

A Survey Paper on Text Steganography

Namrata Singh

P.G Student

Department Of Computer Science & Engineering

ABES Engineering College, Ghaziabad, U.P, India

Corresponding Author: nam2817120@gmail.com

Abstract

Steganography is the art of concealing the way that correspondence is occurring, by concealing data in other data . In this paper, we examine different types of text - Steganography procedures. There are distinctive techniques for text Steganography. We describe it one by one for concealing secret data in text. There exists a huge variety of text Steganography strategies, some are more complex than others and every one of them have individual strong and weak point. Different applications have distinctive necessities of the text Steganography technique used. We describe a typical application that needs both total imperceptibility and extensive secret information get covered up. We analyse some fundamental strategy for text based - Steganography technique.

Keywords: *Text Steganography, Cryptography*

I. INTRODUCTION

Steganography is the art and science of composing shrouded or hidden messages in such a way that nobody, aside from the sender and expected beneficiary, suspects the presence of the message, a type of security without knowledge of its presence [1]. Our real need is covering up. The word

Steganography is of Greek origin and signifies "concealed writing" from the Greek words steganos signifying "covered or protected", and graphein which means "to compose or write". Initially it recorded the utilization of the term was in 1499 by Johannes Trithemius in his Steganographia, a treatise on cryptography and

Steganography disguised as a book on magic camouflaged as a book on enchantment. For the most part, messages will give off an impression of being something else: images, articles, shopping records, or some other cover content and, traditionally, the hidden message might be in undetectable ink between the noticeable lines of a private letter.

Steganography don't baits the spy as it conceals the message. Steganography can be classified based on the type of media it utilize to hide the text[1]. These are as follows:

A) Text Steganography

It hides the text behind some other text file. It is the a difficult type of steganography as the redundant amount of content to shroud or hide the secret message is rare in text files.

B)Image Steganography

It is a standout amongst the most usually used tecnique because of the limitation of the Human visual Framework(HVS). Human eye can't distinguish the inconceivable scope of colors and an insignificant change in the nature of a image that outcomes from steganography.

C) Audio steganography

It is likewise a difficult type of steganography as people can recognize a moment change in the nature of audio.

II. NEED OF STEGANOGRAPHY

In late couple of years, our web is growing rapidly day by day. One major area which picked up attracted by individuals is security subjected to internet, subjected to correspondence. Now a day for security reason concealing information is most well known techniques which get more consideration than cryptography. This is because encryption - decryption give careful consideration than concealing information strategy.

For establishing hidden correspondence, different techniques including cryptography, coding, - Steganography, and so on are utilized [10]. Steganography is the way toward concealing information inside other information in such a way, to the point that no other individual recipient, even know the presence of it. This is significant advantage of stenography over other coding methods. A large portion of the - Steganography technique takes a shot at image, sound, music, video clips, text, and so on [11].

There are four fundamental type of Steganography

1. Text -Steganography
2. Image -Steganography
3. Audio -Steganography
4. Video -Steganography

We describe distinctive text - Steganography strategies and compare it. text Steganography is most difficult sort of - Steganography ; this is because of the absence of redundant information in a text file, while there is a great deal of redundancy in a image or a sound file, which can be used in Steganography [4].

III. DIFFERENT TECHNIQUES OF TEXT -STEGANOGRAPHY

Text steganography can be broadly classified into three types:

1. Format based Random

Format based techniques include modifying physically the configuration or format of text to hide the data. This technique has certain blemishes. In the event that the stego file is opened with a word processor, incorrect spellings what's more, additional white space will get distinguished. Changed textual styles sizes can stimulate doubt to a

human peruser. Furthermore, if the first plaintext is accessible, contrasting this plaintext and the suspected steganographic text would make controlled parts of the text very noticeable [13].

2. Statistical generation

Keeping in mind the end goal to maintain a strategic distance from examination with a known plaintext, steganographers frequently turn to creating their own cover writings [13]. One technique is covering data in arbitrary looking succession of characters. In another strategy, the measurable properties of word length and letter frequencies are utilized as a part of request to make words which will seem to have same measurable properties as real words in the given language [10, 11].

3. Linguistic methods

Linguistic steganography particularly considers the linguistic properties of created and altered content, and by and large, utilizes semantic structure as the space in which messages are concealed [13]. CFG make tree structure which can be utilized for hiding the bits where left branch represents "0" and right branch relates to '1'. A sentence structure in GNF can likewise be utilized where the primary decision in a

generation represents to bit 0 and the second decision speaks to bit 1. This technique has a few downsides. Initial, a little sentence structure will lead to lot of text reiteration. Also, in spite of the fact that the content is linguistically perfect, yet there is an absence of semantic structure. The outcome is a series of sentences which have no connection to each other [13].

