Developing an Efficient Solution to Information Hiding through Text Steganography along with Cryptography

Md. Palash Uddin[1], Mousumi Saha[2], Syeda Jannatul Ferdousi[3], Masud Ibn Afjal[4], Md. Abu Marjan[5]
Faculty of Computer Science and Engineering, Hajee Mohammad Danesh Science and Technology University (HSTU), Dinajpur-5200, Bangladesh
[1]palash_cse@hstu.ac.bd, [2]mousumi.saha17@yahoo.com, [3]jannatulferdous.tee@gmail.com,
[4]masud@hstu.ac.bd,[5]mdabumarjan66@gmail.com

Abstract—**Now-a-days information security over unsecured communication channel is one of the most challenging issues. Cryptography and steganography, a sub-discipline of information hiding, are two distinct popular security offering techniques. To get the combined chemistry from these techniques, today's real-time security systems include both cryptography and steganography in their working principle. To cope up with those systems, we have proposed a format-based pure text steganography algorithm including a private key cryptography providing a higher level of security. The cover text has been made as ordinary as possible. After successful embedding of the secret message into the cover text, the stego-text also looks like an ordinary text because only an alphanumeric puzzle is added at the end of the cover text to make up the stego-text. This final stego-text is sent to the receiver through the unsecured communication channel. Instead of correct recipient, if any third party or cracker or hacker obtains the stego-text, they may think that it is an ordinary text to teach English to the kids since we made it like that or if they try to extract the original message, it requires a huge amount of computational time. The proposed algorithm supersedes various text steganographic techniques through combining both cryptography and steganography.**

Keywords— Information Hiding, Text Steganography, Concealing of message, Embedding and Extraction, Higher Level of Security.

I. INTRODUCTION

Increasing the development of the communication sectors the higher level of security of information has become one of the basic needs for transmitting information over the unsecured communication channel. To protect information from the crackers or hackers broadly two types of security techniques are used. These are cryptography [1] and information hiding [2, 17]. Cryptography is a technique where the existence of the message is not concealed although the format of the message is untraceable by the eavesdroppers. Cipher is a pair of algorithm which creates the encryption and decryption.There are mainly three types of cryptographic algorithm. These are private key cryptography such as advanced data encryption (AES) [3, 19], data encryption standard (DES) [4, 18] etc., public key cryptography such as RSA algorithm [5], and hash algorithms [6]. The major security goals of cryptography are confidentiality, authentication, integrity, non-repudiation, and access control [7]. Often, the cryptography methods cannot provide the total security of information transferred over the unsecured communication channel. Hence, the information hiding techniques are required. Information hiding methods conceals the existence of the secret message in the communication channel. The most common information hiding methods are steganography, watermarking, anonymity and covert channel. Steganography [8] is the art or practice of concealing an image, file, message within another image, file or message. In early days various techniqueswere used in steganography such as pin punctures, character marking, invisible ink, typewriter correction ribbon etc. The traditional types of steganography are text, image and audio steganography.Text steganography is used to hide text in another text. It is a complex form because the redundant amount of text to hide the secret message is scare in text files. It is classified into basic three types [9] - format-based, random and statistical generation and linguistic method. Image steganography uses the Human Visual System (HVS). Audio steganography is one of the difficult forms of steganography as humans are able to detect a minute change in the quantity of audio.In this paper, a new pure text steganographic method along with DES cryptography has been proposed which is shown in Fig. 1 in brief.

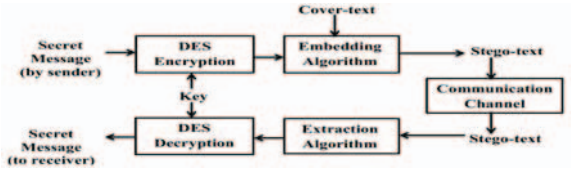

Figure 1.Text Steganography along with DES cryptography

II. RELATED WORKS

Several works have been done regarding information hiding through text steganography. Some researchers conducted a survey on two security tools- cryptography and steganography [10]. A text steganography solution combining word shifting, text steganography and synonym text steganography has been proposed by some researchers [11]. Some researchers tried to employ a set of all synonyms as a way to hide secret message inside a natural language text [12]. Some authors made a review study on text steganography [13]. Some researchers made a comparative study among various steganographic method based on the capacity and security level [14]. Some researchersproposed

