International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 11, pp 648-653, November-2023

https://doi.org/10.47001/IRJIET/2023.711085

Text Steganography Techniques: A Review

[1]Alaa Abdullah Idres, [2]Yaseen Hikmat Ismael

[1]Diploma Student, Department of Computer Science, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq

[2]Lecturer, Department of Computer Science, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq

Authors E-mail: [1]alaa.22csp37@student.uomosul.edu.iq, [2]yaseen-hikmat@uomosul.edu.iq

***Abstract* - The large-scale information interchange that occurs across computer networks in the modern day makes it imperative to protect that information and stop unauthorized individuals from accessing or altering it. Information hiding is one of the critical areas for maintaining information security, preventing unauthorized individuals from accessing and knowing its content. The steganography technique employs various types of media as covers for embedding secret messages. These media can be in the form of images, audio, video, text, and other forms of media. The process of embedding information within text presents several challenges, the most significant of which include the limited available space for concealment and the substantial impact of the hidden data on the text. This research includes a review of the most important techniques and studies used in text-based steganography, presenting their advantages and weaknesses.**

Keywords: Steganography, information hiding, Text Confidentiality.

I. INTRODUTION

Due to the suspicion raised by various encryption methods used in data protection, particularly when sensitive and important data is involved, information steganography has emerged as a fundamental and essential alternative for preserving data security. The process of data steganography involves using a specific medium as a cover to embed secret information. There are various types of media (such as images, audio, video, text, and more). Because of the limited space available for embedding secret data when using text as a cover, as well as the significant sensitivity of text to changes, text-based data steganography has garnered significant attention among researchers.

Many studies and techniques have emerged in this field. Text-based data steganography technology can be divided into three main types: Format-based, Random and Statistical Generation, and Linguistic Methods [1].

II. TEXT STEGANOGRAPHY TECHNIQUES

Text steganography techniques are used to explain the techniques and methods used in hiding within the text, as well as identifying the most important advantages of these techniques [2,3].

2.1 Format based method

The present method includes the use of methods based on actually changing the text format to hide data such as changing font sizes, or adding white distances between words. One of the disadvantages of this method is if the STEGO File is opened using the text processor, the spelling errors and white distances (the spaces added in the concealment process) will be recognized, in addition to that if a comparison between the original text (before concealment) and the text after hiding are made the places of change will be recognized and thus the possibility of discovering the hidden text [2,3].

2.1.1 Line Shift

Line Shift is identified that the distances between the lines of the text are fixed. In this method, for the purpose of hiding the text, the text is initially converted into the values of the ASCII coding and then converting these symbols into the binary formula. For the purpose of hiding the bit (O), the line is removed to the top with a fixed value (the amount of a fixed displacement), while the concealment is the value of (1) the line is removed down and with the same amount of displacement upwards. One of the disadvantages of this method is that if the text (edited) is rewritten using a different text processor, or using the letters of the letters (OCR), hidden information will be destroyed [2]. Line shift can be clarified in Figure (1).

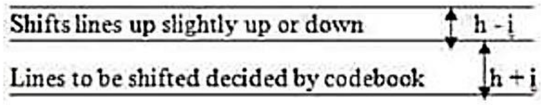
Shifts lines up slightly up or down
h - i
Lines to be shifted decided by codebook
h + i

Figure 1: Line shift text steganography

ISSN (online): 2581-3048
Volume 7, Issue 11, pp 648-653, November-2023
https://doi.org/10.47001/IRJIET/2023.711085

2.1.2 Word Shift

The technique for word shifting relies on altering the ordinary text's word spacing. Throughout the text, there are constant word spacing. The word is moved to the right by a predetermined amount to conceal the value (0), and to the left by a predetermined amount to conceal the value (1), as agreed upon by the two communication parties . While this method is thought to be simple for masking, one drawback is that it is simple to identify the masking process by the use of an optical character recognition application, a different text editor, or by comparing the original text with the masking text[2].Word Shift method can be clarified (2).

