DOI: 10.5604/01.3001.0012.8021

STEGANOGRAPHY METHODS USED IN ATTACKS ON INFORMATION AND COMMUNICATION SYSTEMS

Anna Romanova, Sergiy Toliupa
Taras Shevchenko University of Kyiv, Faculty of Information Technologies

Abstract. *An analysis of steganography methods that are can be potentially used as instruments in attacks on information and communication systems is presented. The possible solutions to ensure resilience to such attacks are presented.*

Keywords: steganography, TEMPEST, covert channel, information protection

ZASTOSOWANIE METOD STEGANOGRAFICZNYCH DO ATAKÓW W SYSTEMACH INFORMACYJNO-KOMUNIKACYJNYCH

Streszczenie. *W artykule został przedstawiony przegląd istniejących i potencjalnie dostępnych technik steganograficznych, które mogą zostać użyte jako narzędzia do ataków na systemy informacyjne i komunikacyjne. Podano możliwe sposoby zapewnienia ochrony przed takimi atakami.*

Slowa kluczowe: steganografia, TEMPEST, ukryty kanał transmisji, ochrona informacji

Introduction

Cryptography is widely used as one of the most efficient and approbated methods of critical information resources protection. Nevertheless, in particular cases it might be more effective to hide the communication channel itself instead of making the information within it unreadable. Such a practice, namely – concealing data within unsuspicious, innocent-looking containers – is called steganography.

While being primarily considered a means of information security assurance, steganography can be used with ill intentions, as well. In fact, several high-tech attacks are based on the hidden data transmission, and contemporary methods of counteraction do not provide satisfactory level of resilience to those. The main advantage of steganography becomes the main source of threat – the channels of the attack, not to mention information about the attacker left in the channels, cannot be identified due to the nature of the method. In other words, attacks become invisible, as does the transmission channel. Even the fact of trespassing cannot be easily proven.

The purpose of this article is to conduct an analysis of attacks that are carried out with the use of steganography methods as their basis, and are directed against information and communication systems. Both existing and potential methods are presented.

1. Steganography as a means of hiding information

1.1. Basic terminology

Steganography is an art and science of storing and transferring secret messages within covert channels that are based on and created inside open channels in such a way that the cover data is perceived as if not having any embedded messages for its unplanned recipients.

The main concepts are:

- Container b (also: carrier) is open data used to conceal secret information.
- Message m (also: payload) is secret information to be concealed.
- Key k is secret information that is known only to a legitimate user and defines a specific concealment algorithm.
- Empty container c (also: unmodified container) is a container devoid of any secret data; it is a sequence of lc-long elements.
- Modified container s (also: package, steganogramme) is the one that contains a secret message.
- Steganographic algorithm means two transforms, a direct F: M x B x K → B and an inverse one F–1: B x K → M.
- Steganographic system (also: steganosystem) is a totality of messages, secret keys, containers and transforms that connect them [8, 13].

Most steganography methods are based on two key principles:

- Human senses cannot distinguish slight changes in colour, shape and sound perception.
- Consequently, there are files that do not demand absolute preciseness and therefore can be modified without losing their functional value.

As a result, said methods imply allocation of insignificant fragments of the container and replacement of the information within them with information that needs to be hidden.

Finally, the process of encoded steganogramme detection is called *steganoanalysis*.

1.2. Popular steganographic solutions

In this section the brief overview of widely used steganographic solutions is presented.

Mostly, steganography uses the data concealment within digital images and audio files, less so video files and text. Electronic communications may also include hiding data inside of a transport layer (program or protocol) [6].

Starting with non-digital methods, physical steganography technics cannot be omitted. They have been developing for centuries and include, for example, blinking one's eyes in Morse code to spell a secret message [10].

Another example is adding tiny yellow dots to each page while printing a document. They are not detectable by the bare eye and contain the model, serial number and timestamps. This information cannot be obtained from a computer file and is embedded in a printout using dot-matrix code. The technology is used by many brand color laser printers, such as Xerox and Hewlett-Packard for traceability reasons [12].

Methods of embedding data within an image container are [5, 8]:

- Least Significant Bit method (LSB) (Sequential Insertion) is the most popular steganographic method. The least siginficant bit of each pixel is in fact a noise. If it is changed, the difference in the image will not be noticed by a human eye. Thus, these bits can be replaced with the bits of a secret message.
- LSB Psuedo Random Insertion. In contrast to the previous method, in which every changed data bit follows the next, this method uses pseudo random distribution of the secret message bits through the container. Thus, the interval between two bits is pseudo-randomly defined, which complicates both visual and statistical attacks, as well as extraction of all the hidden bits.
- Block hiding method. The container is split into disjoint blocks; for each of them a parity bit is calculated. One secret

bit is concealed within one block. If the parity bit does not equal the respective secret bit, then one of the LSB in the block is inverted, so that the parity and the secret bits are the same.

