Ankita Sharma et al. 2015, Volume 3 Issue 6
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Data Transmission and Reception Using Text Steganography with Error Detection

[1]Ankita Sharma, [2]Anjana Sangwan

Abstract

In the research area of text steganography and cryptography, most dominant part was the key used for the hiding, encrypting and placing a data into another data is very crucial, also it is important that an algorithm should not have any limitations regarding the bits size of the secret data, case sensitivity or it should not be language bounded. Approach implemented in this paper suggests the use of text information and cover text to be combined and formed in such a way that it can be used or transformed into any other format including images, audio or into any other multimedia file. Along with the format transformation hiding the encryption key is another important feature of this approach, as the key used for encrypting and hiding data into another data is combined with the transmitted data that it should not be visible from the outside and should not represent any separate data as such.

Keywords: Steganography, cryptography and encryption.

Introduction

Cryptography is a technique of keeping and transmitting data in a specific format so that only the intended user can read and practice it. The term is most often associated with scrambling plaintext into cipher text (encryption), then back again (known as decryption).The limitations of then cryptography overcome by the Steganography, is the hiding of a secret message within an ordinary message and the extraction of it at its destination.

Steganography raised significance as the US and the British government, after the advent of 9/11, the use of cryptography was banned and publication sector needed to cover copyright marks [6]. In steganography, the data to be covered is called embedded text. A harmless medium, such as audio, text, image or video file; which is to be used for hiding embedded message is called cover data. The key which is optional, used in inserting procedure is called as stego-key. A stego-key is used for controlling the concealing procedure so as to limit discovery and/or retrieval of inserted data to the users who know the process [1]. The stego object is an entity i.e. known after concealing the embedded message in a cover media. Steganography can be categorized into audio, image, text and video steganography. Text steganography can contain whatever from varying the formatting of present text, to altering words within a sentence, to producing arbitrary character arrangements or using context-free grammars to produce readable message [7]. Text steganography is assumed to be the complicated due to shortage of redundant info which is present in text, image, audio or a video file. The construction of text data is alike with which is to be observed, whereas in additional categories of documents such as in image, the construction of document is dissimilar from with which is to be observed. Negligible changes can be done to a picture or an audio, video file, but in text files, even an extra letter or punctuation can be noticeable by a reader [9]. Storage of text file requires less memory space and its faster as well as easier communication marks it desirable to other types of steganographic approaches [10]. Text steganography can be generally categorized into three types: Format based Random and Statistical generation and Linguistic methods.

Previous Work

In this sub-section, we present some of the popular approaches of text steganography.

Line Shift

[1]Corresponding Author Email: ankitasharma1989@yahoo.com

© 2015 Ankita Sharma et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

In this method, secret data is concealed by shifting the text lines vertically to some degree [10, 11]. A line noticeable has two unnoticeable control lines one on either sideways of it for spotting the direction of movement of the noticeable line [12]. To conceal a bit 0, upwards the line is shifted and to conceal bit 1, downwards the line is shifted [13]. Determination of where the line has been moved up or down is completed by determining the distance of the centroid of noticeable line and its control lines [12]. The hidden data would get damaged if the word is typed another time or if a (OCR) Character Recognition Program is used [10].

Word Shift

In this method, secret text is secreted by horizontally shifting the words, i.e. right or left to symbolize bit 1 or 0 respectively [13]. Words shift are identified using correlation process that considers a profile as waveform and agrees whether it initiated from a waveform whose center slab has been moved left or right [12]. This technique can be recognized less, because variation of distance between words to fill a statement is very common [10, 11]. But if somebody identifies the procedure of distances, the person can compare or match the stego text with the algorithm and find the secreted content by the difference. Also, retyping or using OCR sequencers abolishes the concealed data [10, 11].

Syntactic Method

This method uses marks of punctuation such as comma (,), full stop (.), etc. to hide bits 1 and 0. But problematic area with this technique is that it needs identification of right places to insert marks of punctuation [10, 11]. Hence, care must be done in applying this technique as person who reads can notify incorrect placement of punctuations [9].

White Steg

This method uses white spaces for concealing a secret data. There are three means of hiding information by using white spaces. In Inter Sentence Spacing, insertion a single space to cover bit 0 and two spaces to conceal bit 1 at the end of individually terminating character [9]. In End of Line (EOL) at the last of every line spaces, insertion of fixed number of spaces is done. Just like, two spaces to convert one bit per line, four spaces to convert two bits and go on. In Inter Word Spacing method, one space afterward a word denotes bit 0 and two spaces afterward a word denotes bit 1. But, uneven use of white space is not transparent [9].

