WhiteSteg: A New Scheme in Information Hiding Using Text Steganography

L. Y. Por[1], T. F. Ang[2], B. Delina[3]
Faculty of Computer Science and Information Technology,
University of Malaya,
50603, Kuala Lumpur,
MALAYSIA
porlip@um.edu.my[1], angtf@um.edu.my[2], delinabeh@yahoo.com [3]

Abstract: - Sending encrypted messages frequently will draw the attention of third parties, i.e. crackers and hackers, perhaps causing attempts to break and reveal the original messages. In this digital world, steganography is introduced to hide the existence of the communication by concealing a secret message inside another unsuspicious message. The hidden message maybe plaintext, or any data that can be represented as a stream of bits. Steganography is often being used together with cryptography and offers an acceptable amount of privacy and security over the communication channel. This paper presents an overview of text steganography and a brief history of steganography along with various existing techniques of text steganography. Highlighted are some of the problems inherent in text steganography as well as issues with existing solutions. A new approach, named WhiteSteg is proposed in information hiding using inter-word spacing and inter-paragraph spacing as a hybrid method to reduce the visible detection of the embedded messages. WhiteSteg offers dynamic generated cover-text with six options of maximum capacity according to the length of the secret message. Besides, the advantage of exploiting whitespaces in information hiding is discussed. This paper also analyzes the significant drawbacks of each existing method and how WhiteSteg could be recommended as a solution.

Key-Words:- Steganography,Text Steganography, Information Hiding, Security, Suspicion.

1 Introduction

Information hiding techniques has become the newest hot spots [27] in security research. New applications and new technologies bring new threats, thus new protection mechanisms have to be invented. Moreover, the need for private and sufficiently secure communications in several applications such as e-banking, e-trading, mobile telephony, medical data interchanging etc., is rapidly increasing [36]. With these forces driving it, research in information hiding has grown explosively.

1.1 Information hiding: A Brief History

Information hiding is a general term encompassing many subdisciplines. One of the most important subdisciplines is steganography [27] as shown in Figure 1.

Steganography, is derived from a finding by Johannes Trithemus (1462-1516) entitled "Steganographia" and comes from the Greek (στεγανό-ς, γραφ-ειν) defined as "covered writing" [20][21]. It is an ancient art of hiding information in ways a message is hidden in an innocent-looking cover media so that will not arouse an eavesdropper's suspicion.

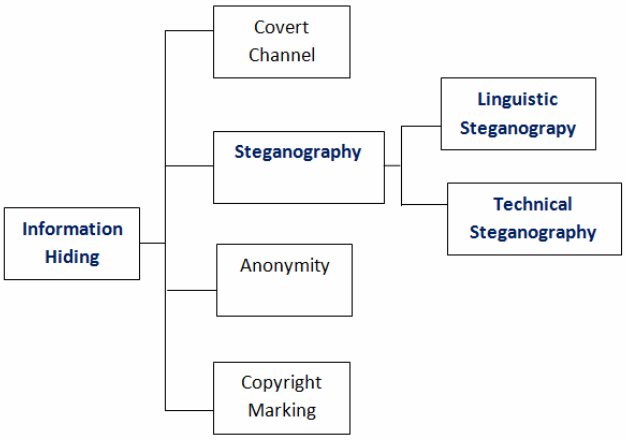


Fig.1: A Classification of Information Hiding Techniques

The goal of steganography is to transmit a message through some innocuous carrier i.e. text, image, audio and video over a communication

channel where the existence of the message is concealed. Based on Figure 1, steganography is one of the information hiding techniques which can be classified into linguistic steganography and technical steganography. Linguistic steganography is defined by Chapman et al. [25] as "the art of using written natural language to conceal secret messages". The main component of the linguistic steganography consists of a medium which required the steganographic cover that is composed of natural language text and the text itself which can be generated to have a cohesive linguistic structure [21]. Conversely, technical steganography is explained as a carrier rather than a text which can be presented, as any other physical medium such as microdots and invisible inks.

During World War II, invisible inks offered a common form of invisible writing. With the invisible ink, a seemingly innocent letter could contain a very different message written between the lines. Therefore, the text document can conceal a hidden message through using null ciphers (unencrypted message), which perfectly camouflage the real message in an ordinary letter. In [13], there is an example on one of the most significant null cipher messages sent by a Nazi spy:

Apparently neutral's protest is thoroughly discounted and ignored. Isman hard hit. Blockade issue affects pretext for embargo on by-products, ejecting suets and vegetable oils [13].

By extracting the second letter from each word, this hidden message can be decoded as:

Pershing sails from NY June 1. [13]

The development of new digital technologies has opened an opportunity to improve message detection that more information can be transferred and even be less conspicuous in transmission such as the microdots technology developed by the Germans. Microdots [16] uses microscopic shrink technique to hide text pictures which can only be read using a microscope. German spies used them in many different ways for instance like messages hidden in letters, on the face of watches and even on spotted ties as shown in [16].

The principle of information hiding was first documented in *On the Criteria to be Used in Decomposing Systems Into Modules* in 1972 [1] whereby Parnas designed a software system and each module's "interface of definition was chosen to reveal as little as possible about its inner workings". Many researchers are trying to carry out research by applying this concept in information hiding. There are three aspects in information hiding systems which contend with each other: capacity, security and robustness [17]. Capacity refers to the amount of information that is able to be hidden in the medium. However, security is important when a secret communication is kept to be secret and undetectable by eavesdroppers. Robustness can be explained as the amount of modification the stego-medium can withstand before an adversary can destroy the hidden information.

