RESEARCH ARTICLE OPENACCESS

Data Hiding Process with Indexes Based on Whitespace Method

Thu Zar Win [1], Aye Wai Oo [2]
Department Of Information Technology Engineering
West Yangon Technological University
Myanmar

ABSTRACT

This research paper proposes to solve the insufficient cover text problem and used the enhancement of white space method for hiding data, which is processed by changing the secret text through extracting the indexes of the characters from the cover text or ASCII code, then converting these numbers from the decimal numeral system into the octal numeral system in order to use the number 8 and 9 as indicators against the remaining numbers. Then, merge these outcomes with the white spaces between the words in the cover text by changing the font size of these numbers to 1pt, and changing the font color to match background color of the cover text. This system composed by three format-based methods such as inter-sentence spacing, end-of-line spacing and inter-word spacing. This system used double encoding process using octal number.

Keywords:-- cover text , index , secret text , steganography and white space.

I. INTRODUCTION

Information hiding is a general term encompassing many subdisciplines. One of the most important subdisciplines is steganography. Steganography is derived from a finding by Johannes Trithemus (1462-1516) entitled “Steganographia” and comes from the Greek words (στεγανός) meaning “covered writing” [1, 2, 3]. Steganography is the art and science of hiding a message inside another message without drawing any suspicion to others so that the message can only be detected by its intended recipient [4].The system presents three white space methods; encoding a binary message into a text after each terminating character, encoding data by inserting spaces at the end of lines and by encoding data that involves justifying format of text where the extra spaces are placed. Open space method is one of the first used methods to hide data in white space between words, lines and paragraph. This system is composed by three formatted-based methods such as inter-sentence spacing, end-osf-line spacing and inter-word spacing. The advantage of this system is the sufficient of cover text.

II. METHODS OF TEXT STEGANOGRAPHY

There are three basic categories of text steganography (Fig. 1) maintained here: format-based methods, random & statistical generation and linguistic methods.

i. **Format-based method :** This method use the physical formatting of text as a space in which to hide information. Format-based methods usually modify existing text for hiding the steganographic text. Insertion of spaces or non-displayed characters, careful errors tinny throughout the text and resizing of fonts are some of the many format-based methods used in text steganography.

ii. **Random and statistical generation method :** This method avoid comparison with a known plaintext, steganographers often resort to generating their own cover texts. Character sequences method hide the information within character sequences.

iii. **Linguistic method :** The affluence of electronic documented information available in the world as well as the exertion of serious linguistic analysis makes this an interesting medium for steganography information hiding.

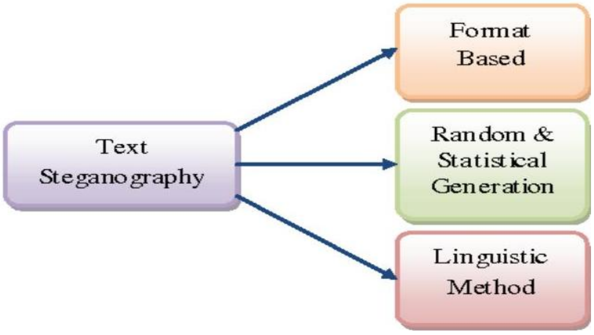


Figure 1. Text Steganography Methods

III. PROPOSED METHOD OF SYSTEM

A.Open Space Method

There are couple ways to employ the open space in text files to encode the information. This method works because to a casual reader one extra space at the end of line or an extra space between two words does not prompt abnormality. However, open space methods are only useful with ASCII format.

- Inter-sentence space method encodes a “0” by adding a single space after a period in English prose. Adding two spaces would encode a “1”. This method works, but requires a large amount of data to hide only little information. Also many word processing tools automatically correct the spaces between sentences.
- End-of-line space method exploits white space at the end of each line. Data encoded using a predetermined number of

spaces at the end of each line. For example two spaces will encode one bit, four spaces will encode two bits and eight spaces will encode three bits and so on. This technique works better than the interspace method, because increasing the number of spaces can hide more data.

B.Challenging in Open Space Method

To hide two words like "Top Secret" requires text size cover of more than 80 words; because each character size is 8 bit "1 byte" and each bit requires one space.That means "T+o+p+ +S+e+c+r+e+t" equal 10 characters, 10 characters multiply by 8 bits equal 80 bits.To be able to hide a large secret message; the result will be a very large messageIn a properly justified format of text, not all spaces are available to be used to hide the required data. Data hiding techniques are obviously suffering some major issues and in some certain cases may become inefficient.

C.Algorithm for embedding

Embedding algorithm
Input : Cover text in richtextfile(.rtf)
Output :Stego-text

Step1:Scan the cover text document to find white space characters:

Index of the secret text characters:

- ❑ Taking each character from the secret text.
- ❑ Looking for character in the cover text.
- ❑ Taking the index of the character from the cover text.
- ❑ Compilation of the indexes taken from the cover text in the array.