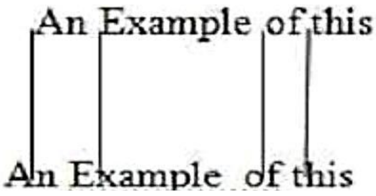

Figure 2: Word shift text steganography

2.1.3 Feature Coding

The method of features coding is used to perform the concealment inside the text by changing some of the features of the text letters. For example, to hide the value (0) the last letter of the word (extends its end) is extended, and to hide the value (1) the end of the last letter of the word is shortened. One of the advantages of this method is that a large hidden space can be provided, but its faults, as is the case in all the methods of Format Bused Method, can be discovered the concealment process in the same previous methods [4].

2.2 Random and Statistical Generation

It is used when a random text is generated by relying on statistical characteristics to include the secret text (the concealment process). One of the methods used in this method is to hide the data in a random display sequence for texts, or use statistical features of the word length and frequency of letters to produce words that have statistical properties similar to actual words in a specific language. One of the most important features of Random Statistical Generation is the lack of original text (before concealment), which is compared by the attacker to discover the concealment [5].

2.2.1 Word Mapping

This technique converts the secret message to be hidden using the Crossover process, after which the resulting encoded text appears in the time in the cover. The concealment process depends on the length of the floor in the cover file is it my husband or individual. The hiding sites map is sent in a separate file from the concealment file to the receiving side [6].

2.2.2 MS Word Document

This method uses the different formatting features of the Microsoft Word, where the text appears after the concealment process as if it is a result of writing a person with a few skills using the Microsoft Word program. The use of various formats to carry out the concealment process depends on the agreement between the two sides of the communication [6].

2.3 Linguistic Methods

The linguistic method depends on the linguistic characteristics of the text that has been created and changed by doing the concealment. For example, (CFG Grammar (CFG Grammar), which is a group of rules that produce a possible trees or analysis and thus rely on it to discover the hidden text. Where the left branch is determined by (0) and the right branch by (1). Greibach normal form (GNF) can also be used to generate a tree brown to achieve the concealment. One of the disadvantages of this method is that the rules of small grammatical lead to the repetition of a text clip several times in addition to that although the text is not tainted with grammatical flaws, there is in the semantic structure where the text is a series of sentences that do not have a link to each other [7].

2.3.1 Syntactic Method

The Syntactic Method of the Text depends on the Use of the Point (.), The Comma (,), The Decisive Comma (;), and so on to Hide the Values of Bits (1,0). The Problem of this Method is that it requires Determining the Right Places to Include Numbering Marks. As the Wrong Use and Punctuation Marks in Incorrect Places LEAD to the Ease of Detective the Concept [8].

2.3.2 Semantic Method

The semantic method is used as the linguistic synonym of certain words and thus the concealment process in the locations of these synonyms. One of the disadvantages of this method is the limited concealment process, as well as this concealment process may change completely the meaning of the text [8].

III. RELATED WORK

Chaw A. has presented a new way to hide inside the text. He depended on the use of the semantic method, in replacement words of the cover of the cover with the words synonymous with it. After converting the text to be hidden to the ASCII system and then to the binary system. In instances where synonyms are used, the value (Zero, One) is concealed. The suggested concealment technique was applied by the researcher using the system for exchanging banking information between banks and clients as a model [9].

Al-Nofaie, S. et al., recommended a way to conceal the texts written in Arabic or similar languages such as Persian and Urdu. For the purpose of performing a safe concealment, the researchers used the fake distance (PS-BetWord) and (Kashida-PS) by integrating them together for efficient formula. The proposed method provided a large hiding space [10].

Alyousuf, F., Roshidi, D. indicated the presentation of a set of concealment methods in the text and the measures of efficiency used in its evaluation, as it depended on the characteristics (Feature-Based Steganography) is the most used in most of the methods It is used to measure the efficiency of the methods of concealment in the text was according to the following percentages: security 34%, capacity 24%, durability 23%, and time included 19% [11].

Mustafa N. studied a new way to hide the text using invisible letters and compare the text and with the hidden text. This research included the generation of a secret message in four main stages such as using the letter from the original message, choosing the appropriate cover text, dividing the text into blocks, and concealing the secret text. One of the advantages of this method is high secret production due to multiple levels of complexity to avoid the attacker [12].