1.2 Text Steganography

Figure 2 shows the basic text steganography mechanism. Firstly, a secret message (or an embedded data) will be concealed in a cover-text by applying an embedding algorithm to produce a stego-text. The stego-text will then be transmitted via communication channel, e.g. Internet or mobile device to a receiver. To recover the secret message sent by the sender, the receiver needs to use a recovering algorithm that is parameterised by a stego-key to extract the secret message. A stego-key is used to control the hiding process as well as to restrict detection and/or recovery of the embedded data to parties who know it [18].

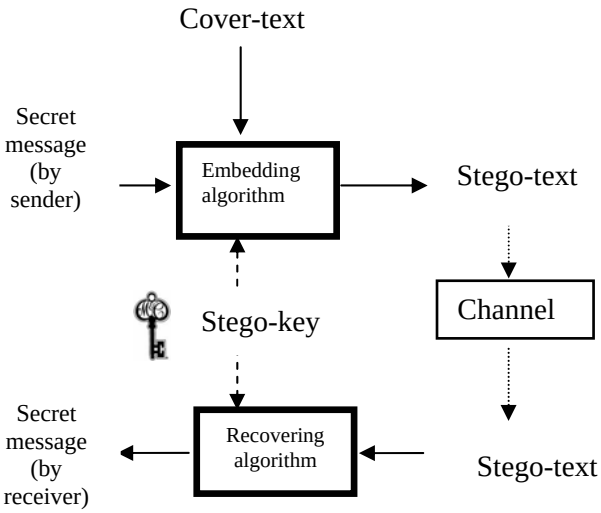


Fig.2: The Mechanism of Text Steganography

The author in [21] has classified the text steganography into three categories - format-based, random and statistical generation and linguistic method.

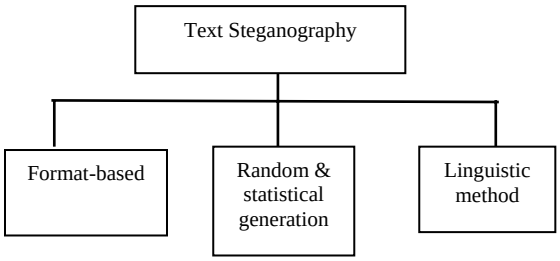


Fig.3: Three basic categories of text steganography

Format-based methods use physical text formatting [19] of text as a place in which to hide information. Generally, this method modifies existing text in order to hide the steganographic text. Insertion of spaces in between words or end of the sentence, deliberate misspellings and resizing of the fonts throughout the text are some of the many format-based methods being used in text steganography. However, Bennett has stated that those format-based methods could not be seen with the human visual system but it is possible to detect with the computer system.

Random and statistical generation is generating cover text according to the statistical properties. This method is based on character sequences and word sequences. The hiding of information within character sequences is embedding the information to be appeared in random sequence of characters. This sequence must appear to be random to anyone who intercepts the message. A second approach to character generation is to take the statistical properties of word-length and letter frequency in order to create "words" (without lexical value) which will appear to have the same statistical properties as actual words in a given language. The hiding of information within word sequences, the actual dictionary items can be used to encode one or more bits of information per word using a code-book of mappings between lexical items and bit sequences, or words themselves can encode the hidden information.

The final category is linguistic method which specifically considers the linguistic properties of generated and modified text, frequently uses linguistic structure [11] as a place for hidden messages. In fact, steganographic data can be hidden within the syntactic structure itself.

1.3 An Overview of Our Proposed Method

In this paper, WhiteSteg is proposed for text steganography by creating a hybrid method in utilising whitespaces between words and paragraphs. This method could be an improvement of open space method [2] because it is not solely using a method of encoding data as what has been mentioned in [2]. By integrating both methods which are inter-word spacing and inter-paragraph spacing into an embedding algorithm, a larger capacity for embedding hidden bits is provided.

The proposed scheme is inspired by Bender's open space method and also a non-commercial used program namely SNOW by Matthew Kwan. Instead of using one method for every embedding mechanism, we propose to create a hybrid method in manipulation of whitespaces so that it is able to hide the secret bits in a dynamic generated cover text to produce a seemingly innocent stego-text.

This paper is divided into three sections. In the first section, it provides an overview for information hiding and text steganography and discusses the new proposed approach. The second section analyzes this hybrid scheme by identifying the improvements that have been made over other approaches of text steganography. In the last section, we draw some conclusions on the proposed approach and also discuss about the future enhancement.

2 Related Work

Peter Wayner proposed a mimicry algorithm [14] that aimed at text steganography in his book Mimic Functions, Cryptologia XVI-3. His approach is to produce mimicked text that looks similar to the real structure of the original text. Peter Wayner used a set of grammatical rules to generate stegotext and the choice of each word determines how secret message bits are encoded. The grammatical rules are based on static grammar which means the grammars must be designed before the algorithm can be used. Apparently, the algorithm generates context-free structures. The system's user must design a grammar that he wishes the text to mimic [10]. The quality of the resulting stegotext directly depends on the quality of the grammar. The grammar acts as the key for hiding data (refer to Figure 4).

