Cybersecurity and National Defence

–

Steganography: steganalysis

Alberto Eusebio, Francesco Imparato, Federico Morro,
Salvatore Orlando, Giulia Rivetti

Abstract

Contents

1 Introduction

Steganography is the technique of hiding a message inside another message or a physical object.[1] The word *steganography* comes from the Greek word *steganographia*, which is the combination of *steganós* meaning "covered" and *-graphia* meaning "writing".

The first testimonials of the use of steganography date back to 440 BC in Greece mentioned by Herodotus in his *Histories*: Histitateus sent a message to Aristagoras by writing a text message on the shaved head of one of his servants and then waited till the hair of the servant had regrow to sent him to Aristagoras. Moreover steganography has been used for centuries in different ways such as secret inks, morse code hidden inside physical objects or encoded in eyes blinking (Jeremiah Denton, tortured prisoner-of-war in 1966 during the korean war, encoded in this way an help message during a TV report) or microdots embedded in paper or in clothes used by espionage agents during and after the World War II.

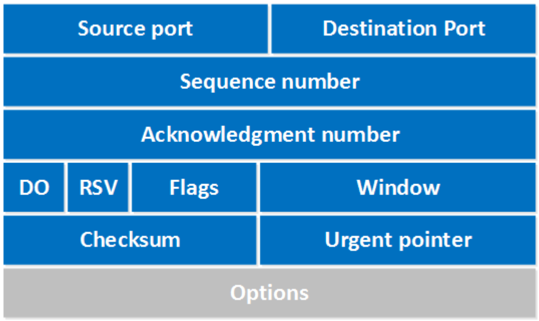
In digital steganography a message or a file is concealed within another file; in particular, electronic communications may contain a steganographic coding inside a information vehicle such as a document, a program or a media file (image, audio or video). Media files are ideal for hiding messages since due to their large size, the modification needed to encode a steganographic coding cause a subtle change that is unlikely to notice for someone who is not looking for it. This is one of the greatest differences of steganography with respect to cryptography: in cryptography the encrypted messages are visible so they attract interest and it's more likely that they will be subject to some type of attacks to be decoded.

In this paper we will focus on steganalysis, the study of detecting and, if possible, recovering hidden messages encoded using steganography, the way in which it can be performed, examples of its application in known cases and the threats and opportunities regarding the cybersecurity implications of its use in digital and communications systems.

We can call steganalysis successful (and consequentially that the steganography is broken) when there is the evidence of an hidden message in the cover signal. Moreover, the embedded information can also be crypted, in that case the attacker (the one who tries to find the hidden message) will have to perform cryptanalysis in order to decrypt the information.

The Field on Information Security Systems is indeed vast and will require much more space than the one we have. Here we attach a map which describes shortly the structure.

Figure 1: Map of Information Security Systems [2]



2 Steganography

As we already stated, steganography is used to hide a message in a *cover type*. We can define a cover type as the mean used to communicate, which can be a text file, an image, an audio recording, but also people, like in the case of the servants that we have mentioned earlier, etc...

In all these cases, the message has been *embedded* into the cover type. Here we can make a distinction between steganography and watermarking since in both cases a message is encoded into a communication's mean, but with different purposes.

2.1 Steganography

Steganographic methods may be grouped under two different classifications.

For the first, we distinguish injective and generative steganography. While injective steganography consists in finding an optimal way of injecting the message for a given cover type, generative steganography aims at generating an adequate container optimal for hiding a sought message.

As for the second classification, we mainly discriminate between substitutive and constructive steganography.

Substitutive steganography targets the noise in signals, to substitute the atmospheric noise with a secret message, whether it concerns audio or digital image signals. Most commonly, it may be achieved by replacing the LSB (Least Significant Bit), which is most susceptible to error, with the hidden message in the matrix of signals. Several drawbacks arise, such as the limited size of the message to be inserted to avoid amplified distorsions of the original signals, the resilience to various degrees of compression and signals transformations, and lastly the transparency of the message. We leave for a further discussion the different available tradeoffs, selectively mitigating some of the aforementioned drawbacks.