- Palette permutation. Any colour palette consists of pairs of indexes. Each pixel of the image correspondsto a certain index in the table. The sequence of colours in the palette is not important, so it is possible to conceal a covert message by changing this sequence.
- Koch-Zhao (Relative DCT (Discrete Cosine Transform) values change method). Initial image is split into blocks of 8×8 pixels. As the result of applying DCT to every block a table of DCT coefficients is formed. Every secret bit is hidden in a separate block. Frequencies quantization causes some rate of distortion in the image, which is still not noticeable by the human eye.
- Benham-Memon-Yeo-Yeung method. Optimized version of the previous method. Firstly, only the most suitable blocks are used. Secondly, three DCT coefficients are selected instead of two, which decreases destortion in the container.
- Fridrich method implies a cascade embedment in low- and high-frequency DCT coefficients.
- Spread-Spectrum method consits of three possible variants:
  - The used frequency band is much wider than needed;
  - Spectrum is expanded by using a special independed (also: code) signal. The signal energy is distributed through all frequency bands, which makes the signal noise immune;
  - Restoration of the initial information is carried out by comparing the received signal and a synchronized copy of the code signal.
- Embedding pictures within video-files [10].

Audio steganography [8]:

- LSB-method for audio-files is the same as for images, but working with the audio-file format. It causes considerable destortions in the container.
- Phase coding method implies the substitution of the initial sound segment phase with the reference phase, which is the data to be concealed. Phases of adjacent segments are agreed to preserve the difference phase between them.
- Echo-signal use. Data is embedded in the container by injecting an echo-signal in it. Three echo-signal parameters are changed: initial amplitude, attenuation and shear rate. The echo-signal is perceived only as an additional resonance [4].

Linguistic steganography [8]:

- Random interval methods. Changing the number of spaces in the end of the text string does not cause significant changes in the meaning of the sentence. What is more, an average reader is unlikely to detect insignificant space modifications:
  - Changing the interval between sentences. One or two additional spaces are added after the sentence.
  - Changing the number of spaces in the end of text lines. Spaces are added according to the secret bit to be hidden. Two spaces encode one bit a line, four spaces – two bits etcetera.
  - Changing the number of spaces between words in a flattened text. When the text is width aligned, spaces between words are not of the same length and some of them can be used to hide data.
- Making the text of the same colour as the background [10].
- Using similarly looking Unicode and ASCII characters [2, 6].
- Using non-printable Unicode characters [2].
- Creating a pattern of deliberate errors and/or marked corrections [6].

Some other methods:

- Converting a file so that it has the statistical characteristics of another one [6].
- Format steganography.
- Blog-steganography. Secret data is added as commentary pin boards on social networks [10].

Finally, there are different software applications that use the methods of steganographic concealment mentioned above:
Using LSB-method: OutGuess, JSTEG, JPHS, Hide-and-Seek, Steganos, Steghide, DC-Stegano.
Using the palette permutation: Gifshuffle.
JPEG format: OutGuess, JSTEG, JPHS.
GIF format: Gifshuffle, Hide-and-Seek.
BMP format: Steganos, Steghide.
PCX format: DC-Stegano.
LSB-method in audio-files: Invisible secrets, Hide4PGP, Steghide, StegoWav, Steghan, S-Tools.
Using parity of quantization of frequency coefficients: MP3Stego.
Using incorrect frames in a compressed stream: UnderMP3Cover [1].