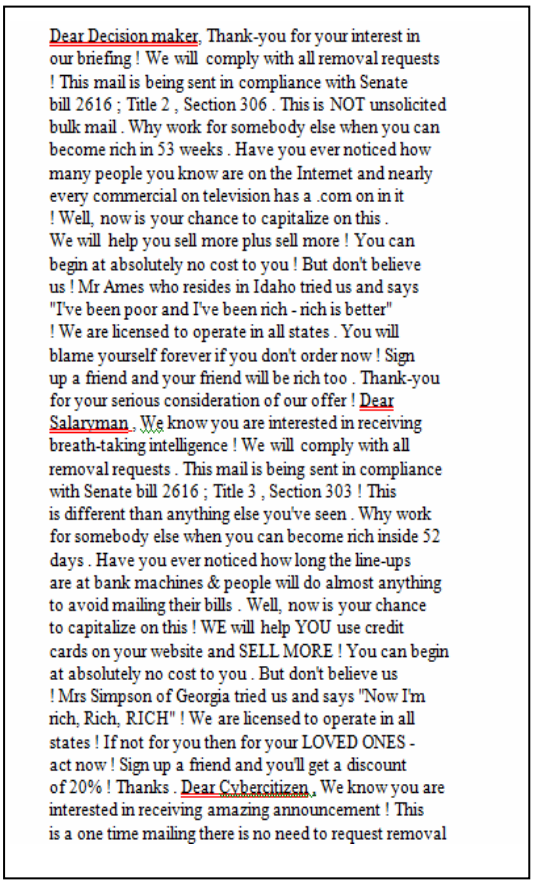
Start → noun verb
Noun → Fred || Barney
Verb → went fishing **where** || went bowling **where**
Where → in **direction** Iowa. || in **direction** Minnesota.
Direction → northern || southern

Fig.4: Example of Context-Free Grammar

In terms of advantage [21], Tenanbaum explained that the context of free grammar approach often produces excellent results because

the grammars always appear to be correct as the algorithm uses a predesigned grammar. One of the major drawbacks will be when both the sender and receiver must have a copy of the grammar to encode and decode the secret message. In fact, Tenenbaum proposed an improvement over the pure context-free grammar method based on mimicry concept. It differs slightly from Wayner's theory as the method does not limit to a single static grammar in order to provide a more flexible framework for grammar based mimicry. Richard Bergmair commented in [22] where the method of Wayner's mimicry completely disregards semantic aspects of language. Data is not hidden in linguistic ambiguity, but in semantically significant parts of language.

Spammimic [12] is another website demonstrating mimicry. This implementation of Wayner's system [9] employs a grammar that mimics the appearance of spam. Based on the author, the advantage of this application is a secret message that can be encoded into an innocent-looking spam message where no one would notice that there is a secret message concealed in it.



Dear Decision maker, Thank-you for your interest in our briefing ! We will comply with all removal requests ! This mail is being sent in compliance with Senate bill 2616 ; Title 2 , Section 306 . This is NOT unsolicited bulk mail . Why work for somebody else when you can become rich in 53 weeks . Have you ever noticed how many people you know are on the Internet and nearly every commercial on television has a .com on in it ! Well, now is your chance to capitalize on this . We will help you sell more plus sell more ! You can begin at absolutely no cost to you ! But don't believe us ! Mr Ames who resides in Idaho tried us and says "I've been poor and I've been rich - rich is better" ! We are licensed to operate in all states . You will blame yourself forever if you don't order now ! Sign up a friend and your friend will be rich too . Thank-you for your serious consideration of our offer ! Dear Salaryman , We know you are interested in receiving breath-taking intelligence ! We will comply with all removal requests . This mail is being sent in compliance with Senate bill 2616 ; Title 3 , Section 303 ! This is different than anything else you've seen . Why work for somebody else when you can become rich inside 52 days . Have you ever noticed how long the line-ups are at bank machines & people will do almost anything to avoid mailing their bills . Well, now is your chance to capitalize on this ! WE will help YOU use credit cards on your website and SELL MORE ! You can begin at absolutely no cost to you . But don't believe us ! Mrs Simpson of Georgia tried us and says "Now I'm rich, Rich, RICH" ! We are licensed to operate in all states ! If not for you then for your LOVED ONES - act now ! Sign up a friend and you'll get a discount of 20% ! Thanks . Dear Cybercitizen , We know you are interested in receiving amazing announcement ! This is a one time mailing there is no need to request removal

Fig.5: Context free generated stego-text using spammimic

Figure 5 is a set of stego-text generated by spammimic after a secret message ("Meet me today for an urgent meeting.") is embedded. The content of the generated text is mainly a commercial publicity for some business scheme. Underlined are the greeting gestures which are repeatedly generating in the cover text that could cause suspicion and it is exposed to attacks such as chosen stego attack and stego-only attack. Chosen stego attack is when the steganography algorithm (tool) and stego-object is defined as only the stego-object is available for analysis [36].

In fact, WhiteSteg keeps this advantage but changes the content of the cover-text which is the lyric extracted from simple and innocent nursery rhymes. The strength of generating nursery rhymes in various length of cover-text is because the chorus of the lyrics could be repeated many times without arousing suspicion.