Alanazi, N. et al., stated a proposed method using (Unicode), in addition to using invisible and visual letters such as Kashida, ZWJs, ZWNJ, also using MMSPS to increase the ability of this method without reducing data safe. This research included the use of the method of contextual forms of Arabic letters to conceal in text. One of the advantages of this method is characterized by a high percentage of safety in relation to other methods [13].

Sadie J. et al., presented a proposal on the basis of the color coding to conceal the text using the first two ways: depends on the substitution (Permutations) and the second: dependent on the numability system. One of the advantages of these two methods provides a better concealment and high capacity in terms of ease of reading [14].

Khekan, A. et al., dealt with a method using the points of Arabic letters and used the Arabic Semantic Disionary to hide many secret texts. Part of the Arabic language features were used to include the secret English message in the text cover to create the information of the information, where the secret text is converted to the bilateral system and then apply the T-5Be algorithm on it to reduce the size of the secret text by 37%. One of the advantages of this research is a high hidden resolution and cover storage capacity [15].

Akbar, F. introduced a new method using (Word-Shift) technology, where the research included providing one of the important factors, which is the level of safety in hiding information, as the message was converted into a series of including bits in the document file by converting distances between words [16].

Thabit, R. et al., presented the color method and the coordinated voids (CSNTSTAGE) to solve the few capacity problems to hide and not reveal the hidden text. The proposed method included the first two phases: the Huffman method is used to reduce the size of the secret text to be hidden and work to increase the number of bits that can be hidden in each location or letter of the original text. The second stage: depends on the coloring of the color and the distance in the original text to solve the problem of the color difference between the colors of the cover and the resulting text after the concealment process [17].

Figueira, J. stated a way to hide information in the text using Winstein's Ideal Coding which has the highest percentage to hide information, where all possible synonyms can be used for words. Also use Huffman coding to reduce the resulting text cover. One of the advantages of this method is one of the advantages of this method is a high hiding rate [18].

Adeeb, O. and Kabudian, S. handled a way to hide secret data using artificial intelligence and long -term long -term memory (LSTM) to increase a capacity of 45%. This research included the use of the Arabic language due to the large number of words, vocabulary and linguistic meanings in it. Where the research relied on the previous Arabic poetry texts, where the linguistic accuracy was increased within the poetry formula with the use of an algorithm (Baudot code), where the secret data was hidden at the level of letters instead of words [19].

Abdul Majeed, M. et al., presented a way to hide the information that combines encryption and pressing using multi -layer FPE coding and coding to reduce the size of secret data before hiding it. This research ensures the use of invisible unicode letters to include secret data in text files in English to produce STEGO files. One of the advantages of this method is a significant improvement in hiding information and lack of doubt about the cover file [20].

Osman, B. et al., offered a method on the basis of color (RGB), red, green, blue, and the second is the theory of dividing the rest (SQRT) to hide in the text, where (RGB) was used on a scale of (0,0,0) to (15,15 , 15) To avoid suspicion of color, in addition to using the generation of random numbers

(PRNG) to make secret messages dynamic with the creation of a table (homophonic) where a quantity of information is hidden by 77.4%. This research guarantees to hide the information by changing the features of the letters: (its size, shape, style). One of the advantages of this method is a high secret ability [21].

Shakir, N. and Mahd, M. used a new, unaccounted method to improve the ability to include and increase the effect by collecting the grammar of the Arabic language and marks of formation (movements) with the Unicode method and discrimination in the use of concealment of information on the basis of using special letters (A – O- D – TH- R - Z) in the Arabic language in the Holy Quran. One of the advantages of this method is a high ability to hide and also high secrecy [22].

Khosravi, B. indicated a new way to hide in the text that depends on that there are different models (RGB, HSL, HSV) are used to represent color values, and that the very little difference between two different colors but close to the RGB system as the same coding in the system (HSL) The researcher relied on this method to do the concealment. The abundance of this method is a large treasury space for the invisible concealment process [23].