There are also several perspective steganography methods, the use of which is still limited, but nevertheless possible [11]:
1) Internet of Things and cyber-physical systems. A cyber-physical system is a mechanism that is controlled or monitored by computer-based algorithms, tightly integrated with the Internet and its users. Examples of CPS are autonomous automobile systems, medical monitoring, smart grids, automatic pilot avionics etc. The Internet of Things (IoT) is the network of physical devices, vehicles and other items embedded with electronics, sensors, software and network connectivity, which enable them to collect and exchange data. It is more or less an instance of a class of cyber-physical systems. The network steganography uses communication protocols' control elements and their functionality to hide information inside. The modifications can be carried out either over a single network protocol (applied to the Protocol Data Unit, the time relations between PDUs or both) or to several protocols at the same time (inter-protocol steganography). Such network steganography methods can be applied to the systems mentioned above, too. The IoT is believed to be a phenomenon that will expand its influence greatly within the next few years. As a perspective network instance it requires thorough attention of steganography specialists. Information circulates within it the same or the fairly similar way as in any other system. Thus, optimal and the most suitable methods of hiding data in communication protocols should be developed specifically for the IoT. What is more, as the items within the IoT possess a vast variety of sensors and software, they can be used to conceal data in. For example, covert messages can be stored in unused registers of the CPS/IoT components or in the states of their actuators.
2) The use of stream containers. As mentioned above, by the type of access to the data one can distinguish fixed and stream containers. All the methods mentioned in Chapter 2.3 use the first ones to conceal information in. Such a container is a constant pre-defined sequence of bits that are displayed before a steganographer all at once. To the contrary, a stream container is a sequence of bits that are continuously changing, as in a phone conversation. A message is embedded in real time so that the final size of the container is never known beforehand. The intervals between the embedded bits are generated by a pseudorandom sequence (PRS) generator and uniformly distributed between readouts. There is hardly a couple of scientific works devoted to this type of steganography, let alone examples of its real-life practical implementation. Despite any reasons, it can be successfully applied as an efficient means of information security. There is a number of solutions for encrypted secure real-time communication. However, what if we could, for example, make a confidential phone conversation not only indecipherable but also seem to be an innocent chat? A stenographic telephone set-top box could be a solution. The same concerns video-conferences. An extraneous observer would only see an average conversation not having any access to the real audio, video or any other embedded data. The unpopularity of the stream-container steganography can be explained by defining major issues concerning its use. First and foremost, it is never known whether the size of the container will be enough to conceal the whole message as the length of the first (and likely of the latter, as well) is undefined. The same property creates and advantage as one carrier file can be capacious enough to contain several messages.

In any case, the secret data has to be somehow synchronized with the container, thus one of the biggest questions is how to define the beginning and the end of the embedded sequence within the container. The problem becomes more serious concerning video communication. The solution would be of extreme complexity, as we would need to synchronize the image-image stream (both open and covert), the sound-sound stream and image and sound respectively. The solution may lie in using special built-in libraries. They would consist of structured groups of words of the same length, which would in ideal case possess pronunciation similarities. Such groups should then be grouped in semantic dictionaries, so that they would form simple, but logically and semantically structured sentences. The linguistic means for this are well-developed and are similar to those of forming synonymic dictionaries and machine translation applications. The words and sentences could then be synchronized with the container using synchronization bits, package headers and/or other means of dividing encapsulated data; the covert message can be embedded after them and be synchronized using the initial properties of the container.The possible situations with video communication would be more complex. If only the content of a given conversation is confidential, then the issue is just to steganographically encrypt the sound and synchronize it with the real video image. On the other hand, if the identities of conversation participants are also a secret, then other methods should be provided. It is not necessary for a steganographic solution to be all-purpose. It is possible to design a system consisting of a cryptographic and a steganographic modules and providing different scenarios according to the situation. The biggest remaining problem is a significant delay which is unacceptable in real-time conversations. Then again, there are numerous solutions in cryptography in this field, that can be adapted to the task.

2. Steganography methods used as instruments for attacks

2.1. Format steganography

Perhaps, the easiest and the most well-known way, which is actually a steganography method, is using legitimate features of file formats to carry hidden malicious software within their structure. A file of every format contains specific fields, which ensure that the former will be processed correctly on the target computer. Some of these fields are optional, or more strictly – information that they contain is not vital for the file. Thus, changing data bits in these fields most probably will not lead to errors while operating with the file. Such characteristics make these formats perfect containers [11].

A vivid example is a virus Win95.CIH – specific malware which is embedded in *.exe files by using Portable Executable format features. This format includes a lot of additional data which are grouped according to their functions. Every group gets its own section in the file structure, and the size of the sections is predefined. If they are not entirely filled with data, it means the file contains a lot of spare space. For example, the first section is only for the PE header, so a big part of the virus uses it as a covert container [3].

2.2. Soft Tempest

In fact, there are a lot of ways to covertly transmit necessary information to the target system. Not only harmless files but also network protocols can be used as efficient containers within the attacker's steganography system. Nevertheless, necessary means depend on the final objective of the attack. If the goal is to steal data, there is need for both an inward and an outward information flow. Getting information into a system is important. A more interesting question, though, is how to get the stolen data out without raising suspicion of a legitimate user.

