International Journal of Soft Computing 12 (1): 20-25, 2017
ISSN: 1816-9503

A Novel Algorithm for Text Steganography

[1]Rehman Ullah Khan, [2]Muh. InamUlHaq, [3]YahyaKhan, [1]Oon Yin Bee,
[1]Shahren Ahmad ZadiAdruce, [1]Mai S. Ishak and [1]Tan Kock Wah
[1]Faculty of Cognitive Sciences and Human Development, Universiti Malaysia, Sarawak, Malaysia
[2]Department of Computer Sciences, Khushal Khan Khattak University, Karak, Pakistan
[3]Institute of Computing and Information Technology Gomal University D.I. Khan,
Khyber Pakhtunkhwa, Pakistan

Abstract: Researchers have used different algorithms to provide safe communication through network but still secure communication is a challenge. In this study, we have introduced a novel algorithm for fast and secure data communication. The new algorithm combines the best features of two algorithms to reduce their limitations and maximize speed and security. The comparative results show that the new algorithm provides best security and faster communication. This algorithm will increase security in the field of communication technology. In the future, this technology can further enhance security by natural text like book page or newspaper text instead of dynamic generated cover.

Key words: Algorithm, information security, steganography, cryptography, communication

INTRODUCTION

Steganography is an important field of computer science which converts data, information or message into codes for secure communication so that the communication is protected from third party. The word "stagnos" means "secret" or "undisclosed" where "graphy" means writing or drawing. Therefore, steganography is a technology of writing messages highly protected and secured (Agarwal, 2013). Actually steganography hides messages inside cover using specific techniques. The final stego message is sent through the web to the destination where the information is recovered using a secret key as shown in the (Fig. 1).

The existing steganography technology is however not without weaknesses-the uses of special characters used to conceal the real meaning of the texts, however could create suspicions that senders are trying to hide certain very valuable message.

The missing letter puzzle algorithm conceals each character of secret message by replacing one or more letters by question marks in the cover. But, it actually makes it suspicious by excessive use of questions marks. Therefore, it is not secure (Agarwal, 2013), hiding data in word list hides a text in a list of words. The initial character and length of the words are determined by ASCHII values of the hiding characters. If the ASCHII sum equal to 1 then a will be the first character. If it equal to 2 then b will be the first character of the word and so on. The cover will be generated dynamically which also reduces the security. The algorithm hiding data using books or newspaper paragraphs as a cover. It hides the secret text characters using initial and ending character of words of cover file (Agarwal, 2013). However, this algorithm also have problems when the first and last characters of word are the identical.

Experts have suggested different techniques for hiding secret messages, cryptography and text Steganography being the common techniques (Purnama and Rohayani, 2015). Kaur and Kaur (2016) proposed a system to hide text in video. A technique which classifies Tamil alphabets into four groups to hide secret data and finally generate summery of the text was introduced by Manimozhi *et al.* (2015). Researchers have used many different techniques for text steganography (Bender *et al.*, 1996; Cummins *et al.*, 2004; Khairullah, 2011; Agarwal, 2013).

State-of-the-art in text steganography: According to Challita and Farhat (2011), steganograpgy and cryptography can be combined for faster more protected communication as compare to conventional methods. So, the secret messages can be hidden from third parties.

Corresponding Author: Rehman Ullah Khan, Faculty of Cognitive Sciences and Human Development, Universiti Malaysia, Sarawak, Malaysia