Following a study of the most significant research projects in the subject of information hiding, Table (1) provides an overview of those projects by outlining the main idea of the approach taken and whether or not other methods, like encryption or compression, were employed to aid in the concealing process. A description of the metrics or techniques that the researchers suggested using to gauge the effectiveness of their suggested approaches is also included in the table.

Researchers [No]	Years	Methods	Compression	Encryption	Criteria
Aye aye Chaw [24]	2019	Synonym substitution based algorithm	-	-	Achieved Capacity
Safia – AL-Nofai [25]	2019	Utilizing pseudo-spaces with kashida	-	-	High capacity High security
Farah Qasim Ahmed [26]	2020	Feature-based and word – role based	-	-	Security 34% Capacity 24% Robustness 23%
Nada Abdulaziz [27]	2020	Invisible Character (White space)	-	-	High secret due use multi-level of complexity
Norah Al-Anazi [28]	2020	Unicode standard with Arabic text (ZWJ , ZWNJ) (MMSPS) (Kashida)	-	-	High security capacity 50%
Juveik, Sadie [29]	2020	Color coding (based on permutation , numeration systems)	-	-	High Capacity
Ahlam R. Khekau [30]	2020	Use data of characters , use Arabic semantic dictionary	(T-5BE)	-	High masking Accuracy and Storage Capacity
Fitrachainl Akbar	2020	Word – shift technique	-	-	-
João Figueira [31]	2022	Weinstein’s ideal coding (14) marks or chain	(Huffman)	-	Highest rate of hidden information
REEMA Thabit [32]	2022	Color or spacing normalization stego	(Huffman)	-	High Capacity 98.85 Improves robustness 94% 94.22%
OMER FAROOQ [33]	2022	Artificial intelligence deep learning and (LSTM) (baudot code)	-	-	Security capacity 45%
Mohammed Abdul Majeed[34]	2022	Unicode Characters	(Huffman)	(FPE) Encryption Format Preserving	High Capacity Maintaining Security
Baharudin Osman [35]	2023	(RGB) color technique (SQRT) Second Quotient Reminder Theorem (PRNG)	-	-	Capacity 77.4%
Nooraldeen Subhi Shakir [36]	2023	Arabic Grammar Diacritics with Unicode	-	-	High Capacity
Bahman Khosravi [37]	2023	(RGB, HSL, HSV)	-	-	High Capacity

ISSN (online): 2581-3048
Volume 7, Issue 11, pp 648-653, November-2023
https://doi.org/10.47001/IRJIET/2023.711085

IV. CONCLUSION

A number of challenges arise when using text files as a cover for information concealment. The most significant ones are the restricted amount of space available for the concealment process, the high sensitivity of any modifications made to the text file, and the possibility that using a different texts editor will cause the concealment process to fail. Text files are no longer used as a means of information concealment as a result of all these limitations. In an effort to get around these challenges, the researchers introduced a hybrid masking method, which leverages encryption and compression techniques to help improve the masking process in text files.

REFERENCES

[1] Sun Bowen and et. al., 2023, "Topic Controlled Steganography via Graph-to-Text Generation", DOI: 10.32604/cmes.2023.025082, pages No.:157-176.

[2] Akbar Chairil Fitra, 2020, "A Study of Text Steganography Methods", pages No.: 369-372.

[3] Hamdan M. Abdullah and Hamarsheh Ala, 2017, "AH4S: an algorithm of text in text steganography using the structure of omega network", DOI: 10.1002/sec.1752, pages No.: 6004-6016.

[4] Mandal Kumar Kumar and Singh Kumar Pradeep, 2019, "Information Hiding in Text Steganography: A Different Approach", pages No.: 593-596.

[5] Khan Yahya and et. at., 2021, "Disbursal of Text Steganography in the Space of Double-Secure Algorithm", https://doi.org/10.1155/2021/7336474, pages No.: 1-9.

[6] Agarwal Monika, 2013, "TEXT STEGANOGRAPHIC APPROACHES: A COMPARISON", DOI: 10.5121/ijnsa.2013.5107, pages No: 91-106.

[7] Abdul Majeed Mohammed and et. al., 2021, "A Review on Text Steganography Techniques", https:// doi.org/10.3390/math9212829, pages No: 1-28.