978-1-4799-6062-0/14/$31.00©2014 IEEE

a text steganography algorithm based on font type in Ms-Word documents [15]. Some researchers made a discussion on various types of information hiding methods specially of steganography and also proposed for text steganography by creating a hybrid method in utilizing whitespaces between words and paragraphs in right-justification of text [16]. Some researchers proposed a text steganography approach based on genetic algorithm [20].

III. PROPOSED TECHNIQUE

The proposed technique for information hiding through the format-based pure text steganography is divided into a cascade of two processes. They are the preparation of the ordinary cover text and the embedding process. Embedding the secret message into the cover text the stego-text is transmitted over the unsecured communication channel. At the intended receiver the extraction process applied to extract the secret message sent by the sender is also discussed here.

A. Preparation of Cover Text

The cover text has been made explicitly in such a way that it looks like an ordinary text consisting of all the English characters which are mostly used to teach English to the kids. The cover text has been treated as a linear array of 224 English characters stored from position 0 to 223.

B. Embedding Process

First, the embedding process takes the prepared cover text which is treated as a linear array of English characters. Secondly, taking the original message the process encrypts it by DES encryption with a secret key of 56 bit. Then counting the frequency with respective positions of each encrypted character in the ciphertext, the algorithm retrieves the position of each encrypted character comparing with the characters of the cover text. Treating each of the positions as an ASCII it takes the equivalent character of each ASCII and thus the algorithm conceals each encrypted character to this newly retrieved character from the cover text. Finally, an alphanumeric puzzle using the counted frequency with respective positions of each character in the ciphertext and the characters, which conceal the encrypted characters, is added at the end of the cover text. In the puzzle, each finally retrieved character is placed directly. Each position guided by the frequency count of each encrypted character in the ciphertext is split into two integers so that the position can be represented by some arithmetic operations and then the operations are added in the puzzle section. The stego-text is sent to the receiver over the unsecured communication channel.

1) Embedding Algorithm

The pseudo-code of the embedding algorithm is illustrated below:

i. Input the cover text
ii. Input the original message to be embedded
iii. Encrypt the original message with DES encryption
iv. Count the frequency with respective positions of each character in the ciphertext
v. Retrieve the position of each encrypted character comparing with the characters of the cover text and treating each of the positions as an ASCII take the equivalent character of each ASCII
vi. Make an alphanumeric puzzle:
    a. Spilt each position guided by the frequency count into two integers
    b. Directly place the final retrieved character and place some arithmetic operations performing by the two split integers
vii. Add the puzzle at the end of the cover text

C. Extracting Process

In the extraction phase, first the counted frequency with respective positions of each character in the ciphertext and the characters, which conceal the encrypted characters, are retrieved from the alphanumeric puzzle of the stego-text. Taking the equivalent ASCII value of the characters from the puzzle the characters stored in the positions equivalent to the ASCII are retrieved. Then, each of the characters is placed to those positions mentioned by the frequency count with respective positions in the ciphertext. Then the DES decryption is applied to the retrieved ciphertext to get the original message sent by the sender.

1) Extracting Algorithm

The pseudo-code of the extracting algorithm is illustrated below:

i. Input the stego-text
ii. Input the alphanumeric puzzle from the stego–text and do:
    a. Retrieving the character from the puzzle take the equivalent ASCII of each character
    b. Retrieve the character from the text of the position equivalent to the calculated ASCII
iii. Retrieve the frequency and associated positions of each character from the puzzle and combine all the characters according to the frequency and positions
iv. Decrypt the combined text with DES decryption

IV. EXPLANATION WITH EXAMPLE

A. Embedding

The embedding process of a typical data using the pseudo-code of the developed algorithm is explained below:

1) Input the cover text

The cover text consists of a title, digits, lowercase letters, uppercase letters, punctuations, operators and special characters. For the developed text steganographic algorithm the ordinary cover text is shown in Fig. 2. The linear array representation of the cover text is shown in Table I.

