al-Qaeda has used or created real good steganographic algorithms, or the claim is probably false.

Steganographic techniques have been used with success for centuries already. However, since secret information usually has a value to the ones who are not allowed to know it, there will be people or organisations who will try to decode encrypted information or find information that is hidden from them. Governments want to know what civilians or other governments are doing, companies want to be sure that trade secrets will not be sold to competitors and most persons are naturally curious. Many different motives exist to detect the use of steganography, so techniques to do so continue to be developed while the hiding algorithms become more advanced.

---

## Uses of steganography

With steganography you can send messages without anyone having knowledge of the existence of the communication. There are many countries where it is not possible to speak as freely as it is in some more democratic countries. Steganography can be a solution which makes it possible to send news and information without being censored and without the fear of the messages being intercepted and traced back to you.

While sending messages can be useful, it is also possible to simply use steganography to store information on a location. For example, several information sources like your private banking information, some military secrets and your mother's special pancake recipe, can be stored in a cover source. When you are required to unhide the secret information in your cover source, you can easily reveal your banking data and the recipe and it will be impossible to prove the existence of the military secrets inside. Steganography can offer *deniable* storage of information. The *Rubberhose* project (<http://www.rubberhose.org>) offers an implementation of this principle.

Because you can hide information without the cover source changing, steganography can also be used to implement watermarking. Although the concept of watermarking is not necessarily steganography, there are several steganographic techniques that are being used to store watermarks in data. The main difference is on intent, while the purpose of steganography is hiding information, watermarking is merely extending the cover source with extra information. Since people will not accept noticeable changes in images, audio or video files because of a watermark, steganographic methods can be

used to hide this.

---

## Implementing steganography

Secrets can be hidden inside all sorts of cover information: text, images, audio, video and more. Most steganographic utilities nowadays, hide information inside images, as this is relatively easy to implement. However, there are tools available to store secrets inside almost any type of cover source. It is also possible to hide information inside texts, sounds and video films for example. The most important property of a cover source is the amount of data that can be stored inside it, without changing the noticeable properties of the cover. When an image is distorted or a piece of music sounds different than the original, the cover source will be suspicious and may be checked more thoroughly.

### *Hiding a message inside a text*

Since everyone can read, encoding text in neutral sentences is doubtfully effective. But taking the first letter of each word of the previous sentence, you will see that it is possible and not very difficult. Hiding information in plain text can be done in many different ways. The first-letter algorithm used here is not very secure, as knowledge of the system that is used, automatically gives you the secret. This is a disadvantage that many techniques of hiding secrets inside plain text have in common.

Many techniques involve the modification of the layout of a text, rules like using every  $n$ -th character or the altering of the amount of whitespace after lines or between words. The last technique was successfully used in practice and even after a text has been printed and copied on paper for ten times, the secret message could still be retrieved.

Another possible way of storing a secret inside a text is using a publicly available cover source, a book or a newspaper, and using a code which consists for example of a combination of a page number, a line number and a character number. This way, no information stored inside the cover source will lead to the hidden message. Discovering it, relies solely on gaining knowledge of the secret key.

### *Images*

Hiding information inside images is a popular technique nowadays. An image with a secret message inside can easily be spread over the world wide web or in newsgroups. The use of steganography in newsgroups has been researched by German steganographic expert Niels Provos, who created a scanning cluster which detects the presence of hidden messages inside images that were posted on the net. However, after checking one million images, no hidden messages were found, so the practical use of steganography still seems to be limited.

To hide a message inside an image without changing its visible properties, the cover source can be altered in "noisy" areas with many color variations, so less attention will be drawn to the modifications. The most common methods to make these alterations involve

the usage of the least-significant bit or *LSB*, masking, filtering and transformations on the cover image. These techniques can be used with varying degrees of success on different types of image files.

### *Least-significant bit modifications*

The most widely used technique to hide data, is the usage of the LSB. Although there are several disadvantages to this approach, the relative easiness to implement it, makes it a popular method. To hide a secret message inside a image, a proper cover image is needed. Because this method uses bits of each pixel in the image, it is necessary to use a lossless compression format, otherwise the hidden information will get lost in the transformations of a lossy compression algorithm.

When using a 24 bit color image, a bit of each of the red, green and blue color components can be used, so a total of 3 bits can be stored in each pixel. Thus, a  $800 \times 600$  pixel image can contain a total amount of 1.440.000 bits (180.000 bytes) of secret data. For example, the following grid can be considered as 3 pixels of a 24 bit color image, using 9 bytes of memory:

```
(00100111 11101001 11001000)
(00100111 11001000 11101001)
(11001000 00100111 11101001)
```

When the character A, which binary value equals 10000001, is inserted, the following grid results:

```
(00100111 11101000 11001000)
(00100110 11001000 11101000)
(11001000 00100111 11101001)
```

In this case, only three bits needed to be changed to insert the character successfully. On average, only half of the bits in an image will need to be modified to hide a secret message using the maximal cover size. The resulting changes that are made to the least-significant bits are too small to be recognised by the human eye, so the message is effectively hidden.