Brassil et al. [4][5][6] gave the initial idea of document coding methods in his paper by proposing line-shift coding, word-shift coding (Figure 6) and feature coding (character coding) to discourage illicit dissemination of document distributed by computer network. Line-shift coding is a method of altering a document by vertically shifting the locations of text lines to uniquely encode the document. Word-shift coding is a method of altering a document by horizontally shifting the locations of words within text lines to uniquely encode the document. Character coding or feature specific coding is a coding method that is applied only on the bitmap image of the document and could be examined for chosen character features, and those features are altered, or not altered, depending on the codeword. A document is marked in an indiscernible way by a codeword identifying the registered owner to whom the document is sent. If a document copy is suspected to have been illicitly disseminated, that copy can be decoded and the registered owner identified. Brassil's work contributes greatly in deterring illicit dissemination of unauthorized electronic publication. Figure 6 indicates the difference of the spacing before and after word-shift coding from an example.

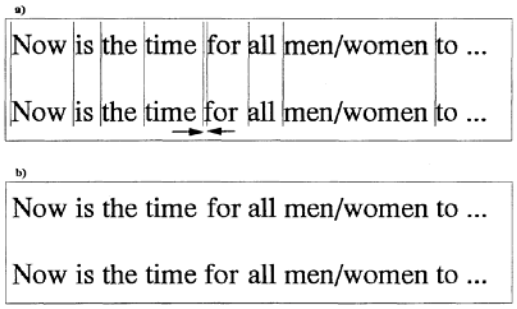


Fig. 6: Example of word-shift coding. **a)** the top text line has added spacing before the "for"; **b)** the original line spacing (adapted from [29])

Bender et al. [2] considered three methods of encoding data when they carried out a study on the techniques for data hiding in text. The three methods are open space methods that encode through manipulation of whitespace, syntactic methods that utilize punctuation and semantic methods that encode using manipulation of the synonymous words. Open space methods derive inter-sentence spacing, end-of-line spaces and inter-word spacing in justified text. Inter-sentence spacing method is to encode a binary message into a text by placing either one or two spaces at the end of each termination character. It encodes a "0" by adding a single space and encodes "1" by adding two spaces. This method works fine but the only disadvantage is its inefficiency because it requires a large amount of text to encode very few bits. One bit per sentence equates to a data rate of approximately one bit per 160 bytes assuming sentences are on average two 80-character lines of text. This method fully depends on the structure of the text. End-of–line space method exploits whitespaces at the end of each line. Data is 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. It works better than the inter-space method because by increasing the number of spaces, more data can be hidden. The third method of using whitespaces to encode data involves the right-justification of text which can also be used to encode data within text files. Data is encoded by controlling where the extra spaces are placed. One space between two words is interpreted as a "0". Two spaces are interpreted as a "1". It is found that not every inter-word space can be used as data due to the constraints upon justification. Bender et al. employed a Machester-like encoding method to determine which of the inter-word spaces represent hidden data bits and which is part of the original text. "01" is interpreted as "1" and "10" as "0". The bit strings "00" and "11" are null. For example, the encoded message in a binary form "1000011110" is reduced to "010," while "110011" is a null string.

Information hiding in whitespaces has inspired Australian programmer, Matthew Kwan to invent a non-commercial use program called SNOW [23]. The encoding scheme used by SNOW relies on the fact that spaces and tabs (known as whitespace) invisibly appear at the end of lines. Since the trailing spaces and tabs occasionally occur naturally, their existence should not be sufficient to immediately alert an observer who stumbles across them.

The SNOW program runs in two modes, first is message concealment and the second one is message extraction. In fact there are other optional modes such as built-in compression and encryption. Similarly, the extraction part will be reversing the process, i.e. extracting data from text, with optional decryption and decompression. Besides, this program is able to inform the user on how much data it can fit in the cover file.

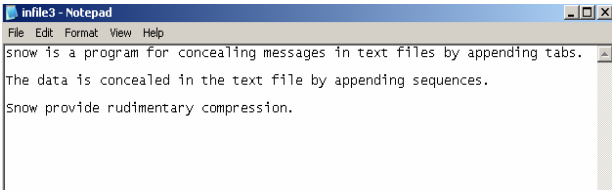


Fig.7: An example of the contents in a cover file

By issuing the command to calculate the storage capacity, it could indicate the estimation of available bits so that we know whether the embedded data could fit the cover file. Therefore, Figure 7 is an example that indicates that it can store between 68 bits to 74 bits of data when the storage calculation command is issued. The size of the embedded data could be fitted into the cover text when it is less that the available storage capacity.

In the embedding scheme, 3 bits are usually coded in 8 columns of text and given the default line length is 80 characters. This allows 30 bits to be stored on empty lines. If a message could not fit into the available cover text, empty lines will be appended and used to accommodate the overflow. The disadvantage of SNOW is the embedded message could be easily destroyed by simply removing the extra spaces at the end of the line.

