Information security vulnerabilities using steganography as the art of hiding information

Javier Guaña-Moya[1][0000-0003-4296-0299], Yolanda Borja López[2][0000-0001-5349-8461], Gonzalo Gutiérrez Constante[3][0000-0003-1484-8558], Paulina Jaramillo Flores[4][0000-0001-7965-9868], Oswaldo Basurto Guerrero[5][0000-0002-4118-872X]

[1] Pontificia Universidad Católica del Ecuador, Ecuador, eguana953@puce.edu.ec
[2] Universidad Central del Ecuador, Ecuador, yaborja@uce.edu.ec
[3] Universidad Central del Ecuador, Ecuador, gfgutierrez@uce.edu.ec
[4] Instituto Tecnológico Superior Japón, Ecuador, pjaramillo@itsjapon.edu.ec
[5] Instituto Tecnológico Universitario Internacional, Ecuador, oswaldo.basurto@iti.edu.ec

Abstract. Due to advances in technological tools, most of the information is kept electronically and because of this information security has become a fundamental issue. Steganography is the art of hiding information and an effort to hide the existence of the embedded information, it serves as a better way to protect the message than cryptography, which only hides the content of the message, not the existence of the message, allowing the original message is hidden within a digital medium that serves as transport, so that the changes that occur in it are not detectable. Consequently, steganography is a useful tool that allows the covert transmission of information through the communication channel, by combining the hidden message with the original message, originating the hidden image, which is difficult to detect without recovering it. The purpose of this bibliographical review is to know steganography as a method to hide information and to determine the various modes of application in the digital area to hide data and information from unauthorized users, thus achieving that the exchange of information is a discreet process and insurance for those interested.

Keywords: Digital archive, hidden file, cryptography, steganography.

1. Introduction

With the growing technological dependence of people and organizations, data represents the most valuable asset to be taken care of, as it is vital to ensure the success of an organization or for the correct use of an individual's personal information. For this reason, there is a growing concern for information security, considering the following aspects to be fundamental in any situation: guaranteeing the availability of resources and information, the integrity and confidentiality of information [1].

All information resources, such as databases, servers, workstations, applications and similar, must be properly registered and protected to ensure the quality of service, considering that this data integrity aims to ensure that the information is always available in the latest version, so the tools must control and monitor versions and changes, thus ensuring the veracity of the information for the organization [1].

The most important term in this work is the confidentiality of information which implies that it should only be available to those who are entitled to it, minimizing attacks, information leaks, access to unnecessary information, thus ensuring that the company's strategic information is secure, which implies that information security must adopt customized physical, technological and human controls, which allow reducing and managing risks, leading the company to achieve the appropriate level of security for the business [2],[3].

In accordance with the above, there are the information hiding techniques that aims to hide messages, in such a way that other people cannot detect it, providing this technique a powerful tool for the secure transmission of information and, fundamentally, it is divided into two parts: cryptography and steganography; however, the difference between them lies in the rationale, since encryption makes it clear that there is data and that it is encrypted, while steganography aims to hide this information, embedding the message in another type of media, making people think that there is nothing hidden [4],[5].

The purpose of this bibliographic review is to learn about steganography as a method for hiding information and to determine the different ways of application in the digital area to hide data and information from unauthorized users, thus making the exchange of information a discreet and secure process for the interested parties.

2. Methodology

The present study was formulated based on the systematic literature review guidelines established by Kitchenham (2004) [6], with the objective of obtaining information related to the research questions posed for the development of the study.

This standard establishes the following stages:

- Planning the review.
- Conducting the review.
- Analysis of results.

2.1. Planning the review

The objective of the research is to learn about steganography as a method for hiding information and to determine the various ways of application in the digital area to hide data and information from unauthorized users.

For the development of the topic, the following research questions were posed:

Q1: What is steganography?

Q2: What are the application criteria of steganography?

Q3: What are the main methods of application of digital steganography?