Similarly, constructive steganography integrates a noise model. The main drawback is that it is hard to produce, and by design fragile to attacks.

2.2 Watermarking

3 Steganalysis

Steganalysis is the branch of study dedicated at analysing the methods and the vectors used to transmit an hidden message in order to retrieve such message whenever it is present. Unlike cryptanalysis where the message may even be apparent but encrypted, in steganalysis the study of the message starts from a suspect.

3.1 Methods

Hereinafter we will present the two type of methods used in steganalysis in order to obtain the secret message.

3.1.1 Statistical

3.1.2 Structural

3.2 Cover types

In the following sections, we will treat several *cover* in which steganography can be applied to hide data. In particular we will focus on some of the most common methods and how is possible to do a steganalysis process in order to detect and in some cases even retrieve the embedded information. Notice that different cover types requires completely different approaches due to intrinsic factors such as size, presence of redudant data, perceptibility of modification by a human or formats in which the bits of the cover are stored.

3.3 Images

Images are likely to be chosen to hide secret messages due to the low sensibility of the *human visual system* (HVS) to some particular attributes such as small changes in luminance or brightness or contrast near figures edges. There are several methods to apply steganography to an image, in the following sections we will treat some of the most common hiding techniques and some of the steganalysis methods to attack them.

3.3.1 Stego-key search and encryption

When in the following sections we will mention modification applied to *LSB* of some data characterizing the images, it's crucial to mention that not always all LSB are modified nor are modified in subsuqeunt blocks. More complex steganographic techniques imply the use of some pseudo-random walk followed when deciding which data to modify, generated by some *stego-key* (usually stego-key will be mapped in a set of possible seeds by a hash map).

It's also possible to apply an *encryption* algorithm before embedding the message in order to make steganalysis harder, since the attacker cannot find any recognizable bit sequence, when searching for the message.

The use of such systems makes the steganalysis process much harder since it becames unfeasible the use of a brute force approach: the complexity would be proportional to the cardinality of the set of possible seeds time the one of the set of possible encryptions.

Moreover, even if the LSB are modified in blocks and no encryption is applied, the steganalysis methods are useful to comprehend (without need to find recognizable bit sequences) if the images contains secret messages in a systematic way.

3.3.2 LSB embedding method

LSB embedding method is arguably the most popular steganographic method, due to its simplicity, high imperceptibility and high capacity. In this method, the image is decomposed into *bits plane* (8 bits per pixel for grayscale and 24 for color images, one for each color channel) and the *least significant bit* (LSB) is substituted with the message to be hidden. Note that even if the message is encrypyted, due to its simplicity, this method is easily detectable with a statistical steganalysis attack.

3.3.3 Difference Image Histogram method

The *Difference Image Histogram method* is derived by the easier idea of analysing the *histogram distribution* of a natural image and its stegoimage. Anyway, when we are steganalysing an image, we *do not* have the natural image, so what we could do is comparing the histogram distribution of the suspect image with a set of natural images, but the problem is that the variation between two different images is *bigger* than the distribution variation between a natural image and its stegoimage. [3]

The proposed way to proceed is the following [4], starting from the test image (that we will call h, considering h a grayscale image or a color image under some assumptions) we generate:

1. an image f given by h with flipped LSB
2. an image g given by h zeroing the LSB
3. images D_h, D_f, D_g, created by the respective image of each one with the following formula:

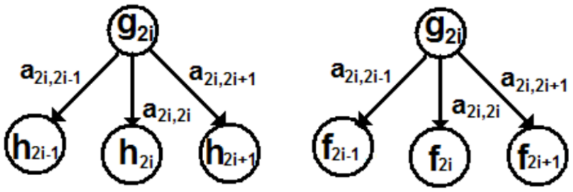
   $$D(i,j) = I(i+1,j) - I(i,j)$$

   where I is the image denoted by the subscript and the couple (i,j) a unique pixel of the image