Hiding information in hypertext is an evolution of character coding and semantic methods whereby

this method modifies the written states of the markup letters based on the fact that there is no request of capital letters or small letters for the markup letters. The Chinese researchers [8] analyzed the structure of hypertext and proposed a more secure method of manipulating the hypertext. The strength of using hypertext in information hiding is that there is no changes and no unwanted symbols in embedded file containing secret information, thus this does not increase the length of the file. In fact, there is a drawback in this method. The inner workings of the coding might be revealed as the source of the page could be easily viewed. For this, our proposed method will never reveal the inner workings of the embedded system as no one will know about the hidden data.

Chen Chao et al. expanded the idea of using TeX, a popular typesetting tool created by Donald E. Knuth [8]. This tool is particularly useful in generating scientific and technical documents with professional page layout. Data hiding without changing the output text can be achieved by using the control sequence ***\begin{equation}\label{eq14}*** represent 0 whereas ***\begin{equation} \label{eq14}*** represent 1. Note the space between ***{equation} and \label{eq14}*** in the latter case [8]. The source file is transmitted to pass the secret information which could be extracted directly by checking whether the space is present. According to [8], another different approach is to slightly modify inter-word spaces by using TeX commands to carry hidden bits. With a **\hspace** command, an inter-word space can be widen or reduced for keeping it unchanged to encode the text format. The embedded data are extracted from the document image obtained from the generated pdf file. Besides, the Chinese researchers proposed that inter-word spaces in a text document can be modified to carry secret information. Each pair of consecutive words carry one bit of secret data. The two words before and after a full stop are grouped and treated as a single word. Then all spaces within a line are approximately equal. With the stego-encoding implementation, space changes are scattered, leading to stealth improvement of the stego-text.

Shirali Shareza has implemented an improved method for hiding information in images [35] using mobile phones. His new approach is to select pixels according to a password. The password would select the pixels at a random manner. In the usual steganographic algorithm, if the size of the information is smaller than the size of the image, the pattern of the altered pixels will be easily identified by an attacker. In fact, the author's approach could solve the problem by putting the pixel in the secret message in random order in each block so that it will be difficult to extract the hidden information. However, the author had identified the major drawback which is by changing the image will cause degradation of the stego-image and the loss of data cannot be recovered.

According to [27] and [28], the Iranian researchers came up with a new method for secret exchange of information through chat and SMS respectively by using and developing abbreviation text steganography with SMS-Texting language [29]. The authors suggested using this collection of abbreviated words over online chatting and mobile SMS. This word can be a usual word such as hospital which is abbreviated as "hosp." or a word from SMS-Texting collection such as "as soon as possible" abbreviated as "ASAP".

Table 1: List of some SMS acronyms

Acronym	Translation
2l8	Too late
ASAP	As Soon As Possible
C	See
F2F	Face to Face
B4	Before
2	To, too.

The algorithm searches the SMS or chat for the words existing in the list according to the algorithm described. If the number of embedded words found were less than the length of array of zero and one bit which made from the data we want to hide, thus the program will indicate that it cannot hide the data in the given SMS and that the size of information given is too large. Our proposed method provides the flexibility where the length of the secret message is determined before fitting in a cover-text which has the most suitable capacity.

3 WhiteSteg

We are proposing a new approach on hiding information through manipulation of whitespaces between words and paragraph as a hybrid method, namely WhiteSteg. WhiteSteg is able to provide more capacity for hiding more bits of data into a cover-text.

In the literature [2], hiding information within spaces appears to have potential as people can hardly identify the existence of the hidden bits which appear in the whitespaces. Bender et al. had shown 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 Bender's method is that it requires a great deal of space to encode few bits. For example, a character is equivalent of 8 bits [30], and it requires approximately 8 inter-spaces to encode one character. However, this problem can be resolved by compressing the character in the secret message from 8 bits to 3 bits in WhiteSteg. With the inter-paragraph spacing, more whitespaces in the cover text document could be utilized effectively. In the embedding method, spaces can be inserted between a newline character and another newline character. Thus, WhiteSteg is a hybrid scheme of both inter-word and inter-paragraph spacing.

Currently, manipulation of whitespaces seems beneficial and has its potential in information hiding because whitespaces appear in a text documents more than the appearance of words. Thus, this can be an advantage when no one will know that a blank piece of document [35] is actually vital secret information which can be retrieved after a decoding process. The syntax that involve in whitespace manipulation are space (ASCII char 32), tab (ASCII 9) and line feed (ASCII 10). Even though these syntax appear to be invisible, but it is an advantage to utilize these syntax in text steganography, especially appending spaces and manipulating them to hide information. According to [37], the above mentioned method is still significant to be implemented due to the weakness of the human optical system.

The cover-text will be dynamically generated according to the length of the secret message. The maximum capacity of hidden bits is determined by the system whether the length of the secret message can be accommodated in the particular capacity.
A minimum input of text by a user is considered where one character or a few characters as a secret message. Thus, the cover text will be generated in 1 byte, 2 bytes, 4 bytes, 8 bytes, 16 bytes, 32 bytes, 64 bytes, 128 bytes, 256 bytes and 512 bytes. This range of maximum capacity is used for the dynamic generated cover text to obtain the preliminary experimental results. The results will be analyzed in order to determine to what extent of the payload can be assigned to accommodate the secret message so that the frequent occurrences of extra spaces may not alert the potential adversaries.

