

Features of using Invisible Signs in the Word Environment for Hiding Data

Zaynalov N.R., Narzullaev U.Kh., Muhamadiev A.N., Bekmurodov U.B., Mavlonov O.N

Abstract: Considered digital steganography is the direction of classical steganography based on concealing or introducing additional information into digital objects, while causing some distortions of these objects. At the same time, images, audio, video, network packets, etc. can be used as objects or containers. Recently, there has been a lot of publication in the field of information hiding in a text container. To embed a secret, steganographic methods rely on redundant information about the used covering media or properties that the human perception system cannot distinguish. Since text documents are widely used in organizations, using a text document as a storage medium may be the preferred choice in such an environment. On the other hand, the choice of using a text document as a storage medium is the most difficult, since it contains less redundant information. In this article, we present textual steganography using invisible characters in a word processor.

Keyword : Cryptography, Information hiding, Steganography, Text steganography

I. INTRODUCTION

The task of protecting information from unauthorized access was solved at all times throughout the history of mankind. Already in the ancient world, there were two main directions for solving this problem, which still exist today: cryptography and steganography.

Steganography is a field of knowledge that deals with hidden information transfer. Steganography (from the Greek steganos (secret, secret) and graphy (record)) literally means "secret writing". Unlike cryptography, the fact of the transfer of information is hidden. Especially effective is the use of steganographic methods in conjunction with cryptographic. A common feature of steganographic methods and algorithms is that the hidden message is embedded in some innocuous, non-eye-catching object that is transported to the addressee openly. When using cryptography, the presence of an encrypted message in itself attracts the attention of the attacker; in the case of steganography, the presence of hidden information remains imperceptible. Plain text where the information encrypted by the steganographic algorithm is hidden is called a container.

The development of computer technology in the last decade has given new impetus to the development of computer steganography. Messages are now embedded in

digital data, usually of analog nature - speech, audio recordings, video, graphic images, and even text files and executable program files.

Steganography is the art and science of hiding a message inside another message without causing suspicion in others, so a message can only be detected by its intended recipient [1], since encrypted messages can attract the attention of attackers [2]. But when steganography is used, even if the eavesdropper receives a stego object, he cannot suspect the message, since it is carried out covertly [3,4].

Modern steganography is usually understood as electronic media, and not physical objects and texts. In steganography, the text we want to hide is called embedded data, and the text, image, audio, or video file that is used as a carrier to hide text is called a container. Container - information intended to conceal secret messages. Empty container - a container without an embedded message; filled container or stego - a container containing embedded information.

Inline (hidden) message - a message embedded in a container.

In parallel with steganography, cryptography also developed, and here I would like to dwell on the contribution of Oriental scientists to these sciences.

Because, in the Islamic world, advances in science in the Middle Ages could not bypass cryptography. One of the oldest works in this field is Abu Bakr Ahmad ibn Ali ibn Wahshiya Al-Nabati (derived from the names of scholars [5]), which deals with the problems of reading ancient manuscripts in the 855s. worked out. This book also explores two alphabetical cryptographic techniques, including various encryption systems that have been in use since the 19th century.

The work on frequency cryptography was created in 855 and belongs to Abu Yusuf al-Kindi, which explores the problems of decrypting cryptographic messages.

Al-Kindi's cryptography manuscript has reached us today, and it also provides encryption methods based on the ancient alphabet. This algorithm was used until the 19th century.

The information collected in this field is compiled in the 14-volume work of Shauba al-Asha, written in 1412 by Egyptian mathematician Shahob Kalkashandi. It addresses seven decryption and decipherment issues. Here are some ways to read unknown encrypted messages based on Arabic-language statistics and linguistic regularities [5]. One of the great features of this encyclopedia is the use of frequency analysis in crypto analysis for the first time. This method reads encrypted text

Revised Manuscript Received on July 22, 2019.

Zaynalov N.R., Narzullaev U.Kh., Muhamadiev A.N., Bekmurodov U.B., Mavlonov O.N

Samarkand branch of the Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Uzbekistan,

using simple replacement methods. The role of the word cipher in science comes from these works.

In the eighth century, Khalil al-Farahidi drew attention to the decryption of text based on keywords in the classification of secret letters. For example, we start the letter with the word "hello", so it is possible to identify the five letters in the text. Khalil al-Farahidi has published this idea in his book *Al-Maumma*.

That is why Shahab Kalkashandi was recognized as the founder of cryptographic analysis, but there is no mention in the national literature.