Spam Text

To cover bits HTML and XML files can be used also. If there are several opening and closing tags, bit 0 is considered and if for starting and closing only tag is used, then bit 1 is considered [13]. In alternative method, bit 0 is denoted by a lack of space in a tag and bit 1 is denoted by placing a space inside a tag [13].

SMS-Texting

SMS-Texting language is a grouping of shortened terms used in SMS [8]. Using full form of name or its abridged form binary data can be concealed. To store words and their respective shortened forms a Codebook is made. Full form of the word is used to hide bit 0, and abbreviated form of word is used to hide bit 1 [8].

Feature Coding

In feature coding technique, the secret text is hidden by altering one or more characters of the text. All the features which can be used to conceal the information are surveyed and picked up by a parser in a text file [13]. Just like, points in letters ‘i’ and ‘j’ can be placed otherwise, length of strike can be altered in letters f and t, or by lengthening or shortening the height of letters ‘b’, ‘d’, ‘h’, etc. [6, 14]. There is an error in this process which is, if an OCR package is used or if re-writing has been done, the concealed message would get damaged.

SSCE (Secret Steganographic Code For Embedding)

This method first ciphers a data using Secret Steganographic Code for Embedding table and then encapsulates the encrypted text in a face file by putting articles a or an with the non-precise nouns in English language using a definite mapping system [15]. The embedding positions are ciphered using the similar SSCE table and set aside in other file which is transmitted to the recipient with the stego file surely.

Word Mapping

In this method a secret message is ciphered using inherited operative crossover and then inserts the resultant cipher text, using two bits at a time, in a mask file by putting blank spaces in between words of even or odd length by means of some mapping method [16]. The embedding positions are stored in some another file and sent to the recipient along with the stego entity.

MS Word Document

In this method, text sections in a article are deteriorated, imitating to be the work of an writer with substandard writing skills, with confidential text being concealed in the choice of deteriorations which are then reviewed with modifications being traced [17]. Data inserting is concealed such that the stego article seems to be the product of combined writing [17].

Cricket Match Scorecard

In this technique, data is concealed in a cricket match record by earlier appending a useless zero previous to a number to symbolize bit 1 and parting the number as it is to signify bit 0 [18].

CSS (Cascading Style Sheet)

This method scrambles a data using RSA public key cryptosystem and secret message text is then fixed in a Cascading Style Sheet (CSS) by using End of Line on each CSS style properties, closely after a (;) semicolon. A space afterward a semicolon inserts bit 0 and a double space afterward a semicolon inserts bit 1 [19].

Related Work

Some of the exceptional characteristics, inflexion, static word order and usage of periphrases, of language i.e. English language are used for the text steganography technique. The meaning of the word inflexion is that it can point out the connection of the words into a sentence by minimum changing of shape. In static order, each word's place in a sentence decides its connection with the others. Periphrases are small phrases and can be stated as the different means to express something.

A specific code called Vedic Numerical Code used in decoding and deciphering Sanskrit text is described. The coding depends on the tongue position. To apply the Vedic code to the alphabet of English language, frequency of letters (alphabets) in English dictionary is used as the base of allocating numbers to the alphabets in the English language.

Frequency of letters varies in between the specified length. No perception is made for allocating coding number to consonants and vowels.

Each letter in the English alphabet is given a number in between the range of 0 to 15 [20].

Proposed Work

This research paper emphasize on hiding data and key in a single and non-separable matrix. Data is first scrambled with the randomly generated data and then further processed.

Data embedding and transmission steps

Step 1: Data input.

Step 2: Display the entered data.

Step 3: Data converted into Machine code.

Step 4: Size and Length calculated.

Step 5: Store the size of the entered data into variables.

Step 6: Range determination of the entered data for the creating the envelope data.

Step 7: Store the range of the secret data.

Step 8: Envelope generation algorithm started.

Step 9: Envelope created with range similar to secret data.

Step 10: Data embedding algorithm started, data is dived into sections and then it is placed along with the envelope data and in between of the envelope data to create a new matrix which includes secret data in a scrambled form along with random data.

Step 11: Determine and store the range and the size of the newly formed encapsulated data from Step 10.

Step 12: A new random matrix is generated by using the size and the range values obtained in Step 11.

Step 13: Random number selection algorithm started based on the number of elements present in the data and the envelope encapsulated matrix.

Step 14: Operations performed with the scalar quantity obtained from Step 13 on the encapsulated matrix obtained from Step 10.