By adapting the source of the generated text from any lyrics of the nursery rhymes, the stego-text will definitely present an innocuous and naive appearance. Besides, the chorus of the lyrics can be duplicated and reused to generate a longer cover-text. There is a greater advantage in using centered-alignment because the appearance of the spacing occurs naturally without arising even the slightest suspicion, refer to Figure 5.

The significance of this research is by merging the two concepts and creating an algorithm for the data embedding system. The capacity of the cover-text depends on the length of the secret message. Based on Figure 4, a user is required to enter the secret message as provided in the text field or by selecting a file that contains secret message. The system will calculate the length of the message and generate a cover-text which is suitable to encode the secret message. Table 1 indicates the size of the payload (amount of hidden information) which is calculated by the system and assigned to the most appropriate cover-text.

As illustrated in Figure 8, the length of the secret ("Arrive on Friday.") is approximately 17 bytes and it can fit into the cover-text which contains 32 bytes of payload. The secret message is then entered and encoded. Eventually, the stego-text is generated to fit the entire secret message which has been entered by a user, as shown in Figure 9. When another user receives this message during a public communication, the hidden bits can only be retrieved by using the extracting algorithm. Table 2 indicates the range of the maximum capacity for the cover text which can accommodate the secret message.

Table 2: Cover-text is generated according to the length of secret text

Size of the text (Byte)	
Secret	Cover
< 1	1
< 2	2
< 4	4
< 8	8
< 16	16
< 32	32
< 64	64
< 128	128
< 256	256
< 512	512

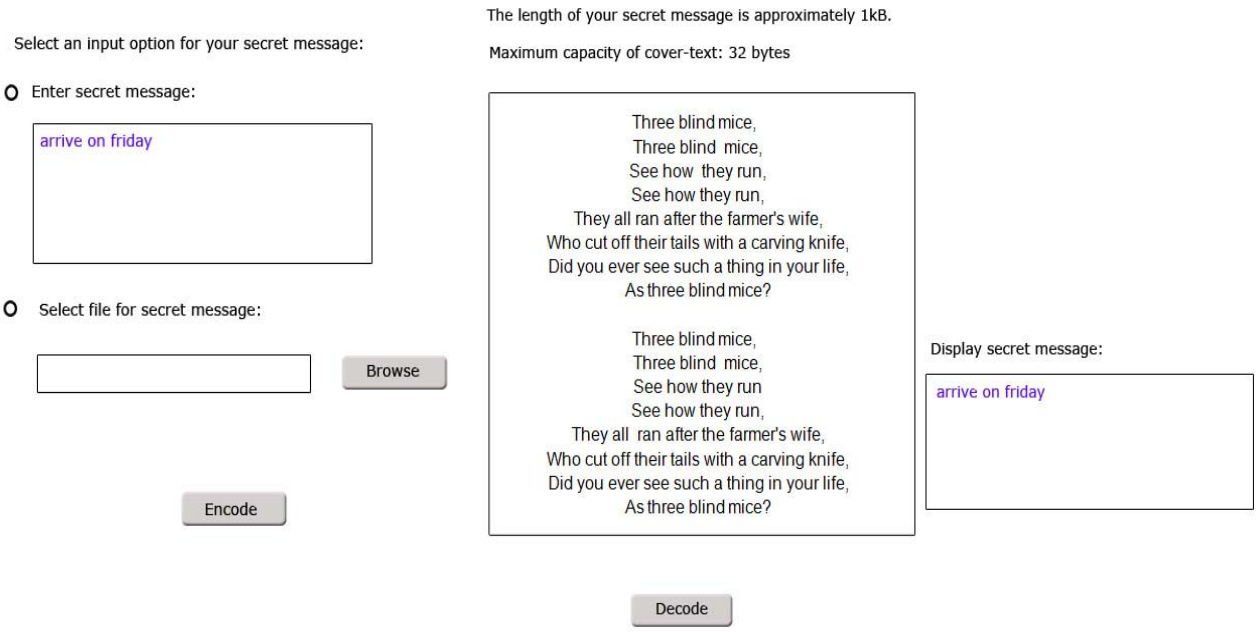


Fig.8: GUI of proposed method (cover-text is adapted from [26])

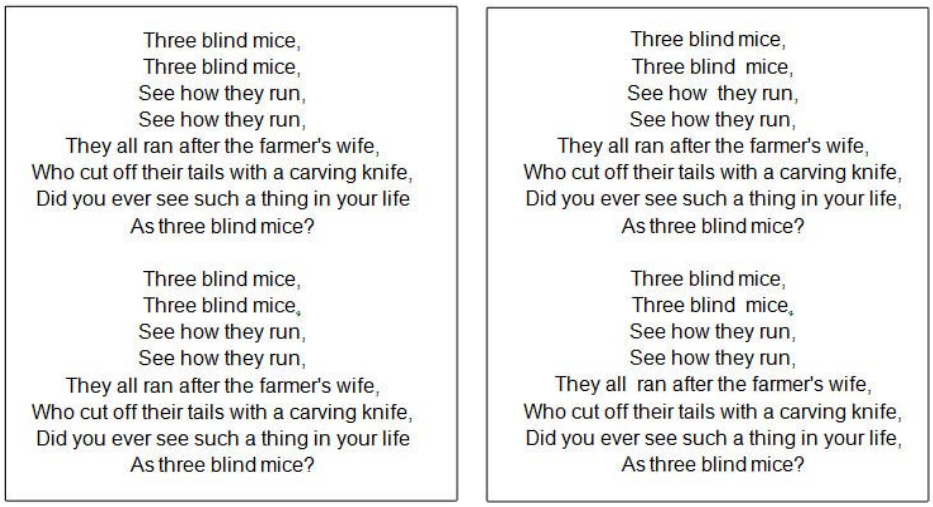


Fig.9: Original cover text (Left), Generated stego-text with hidden data (Right)

