CLASSIFICATION AND WAYS OF DEVELOPMENT OF TEXT STEGANOGRAPHY METHODS

Article · October 2019

CITATIONS
0

READS
21

4 authors, including:

Zaynalov Nodir
Tashkent University of Information Technology
5 PUBLICATIONS **4** CITATIONS
SEE PROFILE

Abdinabi Muhamadiev
Tashkent University of Information Technology
8 PUBLICATIONS **7** CITATIONS
SEE PROFILE

Qilichev Dusmurod
Tashkent University of Information Technology
5 PUBLICATIONS **2** CITATIONS
SEE PROFILE

Impact Factor:					
ISRA (India)	= 4.971	SIS (USA)	= 0.912	ICV (Poland)	= 6.630
ISI (Dubai, UAE)	= 0.829	РИНЦ (Russia)	= 0.126	PIF (India)	= 1.940
GIF (Australia)	= 0.564	ESJI (KZ)	= 8.716	IBI (India)	= 4.260
JIF	= 1.500	SJIF (Morocco)	= 5.667	OAJI (USA)	= 0.350

SOI: 1.1/TAS DOI: 10.15863/TAS

International Scientific Journal

Theoretical & Applied Science

p-ISSN: 2308-4944 (print) **e-ISSN:** 2409-0085 (online)

Year: 2019 **Issue:** 10 **Volume:** 78

Published: 16.10.2019 **http://T-Science.org**

QR – Issue QR – Article

N.R. Zaynalov
Samarkand branch of Tashkent university of information technologies named after Muhammed al-Khwarizmi, Researcher,
nodirz@mail.ru

Sh.A. Aliev
Samarkand branch of Tashkent university of information technologies named after Muhammed al-Khwarizmi, Researcher,

A.N. Muhamadiev
Samarkand branch of Tashkent university of information technologies named after Muhammed al-Khwarizmi, Researcher,

D. Qilichev
Samarkand branch of Tashkent university of information technologies named after Muhammed al-Khwarizmi, Researcher,

I.R. Rahmatullaev
Samarkand branch of Tashkent university of information technologies named after Muhammed al-Khwarizmi, Researcher,
nabi8888@bk.ru

CLASSIFICATION AND WAYS OF DEVELOPMENT OF TEXT STEGANOGRAPHY METHODS

Abstract: *Steganography is based on hiding or embedding additional information in digital objects, while causing some distortion of these objects. In this case, as objects or container can be used image, audio, video, network packets, etc. recently, there is a lot of progress in the field of hiding information in a text container. To embed classified information, steganographic techniques rely on redundant information about the covering medium used or properties that a person cannot visually distinguish. Because text documents are widely used in organizations, using a text document as a storage medium may be a preferred choice in such an environment. On the other hand, choosing to use a text document as a storage medium is the most difficult because it contains less redundant information. In this article, we present a different approach of classifications of existing methods within text stegnography, which allow young scientists to quickly enter the topic for further development of this field of science.*

Key words: *Information hiding, Steganography, text steganography, methods, text documents.*

Language: *English*

Citation: Zaynalov, N. R., Aliev, S. A., Muhamadiev, A. N., Qilichev, D., & Rahmatullaev, I. R. (2019). Classification and ways of development of text steganography methods. *ISJ Theoretical & Applied Science, 10 (78),* 228-232.

Soi: http://s-o-i.org/1.1/TAS-10-78-42 ***Doi***: https://dx.doi.org/10.15863/TAS

Scopus ASCC: *1700.*

Introduction

The task of protecting information from unauthorized access has been solved at all times throughout the history of mankind. Already in the ancient world, there were two main directions for solving this problem, existing to this day: cryptography and steganography.

Steganography is a field of knowledge that deals with the hidden transmission of information. Steganography (from the Greek steganos (secret, mystery) and graphy (writing)) literally means "secret writing" or "covered writing" [1,2]. Unlike cryptography, the very fact of information transfer is hidden. Especially effective is the use of steganographic methods in conjunction with cryptographic. A common feature of steganographic methods and algorithms is that a hidden message is embedded in some harmless, non-attention-grabbing object, which is then openly transported to the recipient. When using cryptography, the presence of an encrypted message itself attracts the attention of the attacker, in the case of steganography, the presence of hidden information remains invisible.

The development of computer technology in the last decade has given a new impetus to the development of computer steganography. Messages are now embedded in digital data, usually having an analog nature-speech, audio, video, graphics, and even text files and executable program files.