Let us return to the classical methods of steganography, which can be classified as follows:

- Hiding the text file of the container in the inter-format spaces is the simplest of the following ways to hide the message file. Most often, the necessary information is entered into empty or initially unreadable areas of the container file. Most often the message is written to the end of the file or between its blocks. It is also possible to use "end of line" and "carriage transfer". These methods are the easiest to implement, but also the most vulnerable.

- Concealment concealments use service areas and special blocks of the container file directly. The basic principle of this approach is to "issue" a message file for all sorts of service information of a container file. There are many ways to create fake areas or data. The most popular for a large number of different formats are the following: hiding in the fields the specifications of the container file, hiding in the fields reserved for expansion, hiding using the properties that are not displayed fields of the container file.

Thus, in textual steganography, symbolic text is used to hide secret information. Storing text files requires less memory, and its easier communication makes it preferable to other types of steganographic methods. Because texts take up less memory, transfer more information and require less printing, as well as some other benefits.

This paper presents a method for hiding data using non-displayable characters in Word.

The rest of the paper is organized as follows: Section 2 describes some of the existing approaches to textual steganography. Section 3 describes the proposed approach. Section 4 compares with other methods. Section 5 concludes and discusses the advantages and disadvantages of steganography.

II. EXISTING APPROACHES

In this section, we present some of the popular text steganography approaches based on non-displayable characters. At the same time, some approaches to classifications of steganography methods are also given in [6, 7].

In paper [7], which can be considered one of the classic works in this area, depending on the type of embedding technique, textual steganography is divided into three categories: 1) embedding at the character level, 2) at the bit level, and 3) mixed type. All these categories and the corresponding subcategories are discussed in detail and

examples are given. Here we consider only those methods that use invisible signs.

2.1 White Steg

There are a lot of methods based on non-displayable characters (for example, spaces, paragraph marks, tabs, etc.). The classic approach here is to change the distance between words using spaces. For example, one space after a word represents bit 0, and two spaces after a word represents bit 1 [4, 8]. As a result of further research in this method, the secret message began to be hidden due to the vertical displacement of text lines to some extent [6, 9]. The marked line has two unmarked control lines, one on each side of it, to determine the direction of movement of the marked line [10]. To hide bit 0, the line shifts up, and to hide bit 1, the line shifts down [11]. In [12], a modification of the well-known textual steganography method, based on the change in the interline distance of an electronic document, is described. With its help, it is proposed to hide a secret message in changing the height of line spacing. The essence of the method modification is to use an electronic document as a container and to change the line spacing not for the entire line or paragraph, but only for the non-displayed characters.

Other methods also use spaces to hide the secret message in XML and HTML files, where bit 0 is represented by the absence of a space in the tag, and bit 1 is represented by inserting a space inside the tag [11]. And in [13], the space after the semicolon is embedded in bit 0, and the tab after the semicolon - bit 1.

III. PROPOSED APPROACH

Once in England, this method was used: under some letters on the front page of the newspaper there were tiny dots, almost invisible to the naked eye. If you read only the marked letters, you get a secret message! Using this idea, a new way of embedding a message in a text container using invisible characters is proposed. In this case, the secret text is written only in lowercase letters. The beginning and end of the secret message is marked in uppercase letters. Thus, each letter of a secret message is marked in turn by an invisible sign in the original text using the technology of creating an index. As an example, consider an excerpt from a gazelle of one of the greatest lyricists of the world literature, Hofiz Sheroziy:

*That beautiful Shirazi Turk, took control and my heart stole,
I'll give Samarkand & Bukhara, for her Hindu beauty mole.*

Introduce the message "I am a girl" into these lines using the technology of the index and if you display all the signs, you will get:

That beautif^uful Shirazi Turk, took control a^u and
m^uly hea^urt stole,
I'll g^uve Samarkand & Bukhara, for^u her
Hindu beauty mol^ue.

In general, this method, which can be referred to as the "Pointer" method, is promising from the point of view of stealth in a visual way.

IV. ASSESSMENT OF RESULTS

Comparison of existing methods is given in more detail in [7]. Methods are compared based on their ability to embed. The amount of embedding is the amount of information that can be hidden in the selected container environment. In the above example, only 103 characters. The embedded message text consists of 8 letters (without spaces). Only for this example, we were able to embed 0.8 bits per substrate character, which is comparable to other methods from [7], where the Character and String Mapping method turned out to be the most acceptable. True, you need to have a solid base of source code for the greatest implementation of the message in a text container.