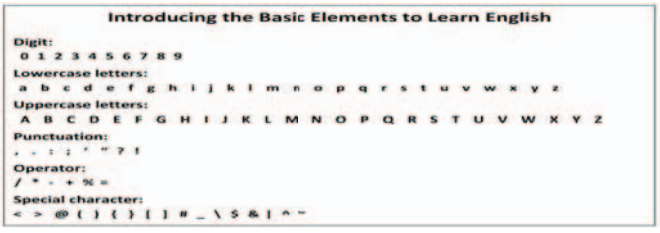

Figure 2. Cover text

TABLE I. OVERTEXT AS A LINEAR ARRAY

Character	Position
I	0
n	1
.	.
0	53
.	.
a	81
.	.
~	223

2) Input the original message to be embedded

Let the original message to be embedded be P = First Australia Next Japan.

3) Encrypt the original message with DES encryption

After performing DES encryption using a 16-digit hexadecimal key, K=AD0B6594EC1320DF the ciphertext is C=0x134ba6386e73d2dc76e546e5076bea6da90dce72296c4c42de6a3472464b01f9. The ciphertext is also treated as a linear array and stored as follows:

TABLE II. CIPHERTEXT AS A LINEAR ARRAY

Character	Position
0	0
x	1
.	.
9	65

4) Count the frequency with respective positions of each character in the ciphertext

Now, the algorithm counts the frequency of each character in the ciphertext and also stores the positions as in Table III.

5) Retrieve the position of each encrypted character comparing with the characters of the covertext and treating each of the positions as an ASCII take the equivalent character of each ASCII

The algorithm retrieves the position of each encrypted character comparing with the characters of the covertext and then treating each of the positions as an ASCII it takes the equivalent character of each ASCII. For example, if the encrypted character is 'x', the algorithm retrieves the position of 'x' comparing with the characters of the covertext as 104. Then the position 104 is converted to its equivalent ASCII which is 'h'. For our example, this is illustrated in Table III.

TABLE III. FREQUENCY, POSITIONS AND RETRIEVING POSITION AND TAKING THE EQUIVALENT CHARACTER TREATING THE POSITION AS AN ASCII

Character in ciphertext	Frequency	Positions against frequency	Equivalent position in cover text	Equivalent ASCII of the position
0	4	0, 26, 36, 62	53	5
X	1	1	104	h
.	.	.	.	.
9	3	35, 43, 65	62	>

So a new representation of the ciphertext is 5h689RQ;8=;U28T7TS<;U:9;U:5<;RUQ;TQ>5TSU<77>;S9S97TU;Q89<79;9R56V>

6) Make an alphanumeric puzzle

Now, the algorithm generated the puzzle in which the first part holds the characters that conceal the message and the second part holds the position of the characters. In the second part the position is represented by several arithmetic operations where the output of the arithmetic operation denotes the position of the character and the number of the arithmetic operation is equal to the frequency of each equivalent character. This is shown in Table IV.

7) Add the puzzle at the end of the cover text

Finally, the algorithm adds the generated puzzle at the end of the cover text. Thus, the stego-text also looks like an ordinary text and it is sent to the receiver.

TABLE IV. ALPHANUMERIC PUZZLE

Characters that conceal the message	Arithmetic operations denoting the frequency with respective positions
5	0+0=0, 23+3=26, 6*6=36, 64-2=62
h	1+0=1
.	.
V	66-2=64

B. Extraction

The extraction process from the stego-text using the pseudo-code of the developed algorithm is explained below:

1) Input the stego-text

In the extraction phase, the algorithm first takes the stego-text as the input.

2) Input the alphanumeric puzzle from the stego–text and do:

First, the algorithm takes the first part, consisting of only characters, of the alphanumeric puzzle from the stego-text. Then, the algorithm takes the equivalent ASCII value of each character. It finds the encrypted and concealed character of that position from the stego-text which is equal to the calculated ASCII. In our example, if the retrieved character is '5', its equivalent ASCII value is 53 and thus '5' is replaced by '0', which was concealed, because '0' is placed at position 53 in the stego-text. For the considered example, this is shown in Table V.