While operating, every electronic device (including those inside a computer) gives off compromising emanations – electromagnetic emanations, which can be demodulated and accordingly processed to illegitimately get the critical information from them. These are called TEMPEST emanations after an American standard on the matter.

Contemporary TEMPEST-based attacks tend to become more and more sophisticated as the countermeasures are being continuously enhanced, as well. Systems are contaminated with the malicious software which then conducts the search of necessary information (key data, passwords, specific files etc) and induces the leak through TEMPEST emanation. For example, if reception of the signal is the one from the monitor, then the information will be, say, amplitude modulated and sent as a visual picture to the monitor. The obvious disadvantage is that such an activity cannot be missed by an operator and will be deemed highly suspicious, which, on its part, will lead to finding and neutralizing the virus.

M. Kuhn and R. Anderson conducted a series of experiment in which they shown a possible solution [9]. The human eye is more sensitive to low-frequency than to high-frequency vibrations, while TEMPEST receivers work vice versa. What is more, any devices primarily perceive luminosity in a linear way, while humans are more sensitive for the dark colours. This difference in sensitivity perception can be used to embed a message in the emanation and make it invisible to an unsuspecting user. The suggested method is to control and modify monitor dithering patterns. Pixels of two colours put in a check pattern are seen as a uniform colour, on the one side; on the other side, they create a high-frequency signal, which is best received by TEMPEST equipment with the following use of gamma-correction. Basically, the target computer is programmed so that it acts as a radio transmitter and emits a compound TEMPEST signal: a legitimate user observes one picture, and the attacker receives another – embedded – one on the monitor of his/her TEMPEST receiver.

The only suggested method of counteraction, which is specific enough for this very type of attack, is still based on using the difference in perception sensitivity between humans and devices. TEMPEST fonts are designed with top 30% of the Fourier transform of the signal removed, which is most probably not noticed by a human eye, but makes it impossible to receive a strong TEMPEST signal [9]. Nevertheless, special equipment with necessary parameters (enhanced sensitivity to low-frequency emanations) might be designed, which will make the use of such fonts ineffective.

2.3. Acoustic emanations as containers

Electromagnetic fields are not the only by-product of the computer systems operation. A. Shamir and E. Tromer published the results of their research, in which they showed that computer emit high-pitched noise while operation, due to vibration in some of their electronic components [7].

A series of experiments conducted by the scientist revealed that acoustic emanations can provide a potential attacker with information about what kind of software is currently running on the target system, as well as leak data on security-related parameters and computations. For example, loops of CPU instructions were highly distinguishable, and different RSA keys appeared to induce different sound patterns. To extract individual keys, the technic of acoustic cryptoanalysis was presented (applicable to GnuPG's implementation of RSA). According to the results, it takes about an hour to extract full keys from a target computer, irrespectively to their models and manufacturers. The key piece of equipment used for the attack is a microphone, and that of a mobile phone was demonstrated to be enough. Apart form acoustics, the scientists demonstrated a low-bandwidth attack, based on the same principles. The main difference was that the attacker had to get the leakage from ends of VGA, Ethernet, USB or other cables [7].

If electromagnetic emanations can be used as containers in steganography systems, acoustic waves can be, too. The first case could be based on the nature of sound perception itself – the classical steganography technic. Human hearing systems cannot distinguish slight variations in an acoustic flow. Here, any known method, mentioned above (Least Significant Bit, Echo-signal use etc) can be used to embed stolen information in parasitic sounds, emitted by the target computer. The second possible scenario is similar to the use of emanations in Soft Tempest. Sound dithering is a widely used method in music digital processing. The principle is the following: any piece of musical record might contain extensive frequency transitions that are too slow and smooth. This is where so called quantization noise can appear. If the level of frequency fluctuation is insignificant, the processing software simplifies the sound by removing the frequencies that exceed some medium limit. To cope with such a situation, special noises are generated and gradually added to the record. In music processing, this technic allows to achieve a natural sound lost during quantization.

It is possible to suggest, that the same technic can be used in attacking steganography systems. The noise emitted by a computer is quite stable. It is not foiled by fan system noise, as critical acoustic signals appear to be mostly above 10 kHz, while a typical fan noise along with other noises lie in a much lower frequency band [7]. Task-switching is not a problem either, as it is the tasks that carry distinguishable acoustic spectral signatures. The same can be said about several computers working simultaneously in a closed space: they can be told apart using different sound patterns, as their depend on specific hardware, temperatures inside and outside the system, humidity, and other conditions. Thus, it acoustic emanations seem to be a sufficient container, while dithering can be accordingly modified and applied as an embedment method.