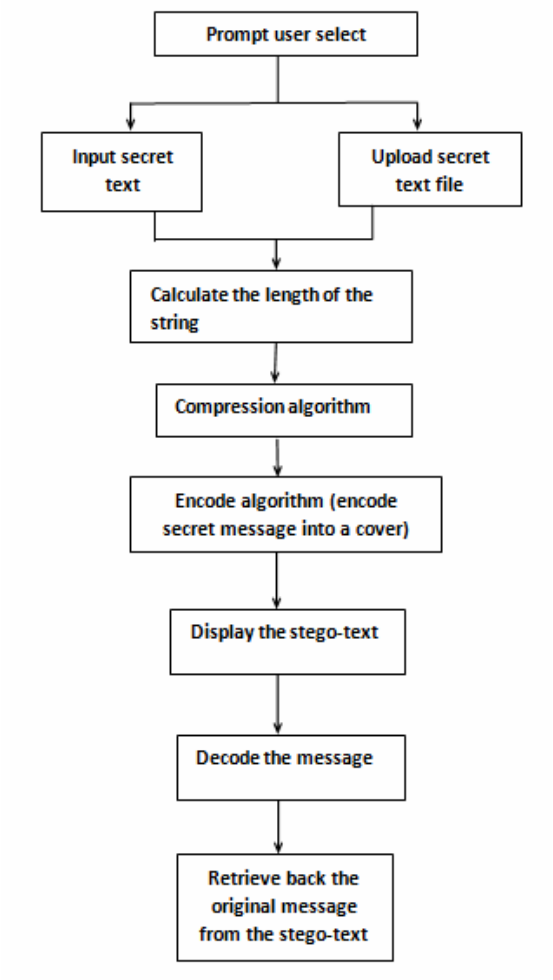


Fig. 10: System flow of WhiteSteg

Figure 10 illustrates the flow of the system in WhiteSteg where the encoding and decoding processes are sequentially revealed.

4 Conclusion and Future Work

A new approach of text steganography method using inter-word and inter-paragraph spacing for hiding information, named WhiteSteg is presented. The unique feature of this method is the system can dynamically generate a suitable cover-text when user enters a secret message. The system will determine which cover-text has the most appropriate capacity to accommodate the secret according to the length of the message.

Nursery rhymes provide flexible repetitive chorus which is an advantage in generating the cover-text. Thus, hiding data in spaces does not attract much attention and this reduces the possibility of detection in the innocent-looking lyrics which contain hidden data.

The future work should focus towards optimizing the robustness of the decoding algorithm. This is because the hidden data will be destroyed once the spaces are deleted by some word processing software. Information hidden in text in the form of appended spaces characters can be revealed by opening the file with a word processor. Extra-spaces and characters can be quickly stripped from text document.

Security wise, encryption should be implemented in the stego-system. This is because even if the embedding and extracting algorithm are discovered, lots of efforts are still needed to decrypt the message.

The range of the payload size can be increased so that more data is able to be embedded in the cover text apart from the range between 1 byte to 512 bytes. Besides, it is important to take other compression methods into consideration. A more detailed research is needed to be carried out on various compression techniques that can be applied in the proposed hybrid algorithm.

5 Acknowledgement

We would like express our gratitude to Dr. Goh Chong Tien for proof reading the paper.

References:

[1] D. Parnas, “On the Criteria to Be Used in Decomposing Systems Into Modules”, *Communication of the ACM*, vol. 15, no. 12, December 1972, pp. 1053-1058.

[2] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, “Techniques for data hiding”, *IBM syst. J.*, vol. 35, nos. 3 – 4, 1996, , pp. 313 – 336.

[3] J. Brassil, S. Low, N. F. Maxemchuk, and L. O'Garman., “*Marking Text Features of Document Images to Deter Illicit Dissemination*”, IEEE, 1994, pp. 315-319.

[4] J. Brassil, S. Low, N. F. Maxemchuk, and L. O'Garman, “*Electronic Marking and Identification Techniques to Discourage Document Copying*”, IEEE, Oct. 1994, pp. 1278-1287.

[5] J. Brassil, S. Low, N. F. Maxemchuk, and L. O'Garman. “Electronic Marking and Identification Techniques to Discourage Document Copying”. *IEEE Journal on Selected Areas in Communications*, Vol. 13, Oct. 1995, pp. 1495-1504.

[6] J. Brassil, S. Low, N. F. Maxemchuk, and L. O'Garman, “Copyright Protection for the Electronic Distribution of Text Documents”, *Proceedings of IEEE*, Vol. 87, July 1999.

[7] J. Brassil, S. Low, N. F. Maxemchuk, and L. O'Garman. "Document Marking and Identification using Both Line and Word Shifting," *Proc. Infocom95, IEEE CS Press*, Los Alamitos, Calif., 1995.