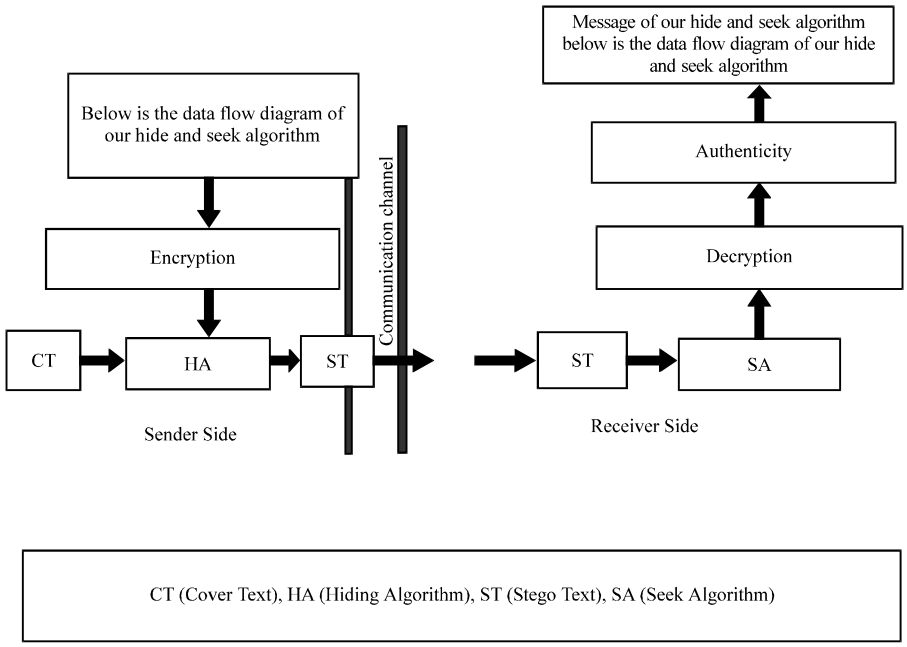

Fig. 1: Text steganography model

Their contribution is the proposed protocol for secure communication. The first protocol secures the cover object while, the 2nd protocol is multiple-cover-object in which more than one cover objects are used for hiding the secret messages.

Similarly, Juneja and Sandhu (2013) have introduced a new algorithm for image steganograpgy. They also proposed a data compression algorithm. This technique couldn't provide a secure method for secret communication even though they claim that this algorithm is efficient in bmp images. Likewise (Tech, 2013) have proposed an algorithm for concealing text in similar bits of the picture. They compared their algorithm with the LSB algorithm which hides the secret information in image pixels. Both of these strategies are analyzed based on the proportion between the likenesses and dissimilarities of the characters values and pixels colors values of the picture.

Gupta *et al.* (2012) in their research have suggested a system for hiding text into gray image (BMP). To get the image having secret message, the system requires both text and cover of the text. The algorithm uses RC4 stream cipher method to encrypt the message and store in non-sequential pixels of the image by using variable hope value power of 2 (Banerjee *et al*., 2012; Bennett, 2004; Juneja and Sandhu, 2013). This algorithm is comparatively faster and having multi-level security regardless of the size of secret message.

Prashanti *et al.* (2013) proposed a modified version of standard encryption algorithm for image steganograpgy. This algorithm uses addition modulo operation to replace XOR operation which is utilized in the sixteenth cycle of the standard cryptography algorithm for image steganography using LSB and MSB techniques. This algorithm is comparatively more secure on both sender and receiver side.

Bennett (2004) have combined Rivest, Shamir, Adleman (RSA) and diffe hellman algorithms to cipher the plain message. It was found that time complexity does not affect the encryption process if the diffe hellman algorithm is replaced by RSA algorithm. This algorithm provides more security at the cost of high processing time. Banerjee *et al.* (2012) introduced an algorithm for Indian regional language (ORIA language) by consolidating specific characters. Their algorithm doesn't change the length of secret message and cover, therefore it is more secure.

The purpose of this study is to eliminate the drawbacks of first to algorithms of Agarwal (2013) and develop more secure algorithm for text stegnography. In this study, we developed a new algorithm to further enhance the security capacity of stego texts so that texts can be sent without special characters that could invite suspicions from the third party. This algorithm combines the best features of first to algorithms of Agarwal (2013) to overcome the above-said drawbacks and maximize the performance in terms of speed and security. The new algorithm that we have developed in this research batter than the above approaches; it has reduced their limitations by combining the best features of the two algorithms into single algorithm.

MATERIALS AND METHODS

Missing letter puzzle algorithm is a mix-up of different words which hides some characters of every word. It replaces few alphabets by question mark in each word of the secret message. The recipient will use a proper key to replace all the question marks with suitable characters to get meaningful words. However, in lengthy messages its security decreases because the cover is dependent on the length of ASCII values of the secret message.

Similarly hiding data in wordlist algorithm uses list of words to hide a text using ASCII values of the text. For example if the sum of ASCHII values equal to 1 then the word will start from a. If it equal to 2 then the word will start from b and so on. So size and initial letter of the word is determined by ASCII values of the secret message. The cover of the stago message will generated dynamically. The limitation of this algorithm is using ASCII values and dynamically generation of cover file. The contents do not look natural and thus, multiple stego files may draw suspicion.

