

# Text -Steganography: Review Study & Comparative Analysis

Swati Gupta and Deepti Gupta

TITS,Bhiwani

**Abstract-**Steganography is the art of hiding the fact that communication is taking place, by hiding information in other information. In this paper, we discuss various types of text -Steganography techniques. There are different methods of Text Steganography. We describe it one by one for hiding secret information in text. There exists a large variety of text Steganography techniques, some are more complex than others and all of them have respective strong and weak points. Different applications have different requirements of the Text Steganography technique used. We describe a common application that needs both absolute invisibility and large secret data get hidden. We analyze some basic method of text based -Steganography techniques and compare it according their performance.

**Keywords-** data hiding, New Text Synonym, Text -Steganography, security, word spelling

## I. INTRODUCTION

Steganography is the art and science of writing hidden messages in such a way that no one, apart from the sender and intended recipient, suspects the existence of the message, a form of security without knowledge of its existence [1]. Our major need is hiding. The word Steganography is of Greek origin and means "concealed writing" from the Greek words steganos (στεγανός) meaning "covered or protected", and graphein (γράφειν) meaning "to write". The first recorded use of the term was in 1499 by Johannes Trithemius in his Steganographia, a treatise on cryptography and Steganography disguised as a book on magic. Generally, messages will appear to be something else: images, articles, shopping lists, or some other cover text and, classically, the hidden message may be in invisible ink between the visible lines of a private letter.

### 1) WHY WE NEED -STEGANOGRAPHY .

In recent few years, our internet growing rapidly day by day. One major area which gained attracted by people is security subjected to internet, subjected to communication. Now a day for security purpose hiding data is most popular technique which receives more attention than cryptography. This is because encryption -decryption pay more attention than hiding data technique.

For establishing hidden communication, various methods including cryptography, coding, -Steganography, etc. are used [2]. -Steganography is the process of hiding data inside other data in such a manner that no other person recipient, even know the existence of it. This is major benefit of steganography over other coding techniques. Most of the -Steganography technique works on image, sound, music, video clips, text, etc [3].

There are four basic type of -Steganography .

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

We describe different Text -Steganography methods and compare it. Text -Steganography is most difficult kind of -Steganography ; this is due to the lack of redundant information in a text file, while there is a lot of redundancy in a picture or a sound file, which can be utilized in -Steganography [4].

## II. DIFFERENT METHOD OF TEXT -STEGANOGRAPHY

### A. Word Spelling

This method is used for hiding data in English Text .In this method Word Spelling in US English are spelt different from

UK English. For Example:” fulfill” have different terms in UK (fulfill) and US (fulfil) [2]. This method is useful for the area where both US &UK type words spelling less used. In this method, our data get hide if no one know both method, but easily detectable. This is because there is little difference between word spelling of UK and US ,like in US (Favorite) and UK (Favourite), minor difference of “u” only, which get easily detectable. This method is considered as less secure method then New Synonym Text method, because in New Synonym Method we use different words like (account) in US and (Bill) in UK. This is clearly visible that both words used synonym of each other cannot detect easily.

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 method is similar to word spelling method. But little change is, in this method we can use synonym of words is used for certain words thereby the information is hidden in the text [4]. The Major benefit in this case is it also protects data in case of retyping or using character recognitions program (OCR) programs also. We have very great advantage over there but we need more secure method, because if a person has great command on a language then easily detect hide data.

### C. Line shifting

In this method, the lines of text are vertically shifted to some degree (like each line shifted 1/300 inch up or down) and information is hidden by creating a unique shape of a text. This instrument of distance assessment and necessary changes will be introduced to destroy the hidden information [5]. In addition, if (OCR) used, the hidden information destroyed. This method is useful for printed text, because on printed test OCR not used,

### D. Abbreviation

Another method for hiding information in Text -Steganography is the use of abbreviation. In this method, very little information hidden in the text, for Example, only

few bytes of data can be hiding in the text [6]. We can hide few lines in hundreds of lines. This method is widely used to hide data in poems, by define some word having other meaning, etc.

*E. Word shifting*

In this method, by shifting words horizontally and by changing distance between words, information are hidden in this text. This method is acceptable for texts where distance between words is varying. This method can be identified less, because change of distance between words to fill a line is quite common [6]. But if somebody was aware of the algorithm of distances, he can compare the present text with the algorithm and extract the difference. The text image can be also closely studied to identify the changed distances. Although this method is very time consuming, there is a high probability of finding information hidden in the text. The same as in the previous method, retyping of text or using OCR program destroyed the hidden information.

*F. Syntactic Method*

In this method, by Placing some punctuation signs such as full stop (.) and comma (,) in proper place, one can hide information in a text file. This method requires identifying proper place or putting punctuation signs. The amount of information to hide data is trivial [7].For example: We use a poem and define full stop(.) and comma(,) as zero and one to hide data in it and send it.

*G. New Synonym Text*

Basic concept of this method is based on Word Spelling method to some extent but different also. In Word Spelling method, there is little change in spelling but in New Synonym Method,different word used for same data[8]. In English some words have different term in US and UK. For example, “rubber” has different term in UK (eraser) and US (rubber).This method is more beneficial then word spelling method, because in this method words have different terms, which cannot detect easily. In word spelling method, words are of different spelling like in US (tire) and UK (tyre). The difference is of one letter only, easily detectable then New Synonym Text Method.