[8] Chen Chao, Wang Shuozhong, Zhang Xinpeng. "Information Hiding In Text Using Typesetting Tools with Stego-Encoding", *ICICIC*, 2006.

[9] P. Wayner, "Mimic functions", *Cryptologia archive*, vol. 16, Issue 3, July 1992, pp.193 – 214.

[10] P. Wayner. "*Strong Theoretical Steganography*". Cryptologia, XIX (3), July 1995, pp. 285-299.

[11] M. Chapman, G. Davida. "*Hiding the Hidden: A Software System for Concealing Ciphertext as Innocuous Text*". Master Thesis, Milwaukee: University of Wisconsin-Milwaukee, 1998.

[12] "Spammimic", 2000. [Online] Available: *http://www.spammimic.com* [Accessed: March 8, 2008].

[13] N. F. Johnson, S. Jajodia, "Exploring Steganography: Seeing the Unseen," *IEEE Computer*, February 1998, pp.26−34.

[14] P. Wayner, Disappearing Cryptography: Being and Nothingness on the Net, *Academic Press*, Inc., 1996.

[15] Richard Bergmair, "Towards Linguistic Steganography: A Systematic Investigation of Approaches", Systems, and Issues, *Technical Report*, Nov 2004.

[16] "Spy Gadgets in World War II: Microdots", 2007. [Online]. Available: *http://www.mi5.gov.uk/output/Page303.html* [Accessed: Feb. 15, 2008].

[17] N. Provos, P. Honeyman, "Hide and Seek: An Introduction to Steganography", *The IEEE Computer Security*, 2003.

[18] Fabien A. P. Petitcolas, Ross J. Anderson and Markus G. Kuhn. "Information Hiding – A Survey", *Proceedings of the IEEE, special issue on protection of multimedia content*, July 1999, pp. 1062 – 1078.

[19] Bret Dunbar, "*A Detailed Look At Steganographic Techniques and Their Use in Open-Systems Environment*", SANS Institute, 2002.

[20] B. Pfiztzmann, "Information Hiding Terminology." pp. 347-350, ISBN 3-540-61996-8, *results of an informal plenary meeting and additional proposals*, 1996.

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

[22] 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].

[23] Matthew Kwan, "The SNOW Home Page", 1998. [Online].Available: *http://www.darkside.com.au/snow/* [Accessed: March 12, 2008].

[24] Sabu M. Thampi. "Information Hiding Techniques: A Tutorial Review", *ISTE-STTP on Betwork Security & Cryptography*, LBSCE 2004.

[25] M. Chapman, G. Davida, and M. Rennhard, "A Practical and Effective Approach to Large-Scale Automated Linguistic Steganography", *Proceedings of the Information Security Conference*, October 2001, pp. 156-165.

[26] Nursery Rhymes - lyrics and origins. [Online].Available:*http://www.famousquotes.me.uk/nursery_rhymes/nursery_rhymes_index.htm* [Accessed: March 31, 2008]

[27] M. Hassan Shirali-Shahreza, Mohammad Shirali-Shahreza. "Text Steganography in Chat", IEEE, 2007.

[28] M. Hassan Shirali-Shahreza, Mohammad Shirali-Shahreza. "Text Steganography in SMS", International Conference on Convergence Information Technology, 2007.

[29] K. Beare, "SMS-Texting", English as 2nd Language. [Online]. Available: *http://esl.about.com/* [Accessed: March 10 2008].

[30] "Bits, Bytes and Bandwidth Reference Guide". [Online]. Available: *http://www.speedguide.net/read_articles.php?id=115* [Accessed: 22 April 2008].

[31] Osamu Takizawa, Akihiro Yamamura, Hiroshi Nakagawa, Tsutomu Matsumoto, Ichiro Murase, Kyoko Makino, Shingo Inoue and Hiroyuki Ohno, "A Proposal of Steganography on Plain Text and XML", *The 1st NLP and XML Workshop*, Nov 2001.

[32] Chen Chao, Wang Shuozhong, Zhang Xinpeng. "Information Hiding In Text Using Typesetting Tools with Stego-Encoding", *ICICIC*, 2006.

[33] Durham University Computing Society, "Whitespace". [Online]. Available: *http://compsoc.dur.ac.uk/whitespace/* [Accessed: April 13, 2008].

[34] Neil F. Johnson and Sushil Jajodia, "Steganalysis: The Investigation of Hidden Information", *Proceedings of the 1998 IEEE Information Technology Conference,* USA, 1998, pp. 113-116.

[35] Mohammad Shirali Shareza, "An Improved Method for Steganography on Mobile Phone", *Proceedings of the 9th WSEAS International Conference on Systems*, Greece, 2005, pp. 955-957.

[36] Klimis S.Ntalianis, "A Short-Message Robust Steganographic Method for Effective Information Recovery Under Transmission Losses of Cellular Networks", *Proceedings of the 9th WSEAS International Conference on Systems*, Greece, 2005, pp. 955-957.

[37] J. J. Liaw, L. H. Chang, Y. S. Liao, "An Improvement of Robust and Blind Data Hiding Based on Self Reference in Spatial Domain", *Proceedings of the 2007 WSEAS International Conference on Computer Engineering and Applications,* Gold Coast, Australia, 2007, pp. 259-263.