Steganography is the art and science of hiding a message inside another message without arousing suspicion in others, so that a message can only be detected by its intended recipient. [3] Since encrypted messages can attract the attention of intruders [4]. But when steganography is used, even if the eavesdropper receives a stego object, he cannot suspect the message because it is carried out covertly. Modern steganography is usually understood as electronic media rather than physical objects and texts. In steganography, the text we want to hide is called inline data, and the text, image, audio, or video file that is used as a carrier to hide the text is called a container. Container-information designed to hide secret messages. This definition implies any type of media. Empty container - a container without built-in messages; the filled container or stego - the container that contains the inline information. Embedded (hidden) message - a message embedded in a container.

Classical methods of steganography can be classified as follows:

• Hiding a container file in the inter-format spaces is the easiest way to hide a message file. Most often, the necessary information is entered in 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. These methods are the easiest to implement, but also the most vulnerable. This method leads to an increase in the size of the container file, which makes it the most suspicious.

• Hiding-masking uses directly service areas and special blocks of the container file. The basic principle of this approach is to "give" the message file for all sorts of service information of the container file. There are quite a few ways to create fake areas or data. The most popular for a large number of different formats can be considered the following: hiding in the specification fields of the container file, hiding in the fields reserved for the extension, hiding using properties that are not displayed fields of the container file.

In a special group, you can also select methods that use special properties of file presentation formats:

• fields of computer file formats reserved for expansion, which are usually filled with zeros and are not considered by the program;

• special data formatting (offset words, sentences, paragraphs, or the choice of certain positions of the letters);

• remove identifying headers for a file.

As a rule, such methods are characterized by a low degree of stealth, low bandwidth and poor performance.

In text steganography, symbolic text is used to hide classified information. Storing text files requires less memory, and its easier communication makes it preferable over other types of steganographic methods. Because texts take up less memory, transmit more information and require less printing costs, as well as some other benefits.

This article presents a new approach to the systematization of textual steganography methods by combining existing methods into groups, with further discussion of some of them.

The rest of the paper is organized as follows: Section 2 describes some of the existing approaches to classifying text steganography. Section 3 describes the proposed classification approach. Section 4 provides an assessment of the proposed classification method. In section 5, we digress and discuss the advantages and disadvantages of steganography and draw appropriate conclusions.

II.EXISTING APPROACHES

It should be noted that the popularity of text steganography methods has led to a variety of approaches that need to be somehow systematized. Research in this direction are found in the scientific literature, so the classification of steganography methods is given in [5, 6]. It is necessary to understand that classification is a system of distribution of objects (objects, phenomena, processes, concepts) by classes in accordance with certain characteristics.

In the work [6], which can be considered one of the classic works in this area, where the volume of implemented information is chosen as a sign. Thus, depending on the type of embedding technique, text steganography is divided into three categories: 1) embedding at the character level, i.e. by-character, 2) at the bit level, i.e. by-bit, and 3) mixed type. In [7] all these categories and their corresponding subcategories are discussed in detail and examples where the classification is made on the basis of the concept of the essence of the method are given.

III. THE PROPOSED APPROACH

Classification is one of the powerful tools in understanding processes, phenomena and objects. As indicated in section 2, classification can be carried out depending on the type of category. Here we take as a category that each object has certain properties. For example, if you take a file as an object, then the essence of the file can be such properties as: textual or binary. In the proposed approach, it is recommended to classify by such properties as: file Structure or format, text data Format and Linguistics (see Fig.1.). Let's consider these concepts in more detail.

A file format is a specification of the structure of data recorded in a computer file. The file format identifier is usually specified at the end of the file name as an "extension". The file name extension helps identify the format of the data contained in the file to programs that can work with it. Sometimes the data format is additionally specified at the beginning of the file content. For example, the " doc " format is a Word 97-2003 document, with the file format being a binary file.At the same time, the "docx" format is also a Word document, but the file format is XML-based files, which by default is inherent in Word 2007 and above.

Thus, Word document files are complex objects organized according to the rules of structured storage. In fact, structured storage is a separate "file system" from Microsoft, roughly the same as FAT or NTFS.

This technology is related to Excel 2007 and later, where the XML format is used to create workbooks, templates, and add-ins. In fact, these files are ZIP archives. If necessary, they can be unzipped and viewed.

So the XML file format belongs to the category of so-called open formats and such files can be created and processed using any text editors that are not related to Office.

Thus, taking into account these features, this article proposes a new classification of text steganography methods and gives a brief overview of some of these methods. This allows to compare modern methods of text steganography and identify trends in the development of this area, as well as the choice of the best algorithms for use in research and commercial tasks.