While using a 24 bit image gives a relatively large amount of space to hide messages, it is also possible to use a 8 bit image as a cover source. Because of the smaller space and different properties, 8 bit images require a more careful approach. Where 24 bit images use three bytes to represent a pixel, an 8 bit image uses only one. Changing the LSB of that byte will result in a visible change of color, as another color in the available palette will be displayed. Therefore, the cover image needs to be selected more carefully and preferably be in grayscale, as the human eye will not detect the difference between different gray values as easy as with different colors.

Disadvantages of using LSB alteration, are mainly in the fact that it requires a fairly large cover image to create a usable amount of hiding space. Even nowadays, uncompressed images of  $800 \times 600$  pixels are not often used on the Internet, so using these might rise suspicion. Another disadvantage will arise when compressing an image concealing a secret using a lossy compression algorithm. The hidden message will not

survive this operation and is lost after the transformation.

### *Masking and filtering*

Masking and filtering techniques, usually restricted to 24 bits or grayscale images, take a different approach to hiding a message. These methods are effectively similar to paper watermarks, creating markings in an image. This can be achieved for example by modifying the luminance of parts of the image. While masking does change the visible properties of an image, it can be done in such a way that the human eye will not notice the anomalies.

Since masking uses visible aspects of the image, it is more robust than LSB modification with respect to compression, cropping and different kinds of image processing. The information is not hidden at the "noise" level but is inside the visible part of the image, which makes it more suitable than LSB modifications in case a lossy compression algorithm like JPEG is being used.

### *Transformations*

A more complex way of hiding a secret inside an image comes with the use and modifications of discrete cosine transformations. Discrete cosine transformations (*DST*), are used by the JPEG compression algorithm to transform successive 8 x 8 pixel blocks of the image, into 64 DCT coefficients each. Each DCT coefficient  $F(u, v)$  of an 8 x 8 block of image pixels  $f(x, y)$  is given by:

$$F(u, v) = \frac{1}{4}C(u)C(v) \left[ \sum_{x=0}^7 \sum_{y=0}^7 f(x, y) * \cos \frac{(2x+1)u\pi}{16} \cos \frac{(2y+1)v\pi}{16} \right]$$

where  $C(x) = 1/\sqrt{2}$  when  $x$  equals 0 and  $C(x) = 1$  otherwise. After calculating the coefficients, the following quantizing operation is performed:

$$F^Q(u, v) = \left\lfloor \frac{F(u, v)}{Q(u, v)} \right\rfloor$$

where  $Q(u, v)$  is a 64-element quantization table. A simple pseudo-code algorithm to hide a message inside a JPEG image could look like this:

```
Input: message, cover image
Output: steganographic image containing message
while data left to embed do
    get next DCT coefficient from cover image
    if DCT  $\neq$  0 and DCT  $\neq$  1 then
        get next LSB from message
        replace DCT LSB with message bit
    end if
    insert DCT into steganographic image
end while
```

Although a modification of a single DCT will affect all 64 image pixels, the LSB of the quantized DCT coefficient can be used to hide information. Lossless compressed images will be susceptible to visual alterations when the LSB are modified. This is not the case with the above described method, as it takes place in the frequency domain inside the image, instead of the spatial domain and therefore there will be no visible changes to the cover image.

### *Audio and video*

Hiding information inside audio files can be done in several different ways. Using the least-significant bit is possible, as modifications will usually not create audible changes to the sounds. Another method involves taking advantage of human limitations. It is possible to encode messages using frequencies that are inaudible to the human ear. Using any frequencies above 20.000 Hz, messages can be hidden inside soundfiles and will not be detected by human checks.

Also, a message can be encoded using musical tones with a substitution scheme. For example, a *Fis* tone will represent a 0 and a *C* tone represents a 1. A normal musical piece can now be composed around the secret message or an existing piece can be selected together with an encoding scheme that will represent a message.

Video files are generally a collection of images and sounds, so most of the presented techniques on images and audio can be applied to video files too. The great advantages of video are the large amount of data that can be hidden inside and the fact that it is a moving stream of images and sounds. Therefore, any small but otherwise noticeable distortions, might go by unobserved by humans because of the continuous flow of information.

---

## Detecting steganography

As more and more techniques of hiding information are developed and improved, the methods of detecting the use of steganography also advance. Most steganographic techniques involve changing properties of the cover source and there are several ways of detecting these changes.

### *Text*

While information can be hidden inside texts in such a way that the presence of the message can only be detected with knowledge of the secret key, for example when using the earlier mentioned method using a publicly available book and a combination of character positions to hide the message, most of the techniques involve alterations to the cover source. These modifications can be detected by looking for patterns in texts or disturbances thereof, odd use of language and unusual amounts of whitespace.