At this point, we can analyse the histograms of the three generated images D_h, D_f, D_g:

$$H = \{h_i | i = -255,\ ...,\ +255\},\ F = \{f_i | i = -255,\ ...,\ +255\},\ G = \{g_i | i = -255,\ ...,\ +255\}$$

Figure 2: Transition values from G to H and F



then, we define the following values:

$$\alpha_i = \frac{a_{2i+2,2i+1}}{a_{2i,2i+1}}$$

$$\beta_i = \frac{a_{2i+2,2i+3}}{a_{2i,2i-1}}$$

$$\gamma_i = \frac{g_{2i}}{g_{2i+2}}$$

As found out by X. Ping and T. Zhang in [4]:

- if $\alpha_i \approx 1, \forall i \in \{-255, -254, ..., 255\}$ then the image contains some hidden message;
- otherwise, for natural images, $\alpha_i \approx \gamma_i$ is satisfied.

3.3.4 Closest Color Pair method

Another method used to detect hidden messages on the LSB plane is the *Closest Color Pair method.*

Whhen an image has a steganographed message in te LSB plane, the number of close colors increases [5]. Given two pair of colors $C_1 = [R_1, G_1, B_1]$, $C_2 = [R_2, G_2, B_2]$, the condition of them being close is:

$$(R_1 - R_2)^2 + (G_1 - G_2)^2 + (B_1 - B_2)^2 \leq 3$$

We apply a LSB embedding steganography algorithm on the image and we compute the number of close color pairs in both images.

Now, we define:

- P as the number of close color pairs; P' as the number of close color pairs of the stegoimage;
- U as the number of color pairs; U' as the number of color pairs of the stegoimage

Then, we compute:

$$R = \frac{P}{\binom{U}{2}},\ R' = \frac{P'}{\binom{U'}{2}}$$

We know that if

$$\frac{R}{R'} \geq Th,\ Th = 1.1$$

then the image is a natural image, otherwise it contains an hidden message. All the proofs are in [5].

3.3.5 DCT Domain embedding method

The DCT Domain embedding method works on *JPEG* images, one of the most used format on Internet web sites. The JPEG format uses a *discrete cosine transform* (DCT) to transform every 8x8-pixel block into 64 DCT coefficients, that are used to calculate the pixels when the image is displayed. The DCT Domain embedding method substitute the *LSB* of the coefficients with the secret message: since the modification is done in the frequency domain, there is no human perceivable change in the image.

3.3.6 Chi-square test

3.4 Audio

Steganography in audio is more challenging with respect to images, because the *human auditory system* (HAS) operates over a wide dynamic range, while maintaining a high sensitivity to perturbations and noises. Howerer, there are still some "holes" where data can be hidden. The HAS has a quite small differential range, so loud sounds mask out quiet sounds, moreover it is unable to perceive absolute sound phase, but only relative one.

Another important factor to consider when dealing with sound are the transmissions enviroments. Audio signals can be transmitted through a digital channel (eventually being resampled), through an analog channel or "over the air" played by a speaker and received by a microphone. Depending on the transmission channel there could be huge modification that can make the steganalysis process impossible, but also that can compromise irreparably the hidden message, damaging also the steganographer.

Due to the several issues presented above, steganalysis in audio signals is more complex with respect to other cover types. In the following sections we will cover some of the possible ways in which steganography and steganalysis is applied in audio signals, without entering too much in details.

3.4.1 Low-bit coding

3.4.2 Phase coding

3.4.3 Spread spectrum

3.4.4 Echo data hiding

3.5 Text

Soft-copy text is the most difficult place in which steganography is performed due to the fact that there are no redundant informations nor large set of data that can be slightly modified without affecting the overall look of the cover.

