IOSR Journal of Computer Engineering (IOSRJCE)
ISSN: 2278-0661 Volume 3, Issue 4 (July-Aug. 2012), PP 11-17
www.iosrjournals.org

"A Novel Approach of Text Steganography based on null spaces"

Prem Singh, Rajat Chaudhary and Ambika Agarwal
Department of Information Technology, DIT (Uttrakhand Technical University)

Abstract: *Steganography is the art and science of covered or hidden writing. The purpose of steganography is covert communication to hide the existence of a message from an intermediary. Digital Steganography algorithms have been developed by using texts, images and audio etc as the cover media. In this approach, a security model is proposed which imposes the concept of secrecy over privacy for text messages. In the recent years, we have plenty of security tools which are developed to protect the transmission of multimedia objects. But approaches for the security of text messages are comparatively less.*

In this approach we present a new approach on text steganography through the null space in the cover message for hiding the secret message. In this method, hiding bits of secret message is done through adding extra null spaces in the plaintext of the cover file. These null spaces placed, when the binary bit of secret message is equal to 1 in plaintext of the cover file. And null space remains unchanged when the binary bit of the secret message is equal to 0.

Keywords: *Steganography, cover message, secret message*

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 through obscurity. The word steganography is of Greek origin and means "concealed writing" from the Greek words steganos meaning "covered or protected", and graphein meaning "to write".

Steganography has been widely used, including in recent historical times and the present day. Some possible known examples include: Hidden messages within wax tablets: in ancient Greece, people wrote messages on the wood, and then covered it with wax upon which an innocent covering message was written. Apparently neutral's protest is thoroughly discounted and ignored. Isman hard hit. Blockade issue affects pretext for embargo on by-products, ejecting suits and vegetable oils. By extracting the second letter from each word, this hidden message can be decoded as:

Pershing sails from NY June 1. [1]

The development of new digital technologies has given an opportunity to improve message detection that can pass more information and even be less conspicuous in transmission such as the microdots technology developed by the Germans. Microdots uses microscopic shrink technique to hide pictures of text which can only be read using a microscope. German spies used them in many different ways like messages hidden in letters, on the face of watches and even on spotted ties.

The application of computer in real life is increasing day by day. So, the need to secure data is becoming more and more essential part of message or data transfer. Information security became a part of our daily life. Among the different techniques, hidden exchange of information is one of the concerns in the area of information security. Various methods like Cryptography, steganography, coding and so on have been used for this purpose. However, during recent years, steganography, perhaps has attracted more attention than others.

In implementing steganography, the main objective of Steganography is to hide the information under a cover media so that the outsiders may not discover the information contained in the said media. This is the major distinction between Steganography and other methods of hidden exchange of information. For example, in cryptography method, people become aware of the existence of data by observing coded data, although they are unable to comprehend the data. However, in Steganography, nobody notices the existence of data in the resources.

Most Steganography works have been performed on images, video clips, text, music and sounds. However text Steganography is the 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 used in Steganography. In this approach we presents a new approach on text steganography through the white/null space in the cover message for hiding the secret message. In this method, hiding bits of secret message is done through adding extra white/null spaces in the plaintext of the cover file. These white/null spaces placed, when the binary bit of secret message is equal to 1 in plaintext of the cover file. And white/null space remains unchanged when the binary bit of the secret message is equal to 0.

The structure of text documents is identical with what we observe, while in other types of documents such as in a picture, the structure of document is different from what we observe. Therefore, in such documents, we can hide information by introducing changes in the structure of the document without making a notable change in the concerned output.

Text Steganography is one of the most difficult methods because a text file is not a proper media to hide data in it. Among the most important of these technologies, one can name the hiding information in electronic texts and documents (e-documents). The use of hiding information in text for web pages is another example. Text steganography is the art or process of hiding one text into another text for the purpose of secure communication so that the illegitimate user/ unauthorized user can't get the secret message for their own use.