3) Retrieve the frequency and associated positions of each character from the puzzle and combine all the characters according to the frequency and positions

Now, the algorithm takes the second part of puzzle section which contains the arithmetic operations. The output of each arithmetic operation in each row indicates the positions of each finally retrieved character for the same row. Then, the algorithm places the character in those positions. For example, the output of 0+0=0, 23+3=26, 6*6=36, 64-2=62 contains the position 0, 26, 36 and 62 for character '5' and '5' is equivalent to '0' as by step 2. So, '0' is placed in positions 0, 26, 36 and 62. This is shown in Table V.

TABLE V. RETRIEVING OF FREQUENCY, ASSOCIATED POSITIONS AND THE FINAL CHARACTER

Puzzled character	5	h	6	……	V
Equivalent ASCII of the character	53	104	54	…… ….	86
Retrieved Character	0	x	1	……	f
Arithmetic Operation	0+0=0, 23+3=26, 6*6=36, 64-2=62	1+0=1	2+0=2, 66-3=63	……	66-2=64
Frequency	4	1	2	……	1
Positions against frequency	0, 26, 36, 62	1	2, 63	……	64

Combining the finally retrieved characters according to their frequency and respective positions the text looks likeC, which is undoubtedly same as the ciphertext performed in the embedding process.

4) Decrypt the combined text with DES decryption

Finally, the DES decryption with the same key as used in the encryption phase is applied to the combined text to get the original message, P = First Australia Next Japan.

V. Experimental Result and Discussion

The performance study with the results of the experiments of the proposed method has been presented here. All experiments are executed on a 2.30 GHz Inter(R) Core™ i3-2350M CPU with 2GB RAM running Windows 7 Ultimate. For both data hiding and data extracting the algorithm has been implemented by a simple JAVA program. The proposed algorithm has also been evaluated for its accuracy using different sets of sample data. The comparison of the developed steganography with other methodsis shown in Table VI. Thus it is clear that the steganographic model offers a higher level of security. In most of the existing methods the size of the cover text is very bulky. But for our proposed method the comparative size of the cover text is very smaller as it is not system generated and interactively very simple.Thus, it ensures high transfer speed.

TABLE VI. Comparison with Other Methods

	Proposed Method	**GATS [20]**	**wbStego [21]**	**SNOW [22]**	**Stego [23]**
Use of encryption/ decryption key	Yes	Yes	Yes/No	Yes/No	Yes
Cover file	Not system generated but simple and interactive	System generated	Not system generated	Not system generated	Not system generated
File types	.txt	.txt	Image, pdf, txt	-	-
Visibility of secret message	Not visible	Not visible	Not visible	Visible	Not visible
Type of Encryption	DES-Data Encryption Standard with 64 bit key	Playfair	Various	ICE-Information Concealment Engine with 64 bit key	-

Moreover, none of the cover text or the steganographic algorithm will not be easily available for various steganographic attacks like the known carrier attack, known message attack, steganography only attack , known steganography attack, or statistical attack. Hackers or crackers can proceed to extract the secret message with only the stego-text, but it requires a huge computational time.

However, the developed algorithm can be applied in various security systems such as E-mail communication system, cloud-based system, banking security system, mobile communication system, key or password management system, administrative security system, network security system, etc.

VI. Conclusion

Format based text steganography has become one of the important methods in hiding information on the cover text because it uses a predefined cover text to hide information. Thus, we have developed a format based pure text steganographic approach which adds extra security in data transfer. The result of this method proved that it is more secure when it is compared with the existing methods. In future, we will try to implement some real time security systems using this algorithm. Besides the method was successful in extracting and retrieving the hidden secret massage out of the text. Finally, it can be concluded that the results we have obtained are promising and thus given us high inspiration to carry out a research in this area.

References

[1] E. Cole, R. Krutz and J. W. Conley, *Network Security Bible*, Wiley Publishing Inc., 2005.

[2] 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.