The information provided by digital databases, such as ACM Digital Library, IEEE eXplorer, Science Direct Elsevier, Scopus, Google Scholar and Springer Link, was accessed, referring to topics related to steganography, definition, origin of the technique, application models and modes of application in computer science, identifying among the sources of information found academic journals and technical publications, between 2010 and 2022, with a search strategy based on aspects related to the research questions. The inclusion criteria considered in the selection of documents are: articles that develop the topic of steganography, conceptualization and application of the technique in the digital field in any of its modalities; while as exclusion criteria, information published on general websites, documents with irrelevant contributions and blogs were considered as exclusion criteria.

1.1 Conducting the review

In this phase, the articles were selected based on the inclusion and exclusion criteria, reviewing the titles of the articles, content and conclusions in order to determine the contribution to the questions posed.

As a result of the search, 43 documents were identified, of which 15 were selected that met the established criteria.

1.2 Analysis of results

Understanding what is meant by steganography was possible by obtaining answers to Q1: What is steganography?

The art of camouflaging the presence of hidden messages in legitimate carriers, known as steganography, has in recent years become a commercial tool for malware vendors, as evidenced by recent attacks against major global targets. Although steganography has been known for centuries, it has recently proliferated in new terrains: digital media, computer networks and popular telecommunications services. There is currently no silver bullet for steganography abuse, other than meticulously searching for any loopholes that can be exploited for the purpose of embedding illicit information, or any means to alter the potential carrier in a way that escapes human perception [7].

According to the above, steganography can be defined as the art of communicating in secret, hiding messages within other unimportant information, so that there is no way to detect that there is a hidden message, specifically in the computer area, this information can be a sound, image or text file [8].

Therefore, steganography is the set of techniques that allow us to hide any type of information. However, it is important not to confuse it with cryptography, due to the fact that the latter modifies data to make it incomprehensible, while steganography simply hides it among other data. Despite the different approach of each technique, it is very common to combine the two to achieve better results [9].

Q2: What are the criteria for the application of steganography?

The reasons for the use of steganography can be very varied, but they can appear because there is no support for encrypting data or because there is an authority that does not allow the passage of certain information. Thus, the information travels in the files without anyone knowing what it really carries inside, finding that one of the applications that is currently arousing more interest is the application of watermarking, which involves a copyright notice or trademark hidden in images, music or commercial software [10],[11].

On the other hand, from the point of view of some authors some of the techniques for hiding viruses in files can be considered strictly speaking as steganographic techniques for hiding the malicious code from the user or even from antivirus programs, which indicates that the arrival and evolution of computer science has made it possible to achieve great advances in the possibilities of steganography, and also to automate tasks that used to be quite costly in terms of time and money. Moreover, the type of information that can now be concealed is not limited to written messages; it is now possible to digitally conceal text, images, sounds and even executable programs, as in the case of computer viruses [12].

Regarding the objectives of hiding information, Kumar et al. (2019) [13] express that they can change in subtle ways, however, they can be classified according to the following criteria:

- Robustness: guarantees that the secret information cannot be destroyed without seriously degrading the visible message, allowing quantifies the resistance of the hidden message to the various attacks or transformations made to the steganographic medium.
- Invisibility: Aims to ensure that the affected medium is not disturbed by the inserted secret information.
- Insertion capacity of a steganography system: Defined by the size in bits of the secret message that can be embedded in a medium of a given size, the relative insertion capacity being the ratio between the size of the secret message to be hidden and the size of the medium used. Therefore, the capacity defines the amount of information that can be embedded in the medium without obvious deterioration.

These three characteristics are closely related and are inverse, because the improvement of the capacity generally has a negative influence on invisibility [14].

On the other hand, it is important to mention that the steganography technique should be based on the following basic principles: firstly, to select very well the medium in which it is to be applied, meaning that the covered file, although it loses quality, is not perceptible and, secondly, to take advantage of the limitations of man in terms of perception, such as the range of colors, which although they vary a little the human eye cannot perceive, there are also frequencies that the ear cannot decode.