3.6 TCP/IP

3.6.1 Introduction

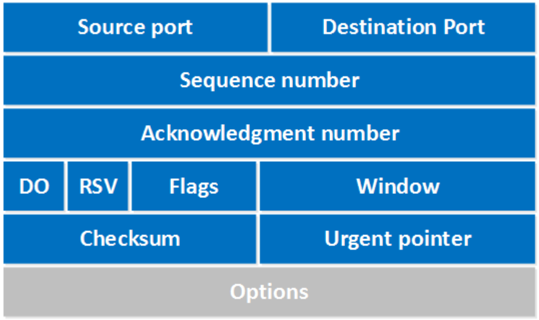
Before treating such topic we must briefly discuss the behaviour of the TCP/IP (**Internet protocol suit**) communications.

IP (Internet Protocol) is at the basis of internet communication. The protocol exploits the principle of encapsulation, sending packets composed by a *header* which contains information such as the **destination address** and **source address** and a *payload* which represents the actual data to be transmitted. Since physical channels usually have a limited MTU[1], then the data is usually fragmented in smaller chunks, wrapped into an header and only then transmitted.

TCP (Transmission Control protocol) is a protocol used for reliable (little information loss), ordered, error-checked data transmission between a machine hosting the data (**Server**) and another requiring such data (**Client**). It is performed previous a three-way handshake which estabilishes a connection between server and client, thus preparing a reliable communication channel.

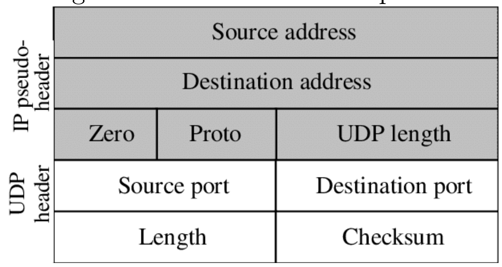
UDP (User Datagram Protocol) , contrary to TCP, is less reliable but faster. It broadcasts the data in an unordered and uncontrolled way to the receiver. There is no need to estabilishing a connection in order to implement such protocol.

Figure 3: Structure of TCP packet



[1] Maximum Transmission Unit

Figure 4: Structure of UDP packet



3.6.2 Exploits and protection tools

As you could notice from the previous figures, the first part of the header is the same. This is because it is part of the IP protocol itself. In the second part of each header we find different types of metadata which are specific for each protocol. Generally the metadat contained in the header is redundant for the transmission of the message itself and this redundancy may become a target for a *stego-algorithm*[2] which starting from a *cover-network packet sequence*[3] and a *covert message*, can generate a sequence of packets(each one embedding a portion of the covered message) which will be sent over the network.

This procedure is not a *risk-free* approach for the attacker who wishes to send the *cover-network packet sequence*, since it is possible that the data will be currupted during transmission or there could be losses using non-reliable transport protocols(UDP). Other criticalities of such method stay in the fact that the sequence of packets will most likely traverse multiple nodes in the network before reaching the target, and the message may be detected by these nodes.

The defense against Steganography in TCP/IP communication consists in series of standards which are implemented by the devices on the internet and that we will illustrate through some examples.

Tof(Type of service) bits are a field in the TCP header which are nowdays rarely used. This could open doors to a steganographic attack if modern operating systems did not set them to zero by default. A warden monitoring the channel could immediately signal an error.

IP ID is a field in the Internet Protocol which is used to assist the receiver in reassemblying the fragmented packets. This field consist of randomly unique numbers representing a packet. It is possible to insert other types of information in this field by simply conform to the uniqueness constraint. Since in many cases the numbers used for the IP ID are not random, by knowing the charachteristics of the sender it is possible to detect an infiltration.

IP Fragment Offset is an offset which is present in the IP header which helps the receiver to reconstruct the sequence of bits from the fragmented sequence. Modulating the size of the fragments changes the offset field in the header and thus a message could be sent. The protection against this method is simply checking the size of the packets relative to the MTU and so even in this case a warden can easily detect an error.