II. Related work:

Before giving a brief description of the related work on Text Steganography, it will be appropriate to discuss the frequent terms used in this context. The secret message to be hidden is referred to as **embedded data** and the innocuous text / audio / image used for embedding is called as **cover**. The resultant output object after embedding is referred to as **stego-object**. It is a priory for the sending and receiving ends to have agreed upon on a mutual key exchange protocol / mechanism.

A few works have been done on hiding information in texts. Following is the list of different methods of the works carried out and reported so far.

2.1 In general, the Text Steganography methods can be categorized into two groups:

1- Changing the text format.
2- Changing the meaning of the text.

The methods which are based on the changing the meaning of the text are limited. Some examples of these methods are as follows:

2.1.1. Syntactic method :

By placing some punctuation marks such as full stop (.) and comma (,) in proper places, one can hide information in a text file. This method requires identifying proper places for putting punctuation marks. The amount of information to hide in this method is trivial.

2.1.2. Semantic method :

The authors Mohammad Sirali-Shahreza and M. Hassan Shirali-Shahreza in [8] have used synonym words substitution for hiding secret message bits on the anology of 2.1-2.2.

This method uses the synonym of certain words thereby hiding information in the text. The synonym substitution may represent a single or multiple bit combination for the secret information. A major advantage of this method is the protection of information in case of retyping or using OCR programs. However, this method may alter the meaning of the text.

Big	Large
Small	Little
Chilly	Cool
Smart	Clever
Spaced	Stretched

TABLE 1 Semantic method

2.1.3. Text abbreviation or acronym :

Another method for hiding information is the use of abbreviations or acronym. Mohammad Sirali-Shahreza and M.Hassan Shirali-Shahreza from Iran have suggested the use of substitution of words with their respective abbreviations or *viza viz* in [6] to hide bits of secret message. Their suggested method operates as follows:

Acronym (0)	Translation (1)
2l8	Too late
ASAP	As Soon As Possible
C	See
CM	Call Me
F2F	Face to face

TABLE 2 Text abbreviation or Acronym

Methodology: *A table of two columns is organized with a preselected list of words and their corresponding acronyms in such a way that the column under which words or its translation will appear is labeled as '1' while that containing respective acronym is labeled as '0'.The message/information required to be hidden is converted into its equivalent binary (Table 2 refers for detail).*

In this method, very little information can be hidden in the text. For example, only a few bits of information can be hidden in a file of several kilobytes.In this method words can be substituted with their abbreviations to represent the binary bit pattern of zero or one corresponding to the bits of secret information.

2.1.4. Change of spelling :

In his research paper [7] Mohammad Shirali-Shahreza presented a method to exploit same words which are spelled differently in British and American English for hiding secret message bits. The concealment methodology elaborated below, where the words spelled in British and American English are arranged in separate columns; is identical to that explained in preceding sub-pera.

Methodology: *The column labeled '1' contain words with British spellings while that containing same words of American spelling is given a label '0'.The secret message is converted into its equivalent binary. The cover message is then iterated to find words that match to those available in pre-defined list. On finding a matching word, the secret message bit is mapped to the column headings and the word at the cross-section of that column and matched word's row is substituted in the cover message for the matched word.*

Whole of the cover message is iterated to find matching words in the list followed by substitution of word under column indicated by secret message bit. This method exploits the way words are spelled in British and American English for hiding secret information bits.

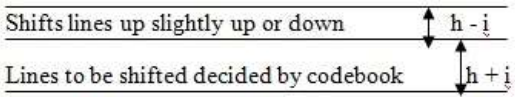
American Spelling	British Spelling
Favorite	Favourite
Criticize	Criticise
Fulfill	Fulfil
Center	Centre

TABLE 3: Change of Spelling

The methods which change the format of the text usually have a large capacity for hiding information. Some examples of these methods are as follows:

2.1.5. Line shifting method :

In this method, the lines of the text are vertically shifted to some degree (for example, each line is shifted 1/300 inch up or down) and information are hidden by creating a unique shape of the text. This method is suitable for printed texts.

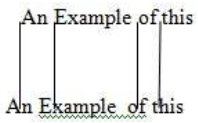


(Figure 4: Line Shifting)

However, in this method, the distances can be observed by using special instruments of distance assessment and necessary changes can be introduced to destroy the hidden information. Also if the text is retyped or if character recognition programs (OCR) are used, the hidden information would get destroyed. This method hides information by shifting the text lines to some degree to represent binary bits of secret information.