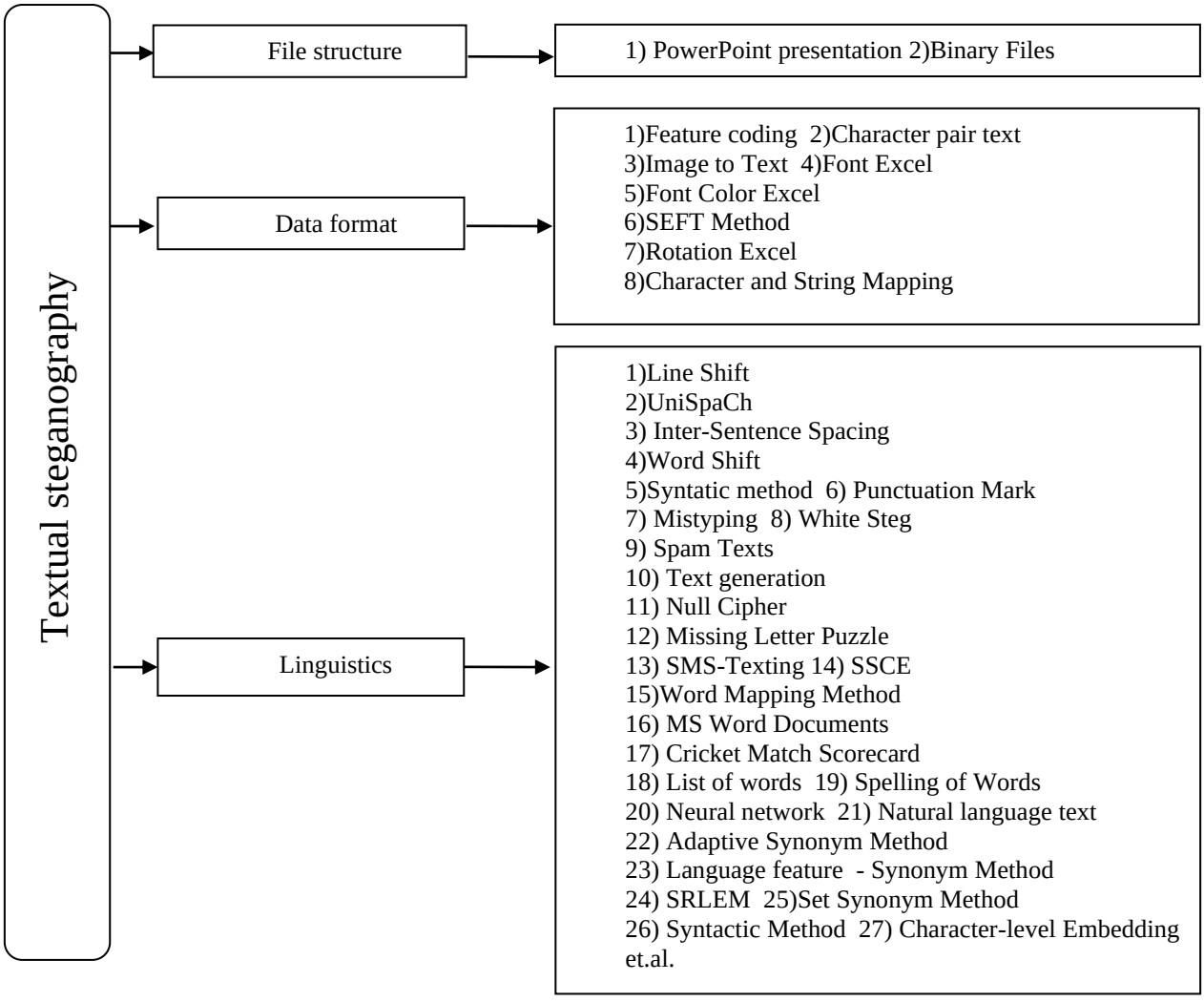

Fig.1. Classification of the steganography

Consider the features of some methods of text steganography shown in Fig.1, proposed by the authors of this work, without in-depth detail of the methods themselves.

3.1 PowerPoint Presentation

This method is based on a PowerPoint file that has a branched structure. The authors of the proposed method, considering the PowerPoint file as a container, found a sufficiently large space, the so-called free zones, where you can embed a large amount of information. Because, physically, a PowerPoint file consists of a header followed by an ordered list of containers and atoms. Once in these containers, you can find enough space to embed sensitive data [8].

3.2 Feature coding

This approach uses character properties and by making specific changes to the fonts of individual letters, you can embed information.

For example, in [9] we propose a modification of the method of spatial-geometric (i.e. changing the style or font size, bold, italic, underline, etc.) and color parameters of text symbols. Here, data is hidden not only in normal, but also in special (soft hyphenation, line break, etc.) characters and spaces. The peculiarity of the method based on color modification is that the processes of inclusion/extraction of information are carried out by comparative change/analysis of color parameters of pairs of neighboring symbols, one of which is the basic one. In [6] these methods are described as Character Marking or, a variation of this method, such as Character and String Mapping, described in [6], uses a font attribute called character spacing in the accompanying document to embed the secret.