V. CONCLUSIONS

Thus, digital steganography, which is inspired by the ancient secret methods of communication, is the art of hiding the secret message within the cover environment in an inconspicuous manner. In connection with the recent development of digital communications, steganography has gained a new paradigm using digital media such as text, image, audio, video, etc. Although other types of media can be used as a covering tool other than text, many organizations prefer text documents.

Until recently, the so-called "prisoners problem" proposed by 1983 by Simmons [14] was used to describe the model of the steganographic system. It consists in the fact that two individuals (Alice and Bob) want to exchange secret messages without the intervention of a security guard (Willie), who controls the communication channel. Moreover, there are a number of assumptions that make this problem more or less solvable. The first assumption facilitates the solution of the problem and consists in the fact that participants in the information exchange can share a secret message (for example, using a code key) before concluding. Another assumption, on the contrary, makes it difficult to solve the problem, since the guard has the right not only to read the messages, but also to modify (change) them. Violation of the first admission can lead to disastrous results, i.e. if the secret message is sent in clear text. As an "example," we will cite the tale "Ali Baba and the Forty Thieves" from the book "A Thousand and One Nights", where Ali Baba accidentally overheard a secret message and gained access to wealth. Namely, once, gathering wood in the forest, Ali Baba accidentally witnesses a conversation of forty thieves. It turns out the entrance to the cave, where the treasures they stole are stored, opens with the help of the magic words "Simsim, open." Learning this secret, Ali Baba, after leaving the bandits, penetrates into the cave and takes with them a bag of gold coins. True, it should be noted that Ali Baba was very "lucky", since the keywords turned out to be symmetrical, because Ali Baba could not hear the words of

the robbers when they left the cave! And the robbers were "unlucky," and the reason was that the key words were transferred not covertly.

Thus, steganographic methods based on the peculiarities of presenting information in computer files gives us the opportunity to talk about the development of a new direction - computer steganography, which will allow them to be applied in areas such as copyright protection, prevention of electronic document falsification and other applications.

REFERENCES

1. Changder S, Ghosh D, Debnath N.C. (2010). Linguistic approach for text steganography through Indian text. In: 2010 2nd international conference on computer technology and development, pp. 318-322
2. Ross J.Anderson, Fabien A.P. Petitcolas. (1998). On the limits of steganography. IEEE J Sel Areas Commun 16(4):474-481
3. Fabien A.P. Petitcolas, Ross J.Anderson, Markus G.Kuhn. (1999). Information hiding—a survey. In: Proc IEEE 87(7):1062-1078
4. Por LY, Ang TF, Delina B (2008) WhiteSteg-a new scheme in information hiding using text steganography. WSEAS Trans Comput 7(6):735-745
5. Wikipedia, The Free Encyclopedia / <http://wikipedia.org/>
6. M. Hassan Shirali-Shahreza, Mohammad Shirali-Shahreza (2006) A new approach to persian/arabic text steganography. In: 5th IEEE/ACIS international conference on computer and information science and 1st IEEE/ACIS international workshop on component-based software engineering, software architecture and reuse, pp 310-315
7. R. Bala Krishnan, Prasanth Kumar Thandra, M. Sai Baba. An overview of text steganography. 4th International Conference on Signal Processing, Communications and Networking (ICSCN -2017), March 16 - 18, 2017, Chennai, INDIA
8. Bender W, Gruhl D, Morimoto N, Lu A (1996) Techniques for data hiding. IBM Syst J 3(3&4):313-336
9. M. Hassan Shirali-Shahreza, Mohammad Shirali-Shahreza (2008) A new dynonym text steganography. In: International conference on intelligent information hiding and multimedia signal processing, pp 1524-1526
10. Brassil JT, Low SH, Maxemchuk NF, O’Gorman L (1995) Document marking and identification using both line and word shifting. In: Proceedings of INFOCOM '95 proceedings of the fourteenth annual joint conference of the IEEE computer and communication societies, pp 853-860
11. Cummins J, Diskin P, Lau S, Parlett R (2004) Steganography and digital watermarking. School of Computer Science, pp 1-24
12. E. A. Blinova. Steganographic method based on the line-shift coding method on non-displayed symbols of the electronic text document. Belarusian State Technological University. BGTU. № 6 2016, p.166-169
13. Kabetta H, Dwiandiyanta BY (2011) Suyoto: information hiding in CSS: a secure scheme text steganography using public key cryptosystem. Int J Cryptogr Inf Secur 1(1):13-22
14. Simmons G.J. The prisoner's problem and the subliminal channel, Proc. Workshop on Communications Security (Crypto'83), 1984, 51-67.