2.1.6. Word shifting method :

In this method, by shifting words horizontally and by changing distance between words, information is hidden in the text. This method is acceptable for texts where the distance between words is varying. This method can be identified less, because change of distance between words to fill a line is quite common.



(Figure 5: Word Shifting)

But if somebody was aware of the algorithm of distances, he can compare the present text with the algorithm and extract the hidden information by using 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. Retyping of the text or using OCR programs destroys the hidden information.

2.1.7. Feature coding :

In feature coding method , some of the features of the text are altered. For example, the end part of some characters such as h, d, b or so on, are elongated or shortened a little thereby hiding information in the text. In this method, a large volume of information can be hidden in the text without making the reader aware of the existence of such information in the text.

This method hides the secret information bits by associating certain attributes to the text characters. By placing characters in a fixed shape, the information is lost. Retying the text or using OCR program destroys the hidden information.

Some other methods are:

2.1.8. Miscellaneous techniques :

A large number of idiosyncrasies methods have been given by the authors in [9] that may be used for hiding secret message bits inside a cover text e.g., by introducing modification or injecting deliberate grammatical word/sentence errors with in text. Some of the suggested techniques / procedures given in this context include:

(i) *Typographical errors - "tehre" rather than "there".*

(ii) *Using abbreviations / acronyms - "yr" for "your" / "TC" in place of "Take Care".*

(iii) *Transliterations – "gr8" rather than "great".*

2.1.9. Steganography of information in specific characters in words :

In this method, some specific characters from certain words are selected as hiding place for information. In the simplest form, for example, the first words of each paragraph are selected in a manner that by placing the first characters of these words side by side, the hidden information is extracted. This has been done by classic poets of Iran as well.

This method requires strong mental power and takes a lot of time. It also requires special text and not all types of texts can be used in this method.

2.1.10. HTML tags :

HTML Tags can be used in varying combination or as gaps and horizontal tabulation to represent a pattern of secret information bits. e.g.,

Stego key:

<img></img> -> 0
<img/> -> 1

Stego data:

<img src=g1.jpg></img>

<img src=g5.jpg></img>
Hidden Bits: 01110

2.1.11. Hiding data by changing case of tag [12][13]**:**

As already stated, HTML Tags and associated members are case insensitive e.g., <html>, <HTML> or <hTmL> will have the same impact on the document's outlook. Bits are hidden in TAGS by changing the case of the alphabets based on the bit as either '0' or '1'.

The draw back is that the changes are frequent and besides eye catching these can also be easily decoded to extract the hidden information in the absence of any pre-agreed stego-key and usage details.

III. Our approach:

3.1 Block Diagram of given Model:

Here we propose a new hybrid model using white/null space technique to hide the secret message in text steganography.

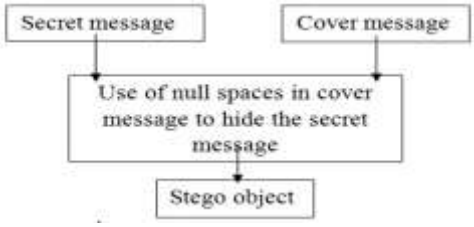


Figure8: Broad level steps in Text Steganography

3.2 Working of the given model:

Hiding information within spaces seems to be potential as people hardly can know about the existence of the hidden bits. This approach shows that one space is interpreted as "0" whereas two spaces are interpreted as "1". This embedding scheme was applied in the space which appears between the words. The major drawback of this method is that it requires a great deal of space to encode few bits. For example, a character is equivalent of 8 bits, and it requires approximately 8 inter-spaces to encode one character.

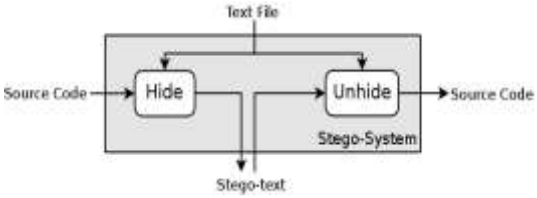


Figure9: How Stego-System works

In this approach we present a new approach on text steganography through the white/null space in the cover message for hiding the secret message.