It is important to mention that there is several software specialized in hiding information through steganography. These programs are used to hide messages or files inside other files, such as images, audio or even documents. Some of the most popular software are:

- Steghide: It is a command line tool that allows you to hide data in image and audio files. It is widely used for its simplicity and effectiveness.
- OpenStego: It is an open-source tool that provides a graphical user interface for hiding data in images. It is easy to use and versatile.
- OutGuess: It is another command line steganography program used to hide data in images. Provides advanced options to customize hiding.
- Steganos Suite: It is a software suite that includes steganography and encryption tools. Provides a comprehensive solution for data security.

- Invisible Secrets: It is a software that combines steganography with encryption to protect files and messages effectively.
- Hide in Picture: This software focuses on hiding data in images and offers an intuitive interface for non-technical users.
- QuickStego: This is a simple option for hiding data in images and is suitable for basic steganography tasks.

Steganography is used for both legitimate and malicious purposes, so its use must be ethical and legal. Furthermore, the choice of software will depend on the specific needs and the level of security required.

Q3: What are the main application methods of digital steganography?

The development of steganography has a wide variety of methods, among the most used we can mention the "least significant bit" or LSB (Least Significant Bit), based on the use of the least significant digit in order to hide the message; there is also the statistical method, which uses the most redundant values of the file to locate there the bits that refer to the message to be hidden, considered by many specialists as one of the most powerful and secure methods [15].

Steganography can be classified into two groups, the first group corresponds to digital steganography that allows hiding a message in another with the help of a computer and the second group is linguistic steganography, which uses natural language to send hidden messages, the latter is based on the ability of people to understand words, symbols, humor and ambiguities that are characteristic of human understanding, an equivalence that does not exist in computers. Although Artificial Intelligence has had a great development in recent times, computers still do not have the ability to recognize characters in an image, for example, the captcha, which is an authentication method to verify that who registers is a human being and not a robot [16].

In relation to digital steganography, hidden information can be inserted in various media, such as images, audios or videos.

Image steganography

The information is hidden inside the cover image according to some algorithm, this image is known as stego image. Often, to increase the security of the embedding, a secret key or special algorithm parameters known to the sender and receiver of the information are used and since steganographic embedding algorithms ensure the invisibility of the embedded data and hide the fact of the presence of such data, the transmission of the steganographic image is usually carried out through an open communication channel [17].

In this case the receiver uses the information extraction algorithm and extracts the embedded message from the stego image, information that can be extracted in the original form and may contain distortions associated with the characteristics of the embeddable procedure or with any destructive effects that arise when transmitting a stego image through a communication channel. The original image can also be restored by the receiver, for this to happen the embedding algorithm must have the property of reversibility, hence the ability to fully restore the original image after extracting the embedded information [17].

On the other hand, embedding additional information in a digital image leads to distortion, consequently, steganographic embedding stealth is characterized, first of all, by the absence of visible distortions of stego-images. Moreover, an additional requirement for steganographic algorithms is robustness: resistance to various destructive effects, which allows it to extract embedded information with minimal distortion even when exposed to any distorting effects on stego-images [17].

Audio steganography

Sound files also allow hiding information, considering that, unlike image files, the human being can only listen to a frequency range; being the .WAV format the most used and recommended for sound storage, because it has two parts, the header containing file information, such as the number of channels, sampling frequency and sample size; and the data sector which is the sound itself, in the form of sequential bits [18].

They point out Tan et al. (2019) [19] that audio steganography methods can be divided into three categories: time domain, frequency domain, and wavelet domain.

Time domain: Most time domain methods employ lowbit coding techniques, including LSB coding, echo hiding and parity coding, finding that the LSB method is one of the earliest used for information hiding, which replaces the least significant part of the binary sequence of each digital audio sample with the binary sequence of the equivalent secret message, so it has a high embedding capacity. For example, when using an embedding deck at a sampling rate of 8 kHz, the capacity is 8 kbps. In general, the length of the secret message to be encoded is less than the total number of samples in the sound file. At the same time, this method is easy to implement and can be combined with other concealment techniques [20].