[8] Ahvanooey Taleby Milad and et. al., 2019, "Modern Text Hiding, Text Steganalysis, and Applications: A Comparative Analysis", doi:10.3390/e21040355, pages No.: 1-29.

[9] Chaw Aye Aye, 2019, "Text steganography in Letter of Credit (LC) using synonym substitution-based algorithm", International Journal of Advance Research and Development, Volume 4, Issue8, pages No.:59-63.

[10] Al-Nofaie Safia and et.al., 2019, "Enhancing Arabic text steganography for personal usage utilizing pseudo-spaces", Journal of King Saud University – Computer and Information Sciences, https://doi.org/10.1016/j. jksuci. 2019.06.010, pages No.:963-974.

[11] Alyousuf Ahmed Qasim Farah and Roshidi Din, 2020," Analysis review on feature-based and word-rule based techniques in text steganography", Bulletin of Electrical Engineering and Informatics, DOI: 10.11591/eei. v9i2.2069, pages No.: 764 – 700.

[12] Mustafa Abdul Aziz Nada,2020, "Text hiding in text using invisible character", International Journal of Electrical and Computer Engineering (IJECE), DOI: 10.11591/ ijece. V10i4. Pp 3550-3557.

[13] Alanazi Norah and et. al., 2020," Inclusion of Unicode Standard seamless characters to expand Arabic text steganography for secure individual uses", Journal of King Saud University – Computer and Information Sciences, https://doi.org/10.1016/j.jksuci.2020.04.011, pp. 1343–1356.

[14] Sadié K. Juvet and et. al., 2020," Two high-capacity text steganography schemes based on color coding".

[15] Khekan R. Ahlam and et. al., 2020, "New text steganography method using the Arabic letters dots", Indonesian Journal of Electrical Engineering and Computer Science, DOI:10.11591/ijeecs. v21.i3. pp1784-1793, pp. 1784~1793.

[16] Akbar Chairil Fitra and et. al., 2020, "Steganography on Text using Word-Shift Coding and Centroid Methods", Journal of Engineering and Applied Sciences, Page No.: 3095-3100.

[17] THABIT REEMA and et. al., 2022, "CSNTSteg: Color Spacing Normalization Text Steganography Model to Improve Capacity and Invisibility of Hidden Data", Digital Object Identifier 10.1109/ACCESS.2022.3182712, pages No.: 65439 – 65458.

[18] Figueira Joao, 2022, "A Survey on Semantic Steganography Systems", arXiv:2203.12425v1, pages No.:1 -7.

[19] ADEEB AHMED AHMED OMER and KABUDIAN JAHANSHA HSEYED, 2022, "Arabic Text Steganography Based on Deep Learning Methods", Digital Object Identifier 10.1109/ACCESS.2022.3201019, pages No.: 94403 – 94416.

[20] Abdul Majeed Mohammed and et.al., 2022," New Text Steganography Technique based on Multilayer Encoding with Format-Preserving Encryption and Huffman Coding", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 13, No. 12, pages No.:163 – 172.

[21] Osman Baharudin and et. al., 2023, "TEXT STEGANOGRAPHY USING THE SECOND QUOTIENT REMAINDER THEOREM AND DARK COLOUR SCHEMES", Journal of Computational

Innovation and Analytics, https://doi.org/10.32890/ jcia2023.2.1.2, pages 21- 40.

[22] Shakir Nooruldeen, 2023, "USING SPECIAL LETTERS AND DIACRITICS IN STEGANOGRAPHY IN HOLY QURAN", Iraqi Journal for Computers and Informatics, Vol. 49, Issue 2, pages No.:1 – 8.

[23] Khosravi Baman, 2023, "Text steganography by changing the black color", Doi: 10.22060/AJMC.2023.21801.1111.

Citation of this Article:

Alaa Abdullah Idres, Yaseen Hikmat Ismael, "Text Steganography Techniques: A Review" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 7, Issue 11, pp 648-653, November 2023. Article DOI https://doi.org/10.47001/IRJIET/2023.711085

****