Step2:Insert ASCII code characters of the secret text characters:

- ❑ If the character does not exist in the cover text or exists with different capitalization:get the character code from the ASCII code characters as it's listed in ASCII Table.
- ❑ Compile the indexes taken from the cover text in the array.

Step3: Change Octal numeral system:

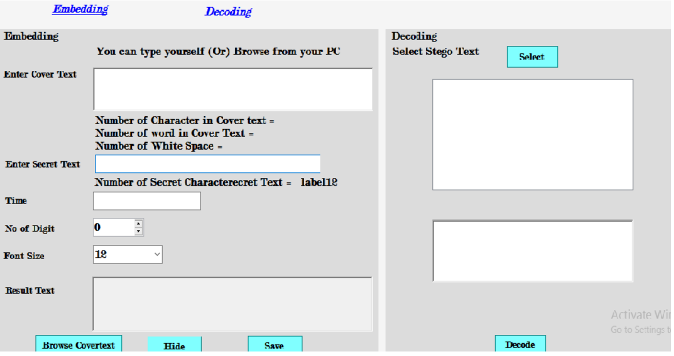
- ❑ Convert the ASCII code from decimal numeral system into octal numeral system.
- ❑ Convert indexes of the cover text from decimal numeral system into octal numeral system.
- ❑ Add number "8" begin of the converted octal number; this step to distinguish the number is from ASCII code characters.
- ❑ Compile the indexes and the ASCII codes for secret text characters in the same order in the text by inserting the number "9" between these numbers to separate the characters indexes and ASCII codes from each other.

Step4:Hide the secret text

- ❑ Detect the number of digits
- ❑ Merge the indexs and ASCII code numbers in the cover text
- ❑ Change font size and color for the numbers

Step5: Repeat step 1 to 4 until complete covered the secret text

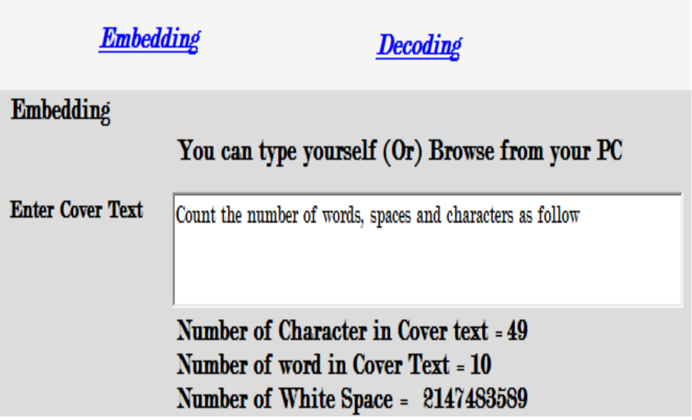
IV. EXPERIMENTAL RESULTS OF SYSTEM



Embedding
You can type yourself (Or) Browse from your PC
Enter Cover Text
Number of Character in Cover text =
Number of word in Cover Text =
Number of White Space =
Enter Secret Text
Number of Secret Characterecret Text = label12
Time
No of Digit 0
Font Size 12
Result Text

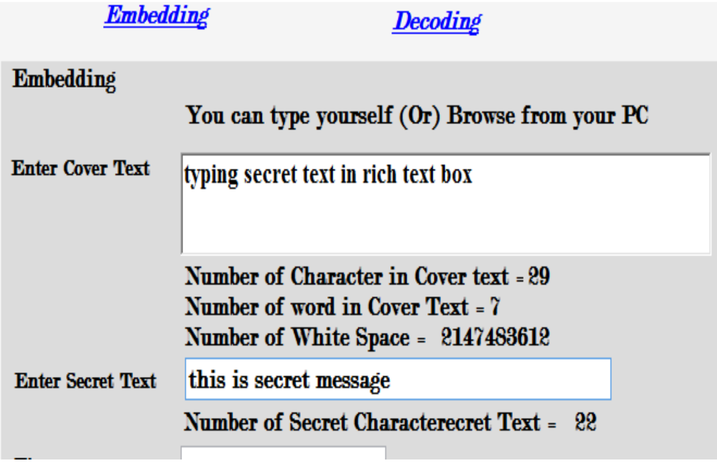
Decoding
Select Stego Text

Fig2: Home screen of system



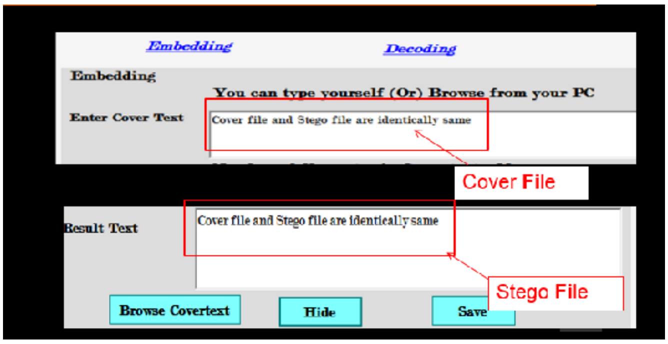
Embedding
You can type yourself (Or) Browse from your PC
Enter Cover Text: Count the number of words, spaces and characters as follow
Number of Character in Cover text = 49
Number of word in Cover Text = 10
Number of White Space = 2147483589