In the echo concealment technique, the secret data is a short echo embedded in the original signal, the essence of the echo being a resonance added to the original audio. To avoid the influence of echoes, stego signals must maintain the same statistical and perceptual characteristics as the original audio after adding echoes.

The method changes three parameters of the echo signal: decay rate, offset and initial amplitude, so that the echo is inaudible, thus making it possible to make the echo indistinguishable from the original signal by controlling the delay. In addition, the amplitude and decay rates can be set below the inaudibility threshold, so that the data can be masked without being perceived [21].

The parity coding method decomposes the original signal into different sample regions and hides the secret message in the parity bits of the sample region. If the secret message matches the verification bit, it continues, otherwise, invert the LSB bit of a sample in this region. With parity coding, the sender has more options for the modified secret bits [22].

Frequency domain: The human auditory system has a masking effect that makes weak frequencies close to strong resonance frequencies inaudible. Many transform domain methods use this feature. The main methods are tone insertion, spread spectrum and phase coding.

The tone insertion technique embeds data by inserting inaudible tones into the original audio. To embed a bit in a speech frame, a pair of tones is generated at two frequencies and it is determined whether the embedded bit is 0 or 1 by comparing the power ratio of the inserted tones [20].

Spread spectrum technology is similar to the system implemented by LSB coding, which randomly distributes message bits in audio files; it spreads the secret information about the frequency spectrum of the audio file using a code that is independent of the actual signal. Therefore, the final signal occupies more bandwidth than is actually required, which will also generate random noise in the audio and cause data loss [23].

The phase encoding method is based on the fact that the human auditory system cannot recognize the phase shift in the signal simply as the noise in the signal, so it can encode the secret message bit in the phase shift in the corresponding spectrum of the digital signal by changing the phase of the original audio segment. It uses the discrete Fourier transform (DFT), which is a transform algorithm for the audio signal. At the receiver, we must know the length of the segment before we can use the DFT to obtain the phase to extract the secret information [23].

Wavelet domain: This method is based on the discrete wavelet transform, where the signal can be decomposed into high and low frequency parts, and the low frequency part can be decomposed again into high and low frequency parts, the number of signal decomposition being determined mainly by the application and duration of the original signal. By combining with wavelet energy, masking effect, adaptive and LSB it is possible to embed secret information in discrete wavelet coefficients [24].

Video steganography

It is a technique that embeds messages in cover contents and is used in many fields, such as medical systems, law enforcement, copyright protection and access control. Since the human visual system is less sensitive to small changes in digital media, especially for digital video, video steganography is a technique that hides the message in a video and conceals the fact of transmission. And it has become more popular recently due to two main reasons: along with the rapid development of computer applications, the problem of information security is becoming more and more serious. Video is an electronic media that may be more eligible than other multimedia due to the rise of powerful digital video content sharing/transmission tools and its size. With the rapid advancement of the Internet and multimedia technologies, digital videos have become a popular field for data hiding because infinite video streams provide a large amount of redundancy space for embedding secret messages in video steganography [25].

Existing video steganography algorithms can be divided into three categories according to the embedding position: pre-embedding, when the message embedding position is the raw video domain; internal embedding, referring to the message embedding position being the compressed domain; post embedding when the message embedding position is the bitstream domain. For all categories, the performances of the algorithm evaluation criteria should always be considered: imperceptibility, robustness, embeddability and algorithm complexity, fundamentally [25].

Finally, Channalli & Jadhav (2009) [26] point out that for this system to be considered robust it must have the following properties:

- The quality of the media should not be noticeably degraded by adding secret data.
- The secret data must be undetectable without secret knowledge, usually the key.
- If multiple data are present, they must not interfere with each other.
- Secret data must survive attacks that do not degrade the perceived quality of the work.