The only suitable countermeasure seems to be the use of sound dampening equipment that can diminish the level of high-frequency leakage. As for means of active protection, strong wide-band noise source can serve for masking the critical data signals. Rough-scale behaving algorithms are another solution: despite somewhat diminishing the level of performance, they can thwart side-channel attacks by shuffling the signal and making it thus useless for the attacker [7]. In addition, electronic components of the system should be those of the highest quality, designed to reduce the level of acoustic and any other leakage.

Nevertheless, at this point, efficiency of such protection methods is rather relevant, as sound-proving degrades other performance features along with being quite expensive. At the same time, due to the need of ventilation, there are still open parts in the cases, so their structure has to be constructed to shuffle outcoming noises very efficiently.

3. Conclusion

Stenganography is a powerful means of information protection. Nevertheless, it has to be also regarded as an instrument for a potential attacker, with all of the advantages turned threats.

Compromising emanations of different physical nature are invisible and can only be noticed with the use of special equipment. Using steganography technics for the attacks ensures that the fact of using those emanations is efficiently hidden, and the system operations remains unsuspicious. This is exactly why there is need to consider technics described above a real threat for information and communication systems, and to join academic and technical potential to develop cost-effective and technically efficient counteracting means.

References

[1] Agranovskiy A. V., Balakin A. V., Gribunin V. G., Sapozhnikov S. A.: Steganografiya, tsifrovyye vodyanyye znaki i steganoanaliz. Vuzovskaya kniga, Moscow 2009.

[2] Ali A. E.: A New Text Steganography Method by Using Non-Printing Unicode Characters. Eng& Tech. Journal 28 (1), 2010, 72–83.

[3] Chekhovskyy S.: Sovremennye metody skrytoy peredachy ynformatsyy putem prohrammnoho upravlenyya yzluchenyem komp′yutery. Pravove, normatyvne ta metrolohichne zabezpechennya systemy zakhystu informatsiyi v Ukrayini. 2003.

[4] Echo Data Hiding (html): http://www.slidefinder.net/a/audio_steganography_echo_data_hiding/24367218.

[5] Ferreira A. M.: An Overview on Hiding and Detecting Stego-data in Video Streams. University of Amsterdam. System & Network Engineering – Research Project II, 2015.

[6] Fridrich J., Goljan M., Soukal D.: Searching for the Stego Key. Proc. SPIE 5306, 2004 [doi: 10.1117/12.521353].

[7] Genkin D., Shamir A., Tromer E.: RSA key extraction via low-bandwidth acoustic cryptoanalysis. Tel Aviv University. 2013.

[8] Konakhovich G. F., Puzyrenko A. Yu.: Computer steganography. Theory and practice with Mathcad. MK-Press, Kyiv 2006.

[9] Kuhn M. G., Anderson R.: Soft Tempest: Hidden data transmission using electromagnetic emanations. University of Cambridge, Computer Laboratory, New Museum Site, 1998.

[10] League C.: An overview of digital steganography, particularly within images, for the computationally curious. Long Island University 2015: https://www.youtube.com/watch?v=-7FBPgQDX5o.

[11] Romanova A., Toliupa S.: Perspective steganographic solutions and their application. Proceedings of the VII Inter University Conference Engineer of XXI Century, volume 2, 2017, 269–278.

[12] Secret Code in Color Printers Lets Government Track You; Tiny Dots Show Where and When You Made Your Print. Electronic Frontier Foundation October 2005: https://www.eff.org/press/archives/2005/10/16.

[13] Zorin Ye.Ye., Chichvarin N.V.: Steganografiya v SAPR. Uchebnoye posobiye. MGTU im. N.E. Baumana. Moscow.

Anna Romanova
e-mail: anitraromanova@gmail.com

Since 2015, has been a student of the Department of Cybersecurity and Information Protection. Has published articles in European scientific journals, as well as took part in conferences, devoted to scientific and applied topics. Academic interests include steganography, TEMPEST, system security, and critical information infrastructure protection. Is a certifies SearchInform DLP-System Specialist.

ORCID ID: 0000-0003-1403-6322

Prof. Toliupa Sergiy
e-mail: tolupa@i.ua

Scientific and practical interests are related to such areas as intelligent control systems, the direction of improving the efficiency of information technology, information security systems, cybersecurity and cyber defense. He is the author of 5 monographs and over 150 scientific and methodological works, 16 textbooks and manuals.

ORCID ID: 0000-0002-1919-9174

otrzymano/received: 1.10.2018 *przyjęto do druku/accepted: 15.12.2018*