TCP sequence number is a field in the TCP header which stores the randomly chosen position(for security reasons) of the first byte to be transmitted through the channel. The steganographic method consists in replacing this field with the data to be sent. Being random it is more difficult to spot a breach in the channel. In this case the usage of a SVM[4], a machine learning tool able to identify

[2] an algorithm used for steganography

[3] a sequence of packets which will be transmitted over the network acting as a cover for the hidden messace

[4] Support Vector Machine

patterns inside the data transmitted could come into hand. However an error could be detected even simply by checking the presence of repetitions in the stream(not admitted by design).

Timestamp modulation is another technique of steganography which operate by modifing the *LSB* of the timestamp of a TCP packet in order to represent a '1'. The covert message is thus embedded into the data stream. Since the TCP Timestamp support is not universal, machines not supporting such feature may detect the hidden message.

3.6.3 Detection of TCP/IP steganography

Each operating system exhibits well defined characteristics in generated TCP/IP fields. These can be used to identify any anomalies that may indicate the use of steganography. For this purpose, a suite of tests which may be applied to *network traces*[5] are defined and they are used to identify whether the results are consistent with the operating systems believed to be installed on the source host. Diffrent methods of covert channel detection are used, employing IP ID characteristics, TCP ISN characteristics and explicit steganography detection.

IP ID characteristics are the features identifying the IP address, a unique address that identifies a device on the internet or local network. They are employed by the following methods. *Sequential Global IP ID* implies the usage of a global counter for the IP ID. To detect this strategy one has to look if connections to different hosts have sequentially increasing IP IDs. *Sequential Per-host IP ID* is characterized by the usage of a per-host counter for packets. If it apperas to be fragmented, the warden can test whether connections to different hosts use apparently unrelated IP IDs; however connections to the same host have a sequentially increasing IP ID. *IP ID MSB Toggle* represents the case where the operating system system toggles the most significant bit of the IP ID every rekey interval, so that the warden can examine the MSB to check if it matches this pattern. *IP ID Permutation* strategy presence can be discarded by the warden if there are duplicates, since within a rekey interval the IP ID is non-repeating.

Explicit steganography detection can be employed by several methods. *Nushu Criptography* is a strategy applied by Nushu, which encrypts data before including it in the ISN field, resulting in a distribution which differs form the one that is normally generated by Linux; therefore, it can be detected. *TCP Timestamp* strategy involves the execution of a randomness test. In particular, if a low bandwidth TCP connection is being used to leak information, this test can be applied to the LSBs of the timestamps in the TCP packets. If an excessive presence of randomness is detected in the LSBs, it can be deduced that a steganographic covert channel is in use. There are also other features which may indicate the usage of steganography: *unsual flags*, *excessive fragmentation*, use of *IP options*, *unexpected TCP options* and *excessive re-ordering.*

[5] function that performs network analysis on a geometric network

References

[1] Merriam Webster. *Steganography definition.* URL: https://www.merriam-webster.com/dictionary/steganography.

[2] Milad Taleby Ahvanooey, Qianmu Li, Jun Hou, Ahmed Raza Rajput, Chen Yini. "Modern Text Hiding, Text Steganalysis, and Applications: A Comparative Analysis". In: (2019).

[3] A. Hernandez-Chamorro, A. Espejel-Trujillo, J. Lopez-Hernandez, M. Nakano-Miyatake, H. Perez-Meana. "A Methodology of Steganalysis for Images". In: (2009).

[4] X. Ping T. Zhang. "A new approach to reliable detection of LSB steganography in natural images". In: (2003). DOI: https://doi.org/10.1016/S0165-1684(03)00169-5.

[5] R. Du J. Fridrich M. Goljan. "Detecting LSB Steganography in Color and Gray Scale Image". In: *IEEE Multimedia* 8.4 (2001).