3.3 Line Shift

The classic approach here is to change the distance between words using spaces. As a result of further research in this method, the secret message began to be hidden due to the vertical displacement of the text lines to some extent [5, 10]. The marked line has two unmarked control lines, one on each side of it, to determine the direction of the marked line. To hide bit 0, the line is shifted up, and to hide bit 1, the line is shifted down. A modification of the known method of text steganography based on changing the line spacing of an electronic document (line-shift coding) is described in [11]. With its help it is offered to hide the secret message in change of height of line intervals. The essence of the modification of the method is to use an electronic document as a container and change the line spacing not of the entire line or paragraph, but only of non-displayed characters (spaces, paragraph marks, tabs, etc.). And you can also use the Inter-Sentence Spacing method [6], based on the fact that a space is inserted between two sentences, where the presence of an additional space represents "1", and their absence - "0". Because the number of sentences in a plain text document is not considered to be large, this method cannot be used to hide more data.

Similarly, the UniSpaCh (Unicode Space Characters) method uses a non-visible space character, but of the Unicode variety [6].

IV.EVALUATION

The proposed classification of steganography methods show that methods based on the use of special properties of data formats have a stable form of the method. While, methods where linguistic features of spelling of different languages are used, have a wide range of forms of methods. At the same time, very little attention is paid to the methods where the file structure is used. A simple analysis of the structure of a Microsoft Word file, at first glance nothing nevidelyayuschiysya object, but nevertheless the analysis shows that how diverse the structure of this type of file. Accordingly, there are different approaches to hiding information in these objects.

The above classification of steganography methods does not take into account cloud technologies, although the research in this area by some scientists [12] deserves the attention of scientists in the field of steganography. The prospects for the use of cloud technologies in the field of text steganography requires a separate study.

V. CONCLUSION

So digital steganography, which is inspired by ancient secret communication techniques , is the art of hiding a secret message inside a covert medium in an undetectable way. Due to the rapid development of digital communications, steganography has received a new paradigm with the help of digital media such as text, image, audio, video, etc. Although other types of media other than text can be used as a covering medium, but many organizations prefer text documents, so this area deserves the attention of scientists.

Thus, the steganographic methods, which are based on the features of the presentation of information in computer files gives us the opportunity to talk about the rapid development of a new direction-computer steganography. Since, with the development of information technology there are a variety of opportunities to hide classified information.

References:

1. Petitcolas, F.A.P., Anderson, R.J., & Kuhn, M.G. (1999). Information hiding-a survey. *In: Proc IEEE 87(7)*, pp.1062-1078.
2. Por, L.Y., Ang, T.F., & Delina, B. (2008). WhiteSteg-a new scheme in information hiding using text steganography. *WSEAS Trans Comput 7(6)*: pp.735-745
3. 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
4. Anderson, R.J., & Petitcolas, F.A.P. (1998). On the limits of steganography. *IEEE J Sel Areas Commun 16(4)*, pp.474-481
5. Hassan Shirali-Shahreza, M., & Shirali-Shahreza, M. (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
6. Krishnan, R. B., Thandra, P.K., Sai Baba, M. (2017). An overview of text steganography. 4th International Conference on Signal Processing, Communications and Networking (ICSCN - 2017), March 16 - 18, 2017, Chennai, INDIA.
7. Zaynalov, N.R., Narzullaev, U.Kh., Muhamadiev, A.N., Bekmurodov, U.B., & Mavlonov, O.N. (2019). Features of using Invisible Signs in the Word Environment for Hiding Data. 2019. *International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8, Issue-9S3, July 2019.* pp.1377-1379. https://www.ijitee.org/wp-content/uploads/papers/v8i9S3/I32950789S319.pdf
8. Tiwari, R.K., & Sahoo, G. (2011). Microsoft power point files: A secure steganographic carrier. *International Journal of Digital Crime and Forensics. 3(4)*, pp. 16-28.
9. Shut'ko, N. P. (2016). The algorithms of realization of text steganography methods based on the modification of the geometric and color text parameters. *Belarusian State Technological University. BGTU. № 6 2016*, pp.160-165.
10. Shirali-Shahreza, M.H., & Shirali-Shahreza, M. (2008). A new dynonym text steganography. In: International conference on intelligent information hiding and multimedia signal processing, pp. 1524-1526.
11. Blinova, E.A. (2016). 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*, pp.166-169.
12. Islomov, S.Z., Mavlonov, O.N., Muhamadiev, A.N., Shodmonov, D.A., & Djumaev, S.N. (2018). New authentication scheme for cloud computing. *Journal of Advanced Research in Dynamical and Control Systems. 10(10)*, pp. 2316-2319. http://jardcs.org/backissues/abstract.php?archiveid=6795