EXISTING APPROACHES OF TEXT STEGANOGRAPHY

A) Word Spelling

This technique is utilized for concealing information in English text .In this technique Word Spelling in US English are spelt not quite the same as UK English. For Instance:" fulfill" have distinctive terms in UK (fulfill) and US (fulfil) [10].This

technique is helpful for the area where both US &UK sort words spelling less utilized. In this method, our information get cover up or hide if nobody know both technique, however effectively distinguishable. This is on the grounds that there is little distinction between word spelling of UK and US ,as in US (favorite) and

UK (favourite), minor contrast of "u" just, which get effortlessly noticeable. This method is considered as less secure technique then New synonym text method, in light of the fact that in New synonyms text method we utilize distinctive words like (account) in US and (Bill) in UK. This is obviously noticeable that both words utilized equivalent word of each other can't distinguish effectively.

Table 1.*List of Some Words Which Have Different Spelling In Uk And Us*

American English	British English
Favorite	Favourite
Criticize	Criticise
Fulfill	Filfil
Center	Centre
Dialog	Dialogue
Medieval	Mediaeval
Check	Cheque
Defense	Defence
Tire	Tyre

B) Semantic method

This technique is like word spelling method. But little change is, in this technique we can utilize synonyms of words is utilized for specific words in this way the data is hidden in the text[14]. The major advantage for this situation is it likewise ensures information if there should be an occurrence of retyping or utilizing character recognition program (OCR) programs moreover. We have extremely awesome advantage over there however we require more secure technique, since if a man has great command on a language then effectively identify conceal information.

C) Line shifting

In this technique, the lines of text are vertically moved to some degree (like each line moved 1/300 inch up or down) and data is covered up by making an exceptional or unique state of a text. This instrument of distance evaluation and vital changes will be acquainted with crush the concealed information[5]. Likewise, if (OCR) utilized, the concealed data annihilated. This strategy is helpful for printed text, because on printed test OCR not utilized.

In this technique, secret message is covered up by vertically shifting the text lines to

some degree [16, 17]. A line checked has two unmarked control lines one on either side of it for distinguishing the heading of development of the stamped line [16]. To conceal bit 0, a line is moved up and to hide bit 1, the line is moved down [17].

Assurance of whether the line has been moved up or down is finished by measuring the separation of the centroid of stamped line and its control lines [12]. On the off chance that the text is retyped or if a character recognition program (OCR) is utilized, the shrouded data would get destroyed. Additionally, the separations can be seen by utilizing exceptional instruments of distance assessment [16].

D) Abbreviation

Another technique for concealing data in text - Steganography is the utilization of shortened form. In this technique, very little information in the content or text, for example, just couple of bytes of information can stow away in the text [12]. We can hide few lines in several lines. This technique is widely used to shroud or hide information in poems, by characterize some word having other importance, and so on.

E) Word shifting

In this technique, by shifting words on a horizontal plane and by changing separation between words, data are covered up in this text. This technique is satisfactory for writings where distance between words is changing. This technique can be recognized less, since change of separation between words to fill a line is very basic [12].

But if somebody knew about the calculation or algorithm of separations, he can compare the present text and the algorithm and extract the difference. The text image can be likewise firmly considered to recognize the changed difference or distance. Despite the fact that this technique is exceptionally tedious, there is a high likelihood of discovering data hidden in the text. The same as in the past technique, retyping of text or utilizing OCR program destroyed the hidden data.

F) Syntactic Method

In this technique, by Putting some punctuation signs, for example, full stop (.) and comma (,) in appropriate place, one can stow away data in a text file. This technique requires recognizing legitimate place or putting punctuation signs. The amount of data to conceal information is insignificant

[13].For example: We utilize a poem and characterize full stop(.) and comma(,) as zero and one to conceal information in it and send it.

G) New Synonym Text

Fundamental idea of this technique depends on Word Spelling technique to some extent yet extraordinary too. In Word Spelling technique, there is little change in spelling however in New synonyms Method, different word utilized for same data[14]. In English a few words have diverse term in US and UK.

For instance, "rubber" has distinctive term in UK (eraser) and US(rubber).This technique is more gainful then word spelling technique, because in this strategy words have different terms, which can't recognize effectively. In word spelling technique, words are of various spelling like in US (tire) and UK (tyre). The difference is of one letter just, effortlessly perceivable then new synonym text Technique.

Table 2 List of Some Words Which Have Different Terms In Uk And Us

American English	British English
Account	Bill
Candy	Sweets
Closet	Cupboard
Faculty	Staff
Fall	Autumn
Gas	Petrol
Incorporated	Limited
Mail	Post
Movie	Film
Package	Parcel
Soccer	Football
Stove	Cooke

F) Selective hiding

This hides the characters in the first (or a particular area) characters of the words. Linking those characters help separating the text. But, this method requires huge measure of plain text[8].