Fig3: We can synchronously count the number of characters, words, and spaces



Embedding
You can type yourself (Or) Browse from your PC
Enter Cover Text: typing secret text in rich text box
Number of Character in Cover text = 29
Number of word in Cover Text = 7
Number of White Space = 2147483612
Enter Secret Text: this is secret message
Number of Secret Characterecret Text = 22

Fig4: Type your secret text in Rich TextBox and check your secret character numbers

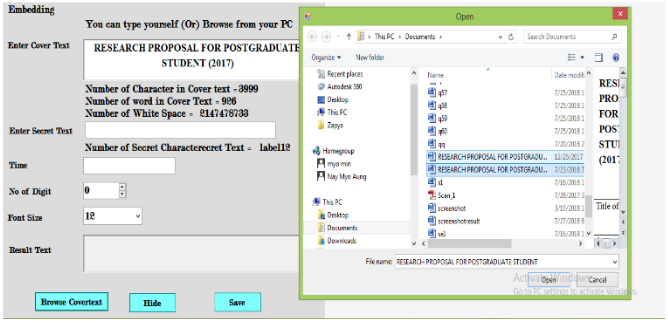


You can type yourself (Or) Browse from your PC

Enter Cover Text: Cover file and Stego file are identically same

Result Text: Cover file and Stego file are identically same

Fig5:Cover file and Stego file are identical



You can type yourself (Or) Browse from your PC

Enter Cover Text: RESEARCH PROPOSAL FOR POSTGRADUATE STUDENT (2017)

Number of Character in Cover text - 3999
Number of word in Cover Text - 926
Number of White Space - 2147478733

Enter Secret Text:

Number of Secret Characterecret Text - label12

Time:

No of Digit: 0

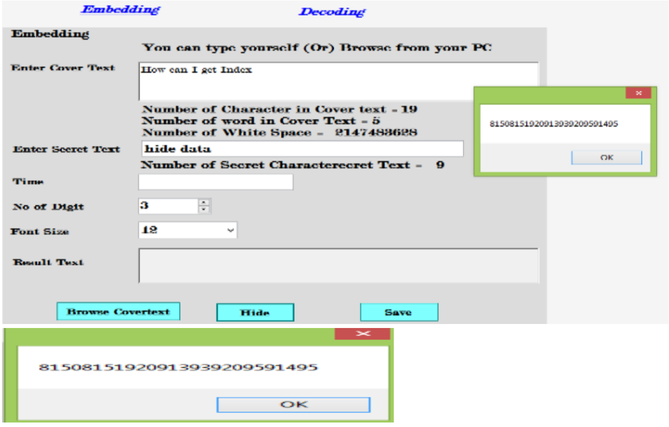
Font Size: 12

Result Text:

q57 7/25/2018 1
q58 7/25/2018 1
q59 7/25/2018 1
q60 7/25/2018 1
qq 7/20/2018 2
RESEARCH PROPOSAL FOR POSTGRADU... 12/25/2017
RESEARCH PROPOSAL FOR POSTGRADU... 7/23/2018 7
s1 7/16/2018 1
Scan_1 7/28/2017 3
screenshot 3/10/2018 1
screenshotresult 7/27/2018 8
sel 7/16/2018 1

File name: RESEARCH PROPOSAL FOR POSTGRADUATE STUDENT

Fig6: We can browse rtf cover file to use as the Cover text



You can type yourself (Or) Browse from your PC

Enter Cover Text: How can I get Index

Number of Character in Cover text - 19
Number of word in Cover Text - 5
Number of White Space - 2147488628

Enter Secret Text: hide data

Number of Secret Characterecret Text - 9

Time:

No of Digit: 3

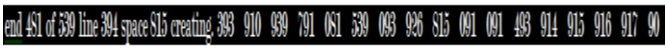
Font Size: 12

Result Text:

81508151920913939209591495

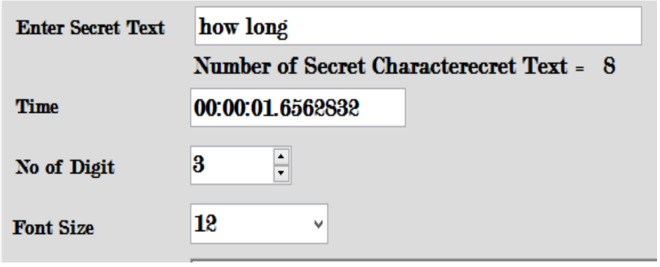
81508151920913939209591495

Fig7: export index of secret text



end 481 of 539 line 394 space 815 creating, 393 910 939 791 081 539 093 926 815 091 091 493 914 915 916 917 90