3.2.1 STEP in PROGRAM:

This approach is composed of two programs mainly:

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).

3.2.1.1 In Hiding program:

- Firstly enter the secret message which we want to hide.
- Enter the name of the file as a cover text in which we want to hide our secret message.
- Generate a coverd text file as a stego-text in which our secret message is hidden.

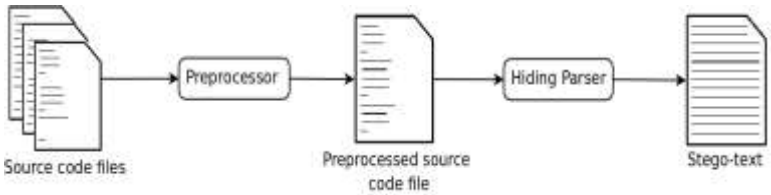


Figure10: Hiding of Secret message

3.2.1.2 In Unhiding program:

- Firstly enter the name of the text file (stego-text) which generate after the hiding of the secret message.
- Extract the secret message.
-

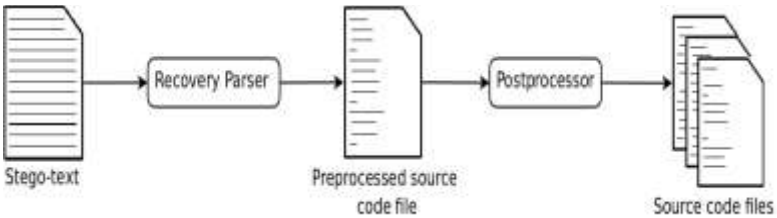
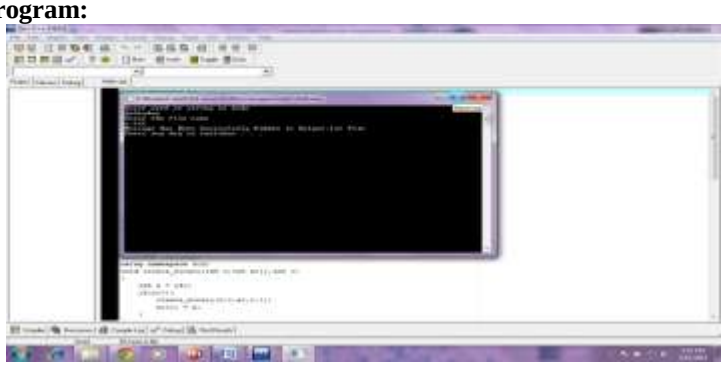


Figure11: Unhiding the secret message

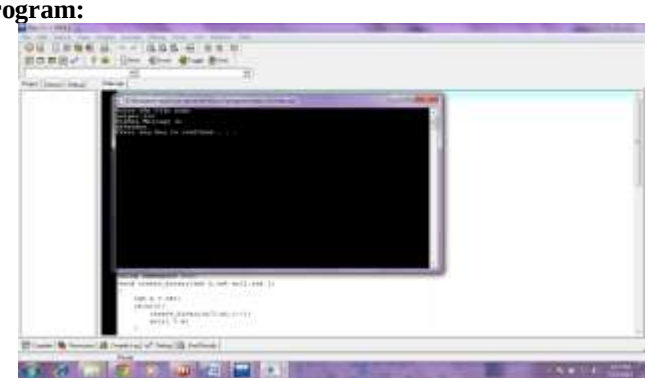
IV. Experimental Result:

In this approch, hiding information within spaces seems to be potential as people hardly can know about the existence of the hidden bits. This thesis shows that one space is interpreted as "0" whereas two spaces are interpreted as "1". This embedding scheme was applied in the space which appears between the words.