G) HTML web pages

This may conceal text utilizing the fact that properties of HTML tags are case insensitive. Those characters can then be utilized to recover the first content[8].

H) Hiding using Whitespace

Less number of whitespaces may indicate a 0 and more number of whitespaces between words may decide a 1[9].

I) White Steg [paper3]

This method utilizes void or white areas for concealing a secret message. There are three strategies for stowing away information utilizing white spaces. In inter Sentence spacing, we put single space to hide bit 0 and two spaces to conceal bit 1 toward the finish of each terminating character [15]. In End of Line Spaces, settled or fixed number of spaces is embedded toward the finish of each line. For example, two spaces to encode one bit per line, four spaces to encode two bits and so on. In inter Word Spacing technique, one space after a word speaks to bit 0 and two spaces after a word represent bit 1. In any case, conflicting use of white space is not straightforward [15].

J) Spam Text

HTML and XML files can likewise be utilized to conceal bits. In the event that there are diverse starting and closing tags, bit 0 is deciphered and if single tag is utilized for starting and closing it, then bit 1 is interpreted [17]. In another technique, bit 0 is represented by an absence of space in a tag and bit 1 is represented to by setting or placing a space inside a tag [17].

K) SMS-Texting

SMS-texting language is a blend of contracted words utilized as a part of SMS [14]. We can stow away paired information by utilizing full type of word or its contracted frame. A codebook is made which contains words and their comparing condensed frames. To conceal bit 0, full type of the word is utilized and to shroud bit 1, condensed type of word is utilized [14].

L) Feature Coding

In feature coding, secret message is covered up by altering at least one elements of the text. A parser looks at a document and chooses every feature of the components that it can use to conceal the data [17]. For example, point in letters i and j can be uprooted, length of strike in letters f and t can be changed, or by amplifying or shortening height of letters b, d, h, and so on

[12, 18]. An imperfection of this technique is that if an OCR program is utilized or if re-writing is done, the concealed text would get wrecked.

M) SSCE (Secret Steganographic Code for Embedding)

This technique first scrambles a message using SSCE table and after that embed the cipher text in a cover document or file by embeddings articles an or a with the non particular things in English dialect using a certain mapping methods [19]. The embedding positions are scrambled using the same SSCE table and spared in another file which is transmitted to the recipient safely alongside the stego file.

N) MS Word Document

In this technique, text sections in a document are declined, imitating to be the work of an writer with mediocre composition aptitudes, with secret message being embedded in the decision of degenerations which are then amended with changes being followed [20]. Information inserting is masked to such an extent that the stego record has all the earmarks of being the result of synergistic written work [20].

O) Cricket Match Scorecard

In this technique, information is covered up in a cricket coordinate scorecard by pre-affixing an insignificant zero prior to a number to speak to bit 1 and leaving the number as it is to speak to bit 0 [21].

P) CSS (Cascading Style Sheet)

This technique scrambles a message using RSA public key cryptosystem and cipher text is then inserted in a Falling Cascading style sheet (CSS) by using End of Line on each CSS style properties, precisely after a semicolon. A space after a semicolon embeds bit 0 and a tab after a semicolon embeds bit 1 [22].

CONCLUSION

A coin has two sides. Steganography although conceals the existence of a message but is not completely secure. It is not meant to supersede cryptography but to supplement it. This paper reviews some text steganography methods that hide text. Among all the Steganography techniques, every technique has individual ability to conceal information in text. Be that as it may, It is identified with the setting of text and its size separately. By using line shifting technique, we can cover up tremendous measure of information, however, line shifting technique proficient

for printed text on the grounds that in this technique, other than printed text character recognition program(OCR) is utilized and concealed data get pulverized. Word shifting technique is very valuable to conceal information. In this technique key term is algorithm made and utilized for word shifting. In the event that this algorithm found by another person than additionally security annihilated.

Syntactic strategy utilized as a part of wars to send critical data and cover up little amount of information. Like, we can utilize (.) and (,) in a poems and hide information in (0) as (.) and (1) as (,) and we can get information by extricated it., However in the event that keen locator analyze then concealed information discover and security get break. Semantic technique is productive and its security is higher than past technique in light of the fact that there are no such simple strategy to identify shrouded information introduce yet peruser know gigantic learning of antonym and synonym, identify it. In shortening technique, we can utilize shortening that officially made and we assembled it all alone. It is a protected strategy as contrast with above strategies for information out of vast information get covered up, in light of the fact that we can't

make condensing of all information .It take huge time too. Presently we can look at Word Spelling technique and New synonym text Technique in light of same idea. Both strategies are utilizing for concealing information in English text.