Fig8: Using end of line space when secret text more than cover text



Enter Secret Text: how long

Number of Secret Characterecret Text - 8

Time: 00:00:01.6562832

No of Digit: 3

Font Size: 12

Fig9: evaluation the speed of data hiding process

V. COMPARISON OF TEXT STEGANOGRAPHY AND DATA HIDING USING INDEXES

TABLE 1

Text Steganography methods	Advantage	Dis-Advantage
Line shifting	This method is suitable only for printed text.	When OCR(character recognition program)applied the hidden information gets destroyed.
Word shifting	Word shifting method identify less because of change of distance between words to fill is quite common.	The algorithm that related to word shifted distance, easily can get hidden data.
Syntactic method	The amount of information to hidden the method is trivial.	Smart reader can find hidden data easily.
Semantic-based hiding method	This method is better than above methods ,line shifting,word shifting,syntactic method 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.
Data hiding using indexes based on white space	One way of hiding data in text is to use white space.Due to the fact that in practically all text editors,extra white space at the end of lines is skipped over,it won't be noticed by casual viewer.We can hide any characters and very secure data because even the unauthorized user found the index, they can not decode the secret text.Since no change is made to the cover ,the cover file and its correspondin stego file are exactly the same..	In a large piece of text and a little change the original file size
Data hiding in paragraph	The approach works by hiding a message using start and end letter of the words of a cover file. Since no change is made to the cover ,the cover file and its correspondin stego file are exactly the same.	The volume of data hiding in the paragraph would be very less.The capacity of hiding the large volume of data leads to the challge.

VI. CONCLUSION

The proposed system takes advantage of the unused white space from the text "Cover Text" to hide the data "Secret Data" on the cover text. Another risk associated with text data hiding is the insufficient of covered text. Security has always been important in electronic applications. The key advantage of stego is the ability to communicate without anyone knowing the true content of communication. Steganography techniques used to hide secret messages in stegoobjects. The Steganography methods applied to different media. In proposed system a text is encrypted before being hidden in order to achieve a better level of secrecy. So we proposed the idea for this paper with data hiding using combination of three formatted based methods such as inter sentence spacing, end-of-line spacing and inter-word spacing with sufficient cover text.

REFERENCES

[1] F. A. P. Petitcolas, R.J. Anderson, and M. G. Kuhn, "Information hiding- a survey," In *Proceedings of IEEE*, vol.87, pp. 1062-1078, 1999.

[2] 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.

[3] 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.

[4] S. Changder, D. Ghosh, and N. C. Debnath, "Linguistic approach for text steganography through Indian text," 2010 2nd *Int. Conf. on Computer Technology and Development*, 2010, pp. 318-322.

[5] R.J. Anderson, and F. A. P. Petitcolas, "On the limits of steganography," *IEEE Journal of Selected Areas in Communication*, vol.16, pp. 474-481, 1998.

[6] K. Rabah, "Steganography-the art of hiding data," *Information Technology Journal*, vol.3, pp.245-269, 2004.

[7] K. Benett, "Linguistic steganography- survey, analysis and robustness concerns for hiding information in text," Purdue University, CERIAS Tech. Report 2004-13, 2004.

[8] Dr. Mohammed Al-Mualla and Prof. Hussain Al-Ahmad, "Information Hiding: Steganography and Watermarking".[Online].Available:*http://www.emirates.org /ieee/information_hiding.pdf* [Accessed: March 12, 2008].

[9] K. Benett, "Linguistic steganography- survey, analysis and robustness concerns for hiding information in text," Purdue University, CERIAS Tech. Report 2004-13, 2004.

[10] M. S. Shahreza, and M. H. S. Shahreza, "Text steganography in SMS," 2007 *Int. Conf. on Convergence Information Technology*, 2007, pp. 2260-2265.

[11] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Systems Journal*, vol.35, pp. 313-336, 1996.

[12] 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 1stIEEE/ACIS Int. Workshop on Component-Based Software Engineering, Software Architectureand Reuse*, 2006, pp. 310-315.

[13] Nosrati, Masoud.,RonakKarimidan Mehdi Hariri.2011. *An Introduction to Steganography Methods.* World Applied Programming, Vol 1, No 3, 191-195, Agustus 2011

[14] Saleh Saraireh. 2013. *A Secure Data Communication System Using Cryptography and Steganography*. International Journal of Computer Networks & Communications (IJCNC) Vol. 5 No. 3, May 2013.