Therefore, we developed a new "Hide and Seek" algorithm by combining best features of first to algorithms of Agarwal (2013). We have used CSharp.net (C#) programming language to redesign and develop the hide and seek algorithms. We have used encipher and decipher algorithms of Agarwal (2013).

Data flow of hide and seek algorithm: The new proposed algorithm works in the same way as its ancestors and therefore the data flow diagram is same as in (Agarwal, 2013) and shown in Fig. 2.

The new proposed algorithm: As missing letter puzzle has its own hide and seek algorithms and hiding data in word list also has its own hide and seek algorithms. The hide algorithm of missing letter puzzle has 8 steps while the hide algorithm of hiding data in word list has 10 steps. We have merged best components of the two hide algorithms into one and developed a new hide algorithm which has 14 steps. Similarly we have modified the Seek algorithm of missing letter puzzle and the seek algorithm of hiding data in word list and combined into a new algorithm. So we have merged best components of both of these two seeks algorithms and developed a new seek algorithm which has 16 steps. Thus, the new hide and seek algorithms inherit the best components from their parents. We developed a prototype application using microsoft csharp.net (C#) programming language to test and validate the new proposed algorithm. The comparative results show that the new algorithm provides best security and faster communication.

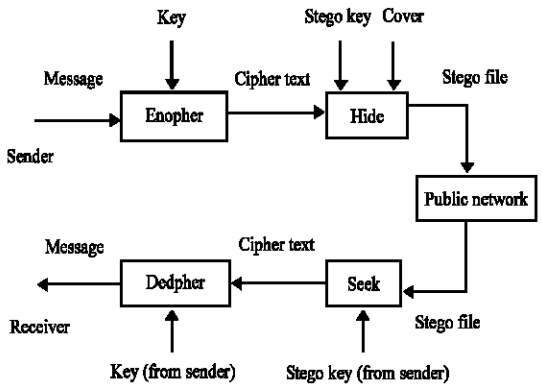

Fig. 2: The proposed model of text steganography (Agarwal, 2013)

Pseudocode of the new proposed algorithm

Hide algorithm:

Step 1-3 are from Monika Agarwal (1) missing latter puzzle hide algorithm.
Calculater length of the input file
Read a character from the file and get its decimal equivalet(n)
If n<100
    a) Flag = 0
//Step 4-14 are from Monika Agarwal (1) hidding data in word list, hide algorithm with some modifications
K = most significant digit of n. Write k in the stago key file
1 = middlw digit of n
If 1<6, then l = 1+10
sum of the digits of n
Get a l-letter word starting with the english sth letter of the alphabet and write it im the stego file
Repeat steps 2-8 till the end of the file
If le<10;
    Insert 10-le ten letter wods in the sego file
    Else
     Flag = 1+least significant digit of n
     q = most significant digit of n
     r = middle digit of n
     l = 10+q
      Read a l-letter word
      If le<10
      Else miss the character at position r of the word
Write the word in the stego file and the flag in the stego key file
Repeat steps 2-11 till the end of the file
If In<10 the insert 10-In ten letter words (missing the character at random position thw word) in the stego file
The stego key are sent, separately, to the receiver

Seek algorithm:

Read a value (k) from the stego key file
Read a word from the stego file and get its length (l)
If l>9, then l = l-10. // Step 1-3 are from Monika Agrwal (l) missing letter Pu: Seek algorithm. But the rest we have modified
    Calculater as by decoding the first letter of the word from the english alphabet:
        r = s-(1+k)
        The extracted value, n = (k*100)+(1*10)+r
        Convert n to its character equivalent
        Repeat above steps till the end of the stego key file
    Else
        A = k-1

Calculater length (l) of the word
l = l-10
R position of the missing character in the word
If r>9 then r = o
The extracted value, asc = (l*100)+a
Convert asc to its character equivalent
Pepeat above steps till the end of the stego key

RESULTS AND DISCUSSION

Time efficiency: The following 7 messages were used as input for all four (missing letter puzzle, hiding data in word list, hiding data in paragraph and the proposed) algorithms using the newly developed prototype application. The execution time was calculated. The results were stored as in Table 1. The last column shows our results:

- Yahya khan
- Yahya khan is working
- Yahya khan is working as lecturer
- Yahya khan is working as lecturer in Gomel
- Yahya khan is working as lecturer in Gomel University
- Yahya khan is working as lecturer in Gomal University deraismail khan
- Yahya khan is working as lecturer in Gomel university deraismail khan kyberpakhtunkhwa