REFERENCES

- 1) Se^ Pahati, Oj (2001-11-29). "Confounding Carnivore: How To Protect Your Online Privacy". Alternet. Archived From The Original On 2007-07-16. Curity Informationshir
- 2) Mohammad Ali-Shahreza "Text - Steganography By Changing Word Spelling"
- 3) N.J Hopper, Towards A Theory Of Steganography,Ph.D Dissertation, School Of Computer Science, Carnegie Mellon University,Pittsburgh,Pa,Usa,July 2004.
- 4) W.BENDER, D.GRUHL, N.MORIMOTO, AND A.LU, "Technique For Data Hiding"IBM Systems Journal, Vol.35, Issue 3&4, 1996, Pp.316-336.
- 5) A.M.Alattar And O.M.Allater, "Water Marking Electronic Text Documents Containing Justified Paragraphs Andirregular Line Spacing". Proceedings Of Spie— Volume 5306,Security, Stagnography,And Watermarking Of Multimedia Contents Vi,June 2004,Pp 685-695.
- 6) Kim,K.Moon,And I.Oh,"A Text Watermarking Algorithm Based On Word Classification And Inter-Word Spacestatistiocs",Prceeding Of The Seventh International Conference On Document Analysis And Recognition,2003,Pp775-779.
- 7) M. Hassan Shirali-Shahreza , And Mohammad Shirali-Shahreza," A New Synonym Text Steganography On International Conference On Intelligent Information Hiding And Multimedia Signal Processing
- 8) H.Kabetta, B.Y. Dwiandiyanta, Suyoto, "Information hiding in CSS: A secure scheme text-steganography using public key

- Cryptosystem”,IJCIS,pp. 13-22
,Vol.1, No.1,December 2011.
- 9) Mohit Garg,” A Novel Text Steganography Technique Based on Html Documents”,International Journal of Advanced Science and Technology,pp.132-138,Vol. 35, October, 2011.
- 10) L. Y. Por, and B. Delina, “Information hiding- a new approach in text steganography,” 7th WSEAS Int. Conf. on Applied Computer and Applied Computational Science, 2008, pp. 689-695.
- 11) L. Y. Por, T. F. Ang, and B. Delina, “WhiteSteg- a new scheme in information hiding using text steganography,” WSEAS Transactions on Computers, vol.7, no.6, pp. 735-745, 2008
- 12) K. Rabah, “Steganography-the art of hiding data,” Information Technology Journal, vol.3, pp. 245-269, 2004.
- 13) K. Benett, “Linguistic steganography- survey, analysis and robustness concerns for hiding information in text,” Purdue University, CERIAS Tech. Report 2004-13, 2004.
- 14) M. S. Shahreza, and M. H. S. Shahreza, “Text steganography in SMS,” 2007 Int. Conf. on Convergence Information Technology, 2007, pp. 2260-2265.
- 15) W. Bender, D. Gruhl, N. Morimoto, and A. Lu, “Techniques for data hiding,” IBM Systems Journal, vol.35, pp. 313-336, 1996.
- 16) M. H. S. Shahreza, and M. S. Shahreza, ”A new approach to Persian/Arabic text steganography,”In Proceedings of 5th IEEE/ACIS Int. Conf. on Computer and Information Science and 1st IEEE/ACIS Int. Workshop on Component-Based Software Engineering, Software Architecture and Reuse, 2006, pp. 310-315
- 17) J. Cummins, P. Diskin, S. Lau, and R. Parlett, “Steganography and digital watermarking,”School of Computer Science, 2004, pp.1-24.

- 18) J. T. Brassil, S. Low, N. F. Maxemchuk, and L. O. Gorman, "Electronic marking and identification techniques to discourage document copying," IEEE Journal on Selected Areas in Communication, vol.1, pp. 1495-1504, 1995.
- 19) I. Banerjee, S. Bhattacharyya, and G. Sanyal, "Novel text steganography through special code generation," Int. Conf. on Systemics, Cybernetics and Informatics, 2011, pp. 298-303
- 20) T. Y. Liu, and W. H. Tsai, "A new steganographic method for data hiding in Microsoft word documents by a change tracking technique," IEEE Transactions on Information Forensics and Security, vol.2, no.1, pp. 24-30, 2008
- 21) M. Khairullah, "A novel text steganography system in cricket match scorecard," Int. Journal of Computer Applications, vol.21, pp. 43-47, 2011.
- 22) H. Kabetta, B. Y. Dwiandiyanta, and Suyoto, "Information hiding in CSS: a secure scheme text steganography using public key cryptosystem," Int. Journal on Cryptography and Information Security, vol.1, pp. 13-22, 2011.