3. Conclusions

This paper presents a general description of steganography, including definitions and basic principles, as well as the various ways of being applied through digital media, under the premise that steganography transmits secrets through apparently innocuous covers in an effort to hide the existence of a secret, so that steganography in digital files in any of its modalities are in permanent use and application.

As technology advances, cutting-edge developments are emerging that allow the creation of novel methods to hide and discover data and even to completely rethink the way in which steganography is used. Equally important are the ethical concerns of its use, considering that software can easily transmit private user information without the user's permission or knowledge.

Steganography, despite its legitimate uses, can be an attack vector, although steganography is a powerful technique for protecting information, it can also be used maliciously by cybercriminals to hide malware or exfiltrate information, posing a significant risk to information security.

Steganographic techniques can easily evade conventional security measures, as the hidden information often does not appear to be malicious. This poses significant challenges for detecting and preventing information security threats, making detection of what is being shared a challenge for many people attempting to decrypt this type of communication.

Steganography can be used to facilitate covert communications, allowing attackers to send instructions to malware or exfiltrate data undetected, adding an additional layer of complexity to cyber defense.

Steganographic techniques can make digital forensics more challenging by making it more difficult to trace the source of an attack and recover stolen information. This is an additional dimension of the vulnerability presented by steganography, so mitigating these vulnerabilities requires a defense-in-depth approach, combining multiple security measures, including steganography detection, intrusion detection systems, firewalls, antivirus, and appropriate information security policies and procedures. In addition, it is essential to provide training and awareness to employees and users about the potential risks of steganography to minimize the likelihood of a security breach.

4. Acknowledgement

We wish to express our sincere gratitude to the Pontificia Universidad Católica del Ecuador for their support of our research, which was conducted as part of the activities of the Applied Information Systems and Technologies research group (SITECIA), registered under GI-Quito-071-2022. Furthermore, we extend our heartfelt appreciation to the Anonymous Reviewers whose thoughtful and constructive recommendations greatly enhanced the quality of this paper. Their expertise and feedback played a pivotal role in refining our research and ensuring its rigor and credibility.