Table 1: Time taken by each messages in millisecond

Algorithm names	Millisecond 1	2	3	4	5	6	7	Avg. time
Missing letter puzzle	462	863	1391	1746	2264	2861	3662	1892.7
Hiding data in wordlist	238	437	697	877	1098	1437	1839	946.1
Hiding data in paragraphs	294	468	702	889	1107	1454	1860	967.7
Proposed algorithm	237	438	697	877	1098	1437	1837	945.8

From the results it is clear that the average execution time of the proposed algorithm is shorter and speedier from the previous technology. Therefore the proposed algorithm is faster than the other three algorithms. Similarly it is clear from the following graph that the proposed algorithm is more efficient. The graph in (Fig. 3) also shows the average time of missing latter puzzle increases with the size of message. The proposed algorithm shows a consistent behavior irrespective to the size of message.

Capacity: The amount of information concealed in cover is called capacity. It can be calculated as follows. (Shirali and Shahreza, 2008):

$$\text{Capacity ratio} = \frac{(\text{Amount of hidden bytes})}{(\text{Size of the cover text in bytes})}$$

If one character uses one byte of memory we have calculated the percentage capacity which is capacity ratio multiplied by 100. From Table 2 and 3 and Fig. 4, it is clear that our algorithm can accommodate more information in cover. Monika Agarwal has proved in that hiding data in paragraph is better than the other two algorithms (Agarwal, 2013). Our experimental results show that our proposed algorithm is better than hiding data in paragraph.

Table 2: Percentage capacity over the seven experimental message

Variables	1	2	3	4	5	6	7
Missing letter puzzle	7.40	7.34	8.12	7.92	7.72	8.09	7.98
Hiding data in wordlist	7.87	7.66	8.04	8.04	8.22	8.18	8.31
Hiding data in paragraph	1.00	2.10	3.30	4.20	5.3	7.00	4.54
Proposed algorithm	7.35	8.07	8.02	8.04	7.81	8.14	8.01

Table 3: Average percentage capacity of the four algorithms

Missing letter puzzle	Hiding data in wordlist	Hiding data in paragraph	Proposed agorithm
7.79	8.04	4.54	7.92

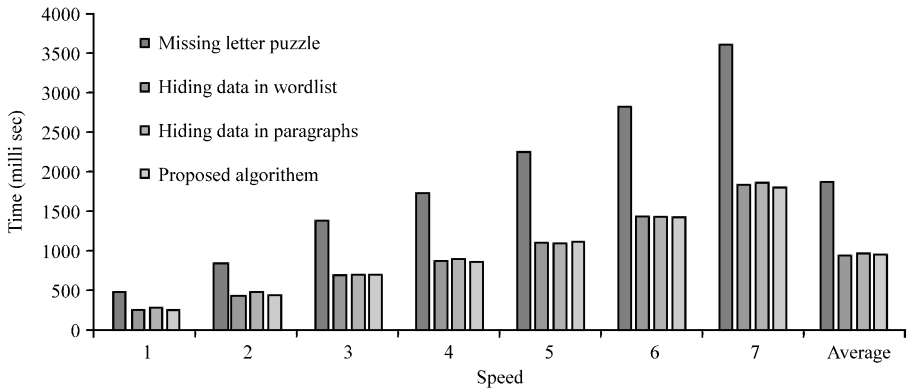

Fig. 3: Speed comparison in milliseconds

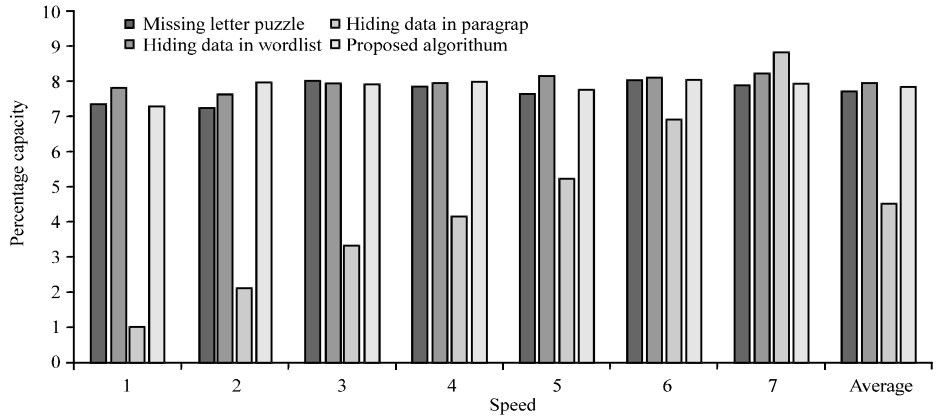

Fig. 4: Percentage capacity comparison; percentage capacity the seven experimental message