4.1 Output of Hide program:



4.2 Output of Unhide program:



The major drawback of this method is that it requires a great deal of space to encode few bits. For example, a character is equivalent of 8 bits, and it requires approximately 8 inter-spaces to encode one character. The challenges concerned with hiding data in text files include:

i. Low storing capacity of text in cover message in text steganography.
ii. Inserting additional spaces to represent information results in an increase in stego-cover object size.
iii. It can store the secret message only in the text file.
iv. Changing a single bit of a byte results in an all together different ACII code that may/may not have any relevancy with the text contents.

V. Conclusion and Future work:

Text Steganography is one of the most difficult methods because a text file is not a proper media to hide data in it. In this there are a lot of shortcoming in the form of low storing capacity of the cover message and the useless increasing size of the cover message with the use of null spaces. So we need to resolve these problems for making text steganography more effective.

- Improve weakness of the system:

* Low storing capacity of text in cover message in text steganography.
* Inserting additional spaces to represent information results in an increase in stego-cover object size.
* Some text editor programs automatically delete extra white-spaces and thus destroy the hidden information.

- Hide any kind of information, not only text file.

In future we can make this more effective by adding some extra feature. In future we will use indexing feature for storing the same letters used in secret message for reducing the low storing capacity of words and reducing the size of the cover message by the limited use of the null spaces in cover message and the scope of the project can be extended by turning the software into a web application and it can be extended by adding an e-mailing feature to the software where the encoded stream can be mailed to the receiver to get it decoded. This feature can make the software really efficient.

References:

[1] Debnath Bhattacharyya, Poulami Das, Samir Kumar Bandyopadhyay and Tai-hoon Kim, "Text Steganography: A Novel Approach", International Journal of Advanced Science and Technology Vol. 3, pp. 79-86, February, 2009.

[2] S. Changder, N.C. Debnath and D. Ghosh, "A New Approach to Hindi Text Steganography by Shifting Matra", 2009 International Conference on Advances in Recent Technologies in Communication and Computing, pp.199-202, 2009.

[3] S. Changder, N. C. Debnath and D. Ghosh, "A Statistical Attack on a Kind of Word-Shift Text-Steganography", 2011 Eighth International Conference on Information Technology: New Generations, pp.30-35, 2011.

[4] S.Changder, D. Ghosh and N. C. Debnath, "Linguistic Approach for Text Steganography through Indian Text", 20IO 2nd International Conference on Computer Technology and Development (ICCTD 2010), pp.318-322, 2010.

[5] Natthawut Samphaiboon, "Steganography via running short text messages", Springer Science+Business Media, LLC 2009, pp.569–596, Published online: 30 December 2009.

[6] Mohammad Sirali-Shahreza, M. Hassan Shirali- Shahreza, *Text Steganography in Chat*, 1-4244-1007/07 © 2007 IEEE

[7] Mohammad Shirali-Shahreza, *Text Steganography by Changing Words Spelling*, ISBN 978-89-5519-136-3, Feb. 17- 20, 2008 ICACT 2008

[8] M. Hassan Shirali-Shahreza, Mohammad Shirali- Shahreza, *A New Synonym Text Steganography* ,International Conference on Intelligent Information Hiding and Multimedia Signal Processing 978-0-7695-3278-3/08 © 2008 IEEE

[9] Mercan Topkara, Umut Topkara, Mikhail J. Atallah, *Information Hiding Through Errors: A Confusing Approach*, Purdue University

[10] Tsung-Yuan Liu, Wen-Hsiang Tsai,and Senior Member, *A New Steganographic Method for Data Hiding in Microsoft Word Documents by a Change Tracking Technique*, 1556-6013 © 2007 IEEE

[11] NeoByte Solutions, "*Invisible Secrets 4*", http://www.invisiblesecrets.com/index.html

[12] Mohammad Shirali Shahreza, *A New Method for Steganography in HTML Files*, Computer, Information, and Systems Sciences, and Engineering, Proceedings of IETA 2005, TeNe 2005, EIAE 2005, 247-251, Springer

[13] K. Bennett, "*Linguistic Steganography: Survey, Analysis, and Robustness Concerns for Hiding Information in Text*", Purdue University, CERIAS Tech. Report 2004-13

[14] HIPS Systems, "Shadow Text", http://home.apu.edu/~jcox/ projects/HtmlStegol