Reference

1. Guaña-Moya, J., Chiluisa-Chiluisa, M. A., del Carmen Jaramillo-Flores, P., Naranjo-Villota, D., Mora-Zambrano, E. R., & Larrea-Torres, L. G. (2022, June). Ataques de phishing y cómo prevenirlos Phishing attacks and how to prevent them. In 2022 17th Iberian Conference on Information Systems and Technologies (CISTI) (pp. 1-6). IEEE.
2. Azevedo, E., Faveri, J. G., & Nunes, S. E. (2015). Esteganografia. Revista de Ciências Exatas e Tecnologia, 10(10), Article 10. https://doi.org/10.17921/1890-1793.2015v10n10p%p
3. Warner, M. (2019). Wanted: A definition of 'intelligence'. In Secret Intelligence (pp. 4-12). Routledge.
4. Carvalho, D. (2017). Esteganografia Digital para transmissão oculta de mensagens. 1(1). https://silo.tips/queue/esteganografia-digital-para-transmissao-oculta-de-mensagens-diego-fiori-de-carva?&queue_id=-1&v=1662746592&u=MTkxLjk5LjE0MS4xNjY=
5. Renner, R., & Wolf, R. (2023). Quantum advantage in cryptography. AIAA Journal, 61(5), 1895-1910.
6. Kitchenham, B. (2004). Procedures for Performing Systematic Reviews. Keele, UK, Keele Univ., 33. https://www.researchgate.net/publication/228756057_Procedures_for_Performing_Systematic_Reviews
7. Zielińska, E., Mazurczyk, W., & Szczypiorski, K. (2014). Trends in steganography. Communications of the ACM, 57(3), 86-95. https://doi.org/10.1145/2566590.2566610
8. Lucas, G. (2017). Esteganografia. http://app.uff.br/riuff/handle/1/5663
9. Dalal, M., & Juneja, M. (2021). A survey on information hiding using video steganography. Artificial Intelligence Review, 1-65.
10. Ruiz, H., Chaumont, M., Yedroudj, M., Amara, A. O., Comby, F., & Subsol, G. (2021). Analysis of the scalability of a deep-learning network for steganography "into the wild". In Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10–15, 2021, Proceedings, Part VI (pp. 439-452). Springer International Publishing.
11. Guaña-Moya, J., Sánchez-Zumba, A., Chérrez-Vintimilla, P., Chulde-Obando, L., Jaramillo-Flores, P., & Pillajo-Rea, C. (2022). Ataques informáticos más comunes en el mundo digitalizado. Revista Ibérica de Sistemas e Tecnologias de Informação, (E54), 87-100.
12. Mezquida, D. A., Alcojor, M., Acero, I., & Sanz, L. F. (2003). Ocultación de imágenes mediante Esteganografía. Novática: Revista de la Asociación de Técnicos de Informática, 163, 52-57.
13. Kumar, S., Singh, A., & Kumar, M. (2019). Information hiding with adaptive steganography based on novel fuzzy edge identification. Defence Technology, 15(2), 162-169. https://doi.org/10.1016/j.dt.2018.08.003
14. Khaldi, A. (2019). Steganographic Techniques Classification According to Image Format. International Annals of Science, 8, 143-149. https://doi.org/10.21467/ias.8.1.143-149
15. Angulo, C. A., Ocampo, S. M., & Blandon, L. H. (2007). Una mirada a la esteganografía. Scientia et Technica, 5(37), 421-426.
16. Sánchez, J. (2017). Esteganografía, Disciplina para ocultar información. Biblioteca Digital UBA, 44.
17. Evsutin, O., Melman, A., & Meshcheryakov, R. (2020). Digital Steganography and Watermarking for Digital Images: A Review of Current Research Directions. IEEE Access, 8, 166589-166611. https://doi.org/10.1109/ACCESS.2020.3022779
18. Kuchkorov, M. A. (2022). Analysis of sound signals in wav format in telecommunication systems and algorithms for them processing. Central asian journal of mathematical theory and computer sciences, 3(12), 27-33.

19. Tan, D., Lu, Y., Yan, X., & Wang, X. (2019). A Simple Review of Audio Steganography. 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 1409-1413. https://doi.org/10.1109/ITNEC.2019.8729476
20. Djebbar, F., Ayad, B., Meraim, K. A., & Hamam, H. (2012). Comparative study of digital audio steganography techniques. EURASIP Journal on Audio, Speech, and Music Processing, 2012(1), 25. https://doi.org/10.1186/1687-4722-2012-25
21. Tekeli, K., & Aşlıyan, R. (2017). A COMPARISON OF ECHO HIDING METHODS. 1, 397-403.
22. Malviya, S., Saxena, M., & Khare, A. (2012). Audio Steganography by Different Methods. International Journal of Emerging Technology and Advanced Engineering, 2(7), 371-375.
23. Alwahbani, S., & Elshoush, H. (2018). Hybrid Audio Steganography and Cryptography Method Based on High Least Significant Bit (LSB) Layers and One-Time Pad—A Novel Approach (pp. 431-453). https://doi.org/10.1007/978-3-319-69266-1_21
24. Shivaram, H., Acharya, D., Adige, R., Deepthi, S., & Upadhya, K. (2015). Audio steganography in discrete wavelet transform domain. International Journal of Applied Engineering Research, 10, 37544-37549.
25. Liu, Y., Liu, S., Wang, Y., Zhao, H., & Liu, S. (2019). Video steganography: A review. Neurocomputing, 335, 238-250. https://doi.org/10.1016/j.neucom.2018.09.091
26. Channalli, S., & Jadhav, A. (2009). Steganography An Art of Hiding Data (arXiv:0912.2319). arXiv. https://doi.org/10.48550/arXiv.0912.2319