CONCLUSION

The novel algorithm is composed of the best features of the first two algorithms of Agarwal (2013). Therefore, it reduces their drawbacks and enhance security and speed. The missing letter puzzle algorithm hides every letter of the text by replacing one or more letters by question marks in the cover. But the excessive use of question marks makes it suspicious. Therefore, it is not secure (Agarwal, 2013). The dynamic generation of cover reduces the security of hiding data in word list algorithm. The algorithm hiding data in paragraphs using books or newspaper paragraphs as covers. It hides the secret text characters using initial and ending character of words of cover file (Agarwal, 2013). However, this algorithm also have problems when the first and last characters of word are the identical. The average percentage capacity of the proposed algorithm was found to be better than its parent algorithms. The stego files have no clue if they are open in any word processing programs. Therefore, the files are secure. We have eliminated the drawbacks of these algorithms by combining the best features of the two algorithms into single algorithm and this is our contribution to propose, design and develop a new algorithm which outperforms than its predecessors.

REFERENCES

Agarwal, M., 2013. Text steganographic approaches: A comparison. Intl. J. Network Secur. Appl., 5: 91-106.

Banerjee, I., S. Bhattacharyya and G. Sanyal, 2012. A procedure of text steganography using indian regional language. Intl. J. Comput. Network Inf. Secur., 4: 65-73.

Bender, W., D. Gruhl, N. Morimoto and A. Lu, 1996. Techniques for data hiding. IBM Syst. J., 35: 313-336.

Bennett, K., 2004. Linguistic steganography: Survey, analysis and robustness concerns for hiding information in text. CERIAS Technical Report, Purdue University, West Lafayette, IN 47907-2086. https://www.cerias.purdue.edu/assets/pdf/bibtex_archive/2004-13.pdf.

Challita, K. and H. Farhat, 2011. Combining steganography and cryptography: New directions. Int. J. New Comput. Archit. Appl., 1: 199-208.

Cummins, J., P. Diskin, S. Lau and R. Parlett, 2004. Steganography and digital watermarking. MSc Thesis, University of Birmingham, Birmingham, England.

Gupta, S., A. Goyal and B. Bhushan, 2012. Information hiding using least significant bit steganography and cryptography. Intl. J. Mod. Educ. Comput. Sci., 4: 27-34.

Juneja, M. and P.S. Sandhu, 2013. An improved LSB based Steganography with enhanced security and embedding/extraction. Proceedings of the 3rd International Conference on Intelligent Computational Systems, January 26-27, 2013, Hong Kong, China, pp: 29-34.

Kaur, J. and J. Kaur, 2016. Hiding text in video using steganographic technique a review. Intl. J. Eng. Sci., 17: 578-582.

Khairullah, M., 2011. A novel text steganography system in cricket match scorecard. Intl. J. Comput. Appl., 21: 43-47.

Manimozhi, K., V. Kalaichelvi, M. Poornima and A. Sumathi, 2015. An approach for text steganography: Generating tamil text summary using tamil phonetics. Intl. Rev. Comput. Software, 10: 137-143.

Prashanti, G., K.S. Rani and S. Deepthi, 2013. LSB and MSB based steganography for embedding modified des encrypted text. Intl. J. Adv. Res. Comput. Sci. Software Eng., 3: 788-799.

Purnama, B. and A.H. Rohayani, 2015. A new modified caesar cipher cryptography method with legibleciphertext from a message to be encrypted. Procedia Comput. Sci., 59: 195-204.

Shirali, S.M.H. and M.S. Shahreza, 2008. A new synonym text steganography. Proceedings of the 2008 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP'08), August 15-17, 2008, IEEE, Yazd, Iran, ISBN:978-0-7695-3278-3, pp: 1524-1526.

Tech, H.S.M., 2013. Analysis and implementation of algorithm to hide secret message. Intl. J. Adv. Res. Comput. Sci. Software Eng., 3: 327-333.