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 r	Football
Stove	Cooke

The above methods are very basic method of data hiding & we analyze their performance rapidly. However, Word spelling & New Text Synonym method need detail study to analyze their performance; both methods have same working process. In both the methods, our goal is to hide data in a text by substituting the words that have different terms & spelling in UK and US. Table I & II shows a number of such words which have different terms and spelling in UK and US. The details of implementing this method are as follows: We have used Java programming language to implement this method [2], [8]. This method is conversion of data from one form to another form. This project is composed of two programsmainly:

- 1- Hiding program, which is responsible for hiding data in text means program of converting data.
- 2- Extractor program which extracts data from the stegno text (text containing hidden data).

At first, we prepare a list or dictionary containing the words that have different terms and spelling in UK and US. The hiding program looks for existing words in the list in the text. Furthermore this program converts the concerned data to an arrangement of 0 and 1 bits. The program will place US term in sentence for hiding of the bit 0 and will place the UK term in the sentence in order to hide the bit 1.This way the data will be hidden in the concerned text. Of course the size of data is hidden in the text in order that the extractor program can work correctly. The extractor program extracts the data from the stegno text. This program identifies the type of words in stegno text by using the list of words having different terms in UK and US and saves the quantity of 0 or 1 in an arrangement according to the fact whether it is a US term or a UK term. Now the hidden data will be extracted through conversion of this arrangement from the bits 0 and 1 to its original format. At the end the extracted data will be saved on the user's computer.

III. PERFORMANCE ANALYSIS

TABLE III. LIST OF ALL TEXT -STEGANOGRAPHY METHOD ANALYSIS ON THE BASIS OF ITS ADVANTAGE AND DISADVANTAGE

	Advantage	Disadvantage
Line Shifting	This method is suitable only for printed text. In printed text OCR(character recognition)never used.	When OCR (character recognition program) applies the hidden information get destroyed.
Word Shifting	This identify less because of change of distance between words to fill line is quite common.	If anyone aware about the algorithm that related to word shifted distance , then easily get hide data.
Syntactic Method	The amount of information to hidden the method is trivial.	Smart reader can find hidden data easily.

	<b>Advantage</b>	<b>Disadvantage</b>
Semantic Method	This method is better than above methods, syntactic, line shifting and word shifting because that cannot detect by retyping or using OCR programs.	Smart reader which has huge knowledge of words their synonyms or antonyms can discover it.
Abbreviation	This method is because it's a kind of any abbreviation present and we built also.	It is limited only for small data means out of large data. Only small part of data can be hidden.
Word Spelling	This is good method for data hiding not only for electronic document but also for printing text. In this method, hidden data is not destroyed.	It is less secure than new synonyms text method because little change in spelling not hides data properly.
New Synonyms Text	This is best method among above methods because in this method we use different terms of words that hide data properly.	This type of method is little bit time consuming because we get search word by word and replace it until we get the result.

**IV. CONCLUSION:-**

Among all the Steganography methods, each method has respective capability to hide data in text. However, It is related to the context of text and its size respectively. By using line shifting method, we can hide huge amount of data, but line shifting method only capable for printed text because in this method, other than printed text character recognition program(OCR) is used and hidden information get destroyed. Word shifting method is quite useful to hide data. In this method key term is algorithm made & used for word shifting. If this algorithm found by someone else than also security destroyed. Syntactic method used in wars to send very important information and hide very small amount of data. Like, we can use (.) and (,) in a poem and hide data in (0) as (.) and (1) as (,) and we can get data by extracted it., But if smart detector

examine then hidden data find out and security get break. Semantic method is efficient and its security is higher than previous method because there are no such easy method to detect hidden data present but reader know huge knowledge of antonym and synonym, detect it. In abbreviation method, we can use abbreviation that already made & we built it on our own. It is a safe method as compare to above methods of data out of large data get hidden, because we cannot make abbreviation of all data .It take large time also. Now we can compare Word Spelling method and New Synonym Text Method based on same concept. Both methods are using for hiding data in English text. In these methods, word spelling in UK English and US English are used. In Word Spelling method, US English and UK English word spelling different from each other to some extent. This method is useful for area where US and UK English language not common to people. This method is less secure because of small change in spelling. In New Synonym Method, US and UK English word purely different from each other. This method hides data properly so New Synonym Text Method is best method among all stenography methods.

**V. REFERENCES:**

[1]SE<sup>^</sup> 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 AND IRREGULAR LINE SPACING". PROCEEDINGS OF SPIE—VOLUME 5306, SECURITY, STAGNOGRAPHY,AND WATERMARKING OF MULTIMEDIA CONTENTS VI,JUNE 2004,PP 685-695.

[6]Y. KIM,K.MOON,AND I.OH,"A TEXT WATERMARKING ALGORITHM BASED ON WORD CLASSIFICATION AND INTER-WORD SPACE STATISTICS",PRCEEDING OF THE SEVENTH INTERNATIONAL CONFERENCE ON DOCUMENT ANALYSIS AND RECOGNITION,2003,PP775-779.

[7]K.BENNETT,"LINGUISTIC STEGNOGRPHY: SURVEY, ANALYSIS AND ROBUSTNESS CONCERN FOR HIDING INFORMATION IN TEXT",PURDUE UNIVERSITY,CERIAS TECH.REPORT 2004-13.

[8]M. HASSAN SHIRALI-SHAHREZA , AND MOHAMMAD SHIRALI-SHAHREZA," A NEW SYNONYM TEXT STEGANOGRAPHY ON INTERNATIONAL CONFERENCE ON INTELLIGENT INFORMATION HIDING AND MULTIMEDIA SIGNAL PROCESSING.