### *Images*

Although images can be scanned for suspicious properties in a very basic way, detecting hidden messages usually requires a more technical approach. Changes in size, file format, last modified timestamp and in the color palette might point out the existence of a hidden message, but this will not always be the case.

A widely used technique for image scanning involves statistical analysis. Most steganographic algorithms that work on images, assume that the least-significant bit is more or less random. This is however, an incorrect assumption. While the LSB might not seem to be of much importance, applying a filter which only shows the least-significant bits, will still produce a recognizable image. Since this is the case, it can be concluded that the LSB are not random at all, but actually contain information about the whole image. When inserting a hidden message into an image, this property changes. Especially with encrypted data, which has a very high entropy, the LSB of the cover image will no longer contain information about the original, but because of the modifications they will now be more or less random.

With a statistical analysis on the LSB, the difference between random values and real image values can easily be detected. Using this technique, it is also possible to detect messages hidden inside JPEG files with the DCT method, since this also involves LSB modifications, even though these take place in the frequency domain.

### *Audio and video*

The statistical analysis method can be used against audio files too, since the LSB modification technique can be used on sounds too. Except for this, there are several other things that can be detected. High, inaudible frequencies can be scanned for information and odd distortions or patterns in the sounds might point out the existence of a secret message. Also, differences in pitch, echo or background noise may raise suspicion.

Like implementing steganography using video files as cover sources, the methods of detecting hidden information are also a combination of techniques used for images and audio files. However, a different steganographic technique can be used that is especially effective when used in video films. The usage of special code signs or gestures is very difficult to detect with a computer system. This method was used in the Vietnam war so prisoners of war could communicate messages secretly through the video films the enemy soldiers made to send to the homefront.

---

## **Defeating steganograms**

While steganograms may not always be successfully detected, there are different ways of removing hidden messages from possible cover sources. Knowledge or certainty of the existence of a hidden message is not needed, since messages can even be destroyed without this. Although there will never be a 100 percent guarantee of success, the number of possible ways of sending hidden messages can easily be reduced using any combination of steganographic defeating techniques.

### *Text*

The best way of removing hidden messages from a plain text might be rewriting and reformulating the contents. Rewriting it using different words and sentence constructions will most certainly remove all ways of reproducing a hidden message, since it will take care of almost every possible way data can be stored inside a plain text. The character position scheme will no longer work because the words have been changed, and the same is valid for the differentiations in whitespacing, since the text will have a new layout.

The only method that will not be covered by this technique is the usage of a publicly available cover source. Since this source cannot easily be altered, there is no effective way of stopping this method, except for intercepting the secret key.

### *Images*

Compressing an image using lossy compression will remove messages that are hidden using the LSB modification technique. This will also happen when the image is resized, the color palette is modified or the colors themselves are modified. Conversion to a different image format, which often uses a different type of compression, will also help in removing hidden messages. And altering the luminiscence for example, will remove watermarks in the visible part of an image.

### *Audio and video*

Most of the techniques that can be used on images, can also be applied on audio files. Compressing an audio file with lossy compression will result in loss of the hidden message as it will change the whole structure of a file. Also, several lossy compression schemes use the limits of the human ear to their advantage by removing all frequencies that cannot be heard. This will also remove any frequencies that are used by a steganographic system which hides information in that part of the spectrum.

Another possible way of removing steganograms is lowering the bitrate of the audio file. In that case, there will be less available space to store hidden data and therefore, at least parts of it will get lost.

For video, once more again, the same methods as for images and audio files can be applied to remove hidden information. To defeat the use of signals or gestures however, human insight is still necessary, as computer systems are not yet capable of detecting this with a reasonable rate of success.

---

## **Conclusion**

Steganography, especially combined with cryptography, is a powerful tool which enables people to communicate without possible eavesdroppers even knowing there is a form of communication in the first place. The methods used in the science of steganography have advanced a lot over the past centuries, especially with the rise of the computer era. Although the techniques are still not used very often, the possibilities are endless. Many

different techniques exist and continue to be developed, while the ways of detecting hidden messages also advance quickly.

Since detection can never give a guarantee of finding all hidden information, it can be used together with methods of defeating steganography, to minimize the chances of hidden communication taking place. Even then, perfect steganography, where the secret key will merely point out parts of a cover source which form the message, will pass undetected, because the cover source contains no information about the secret message at all.

In the near future, the most important use of steganographic techniques will probably lie in the field of digital watermarking. Content providers are eager to protect their copyrighted works against illegal distribution and digital watermarks provide a way of tracking the owners of these materials. Although it will not prevent the distribution itself, it will enable the content provider to start legal actions against the violators of the copyrights, as they can now be tracked down.

Steganography might also become limited under laws, since governments already claimed that criminals use these techniques to communicate. More restrictions on the use of privacy-protecting technologies are not very unlikely, especially in this period of time with great anxiety of terrorist and other attacks.