[3] J. Dray, Report on the NIST Java AES Candidate Algorithm Analysis, [Online]. Available: http://csrc.nist.gov/ encryption/aes/round/r1-java.pdf.

[4] S. Kumar, Addagarla, and Y. Babji, A Comparative Security Study Review on Symmetric Key Cryptosystem Based Algorithms, *International Journal of Computer Science and Mobile Computing*, vol. 2, no.7, pp.146– 151, 2013.

[5] C. P. Pfleeger and S. L. Pfleeger, *Security in Computing*, 3rd Edition, Prentice-Hall, 2003.

[6] A. H. Omari, B. M. Al-Kasasbeh, R. E. Al-Qutaish, and M. I. Muhairat, A New Cryptographic Algorithm for the Real Time Applications, *Proc. of the 7th WSEAS International Conference on information security and privacy (ISP)*, pp. 33-38, 2008.

[7] D. Stinson, *Cryptography Theory and Practice*, CRC Press Inc., NY, USA, 1995.

[8] M. Al-Mualla and H. Al-Ahmad, Information Hiding: Steganography and Watermarking, [Online]. Available: http://www.emirates.org/ieee/information_hiding.pdf.

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

[10] A. J. Raphael and V. Sundaram, Cryptography and Steganography – A Survey, *Int. J. Comp. Tech. Appl.*, Vol. 2 (3), pp. 626-630.

[11] S. R. Govada, B.S. Kumar, M. Devarakonda and M.J. Stephen, Text Steganography with Multi level Shielding, *International Journal of Computer Science Issues*, Vol. 9, Issue 4, No 3, July 2012, pp. 401-405.

[12] A. Alabish, A. Goweder, and A. Enakoa, A Universal Lexical Steganography Technique, *International Journal of Computer and Communication Engineering*, Vol. 2, No. 2, March 2013, pp.153-157.

[13] N. Rani and J. Chaudhary, Text Steganography Techniques: A Review, *International Journal of Engineering Trends and Technology (IJETT),* Volume 4 Issue 7- July 2013, pp. 3013-3015.

[14] A. K. Hmood, H. A. Jalab, Z.M. Kasirun, B.B. Zaidan and A.A. Zaidan, On the Capacity and Security of Steganography Approaches: An Overview, *Journal of Applied Sciences*, Vol. 10(16), pp. 1825-1833, 2010.

[15] W. Bhaya, A. M. Rahma and D. AL-Nasrawi, (2013), Text Steganography Based on Font Type in Ms-Word Documents, *Journal of Computer Science*, Vol. 9(7), pp. 898-904.

[16] L. Y. POR and B. Delina, Information Hiding: A New Approach in Text Steganography, *7th WSEAS Int. Conf. on Applied Computer & Applied Computational Science (ACACOS)*, Hangzhou, China, April 6-8, 2008, pp. 689-695

[17] W. Bender, D. Gruhl, N. Morimoto and A. Lu, Techniques for data hiding, *IBM Systems Journal*, vol. 35, nos. 3&4, 1996.

[18] G. Blelloch, Introduction to Cryptography, [Online]. Available: www-2.cs.cmu.edu/afs/cs/project/pscicoguyb/realworld/crypto.ps.

[19] D. Baudran, H. Gilbert, L. Granboulan, H. Handschun, A. Joux, P. Nguyae, F. Noilhan, O. Poincheva, T. Pornin, G. Poupard, J. Stern and S. Vaudenay, Report on the AES Candidates, *Proc. of the 2nd ASE Conference*, Rome, Italy, 1999.

[20] C. K Mulunda, P. W .Wagacha and A. O. Adede, Genetic Algorithm Based Model in Text Steganography, *The African Journal of Information Systems*, Vol. 2 Issue 5, pp. 131-144, October 2013.

[21] Welcome to wbStego Steganography Tool, [Online], Available: http://wbstego.wbailer.com/

[22] M. Kwan, The SNOW Home Page, [Online], Available: http://www.darkside.com.au/snow/

[23] J. Walker, Stego! Text Steganography, [Online], Available: http://www.fourmilab.ch/javascrypt/stego.html