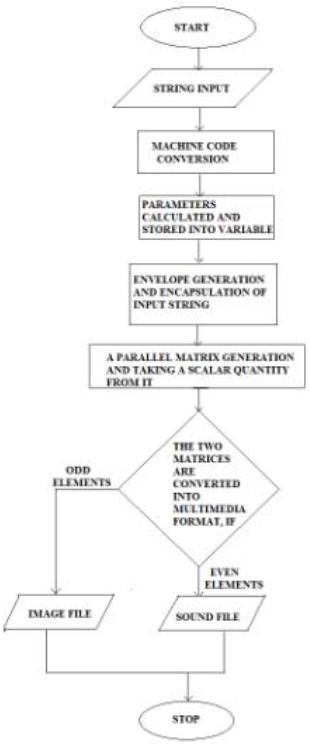

Figure 1: Data Embedding and transmission

Step 15: Now the matrices are converted into multimedia format, the two matrices random matrix and the encapsulated matrix will coincide with each other to form a different matrix so that it can be converted into any form of multimedia file.

Step 16: Now, there will further complex functions to be implemented on the multimedia file generated from step 15.

Step 17: Data transmission started.

Data receiving and extracting steps:

Step 18: Data received.

Step 19: Data extraction started.

Step 20: Level 1 extraction performs the reverse of the operations performed in the Step 16.

Step 21: Level 2 extraction performs the separation of the two matrices which formed the multimedia image in Step 15.

Step 22: Step 13 is again followed at the receiver end.

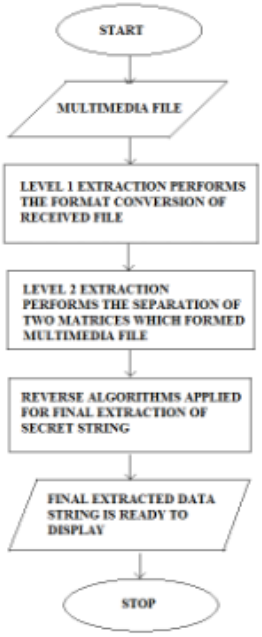

Figure 2: Data Receiving and Extraction

Step 23: Secret data is now been fetched by the extractor by applying the reverse algorithm of the procedure followed at the transmitter end in Step 10.

Step 24: Data extracted and ready for the display.

Results

Step 1: Firstly, taking the secret text as an input which is to be steganographied further by applying encoding algorithm.

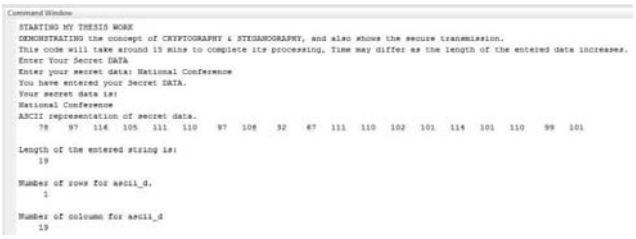

Figure 3(a): Taking Secret Text as an Input.

Step 2: The steganographied secret data is further transmitted from transmission end to receiving end.

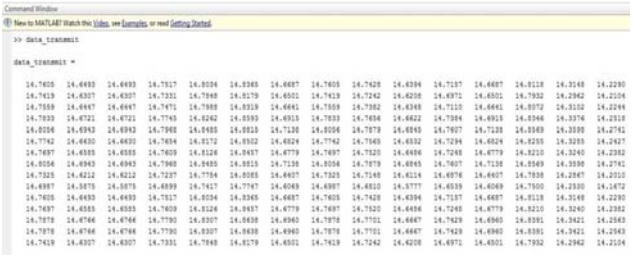

Figure 3(b): Data Transmission.

Step 3: The transmitted data is now received at receiver side and extraction of secret data is performed by applying decoding algorithm and output is received.

Figure 3(c): Extracted output at receiver end.

This is how the work is done by encapsulating the secret data by applying the suitable algorithms and complex operations at the transmission end. Similarly the extraction performed at the receiver end by using the decoding algorithms. The data is extracted in the same format as it was entered as an input.

Table 1: The comparative table of the parameters of papers studied with the proposed methodology

Parameters	Ours	Paper 1[21]	Paper 2[20]
Length	Unlimited	Unlimited but needs to be reduced	Limited
Language Bounded	No	Yes	Yes, Indian Root
Database Requirement	Not Required	Required	Required
Words Dependency	No	Yes	Yes
Format Transformation	Possible	Not Specified	Not Specified
Time	10ms	27ms	Not Specified
Accuracy	99.98%	Not Specified	Not Specified
Error Resistive	Yes	Not applicable	Not applicable

The results shown in this paper are based on the simulation done over the MATLAB. MATLAB runs the code for approximately 10 milliseconds and generated the above presented results.

The system used for the simulation has the following configurations: Processor Intel(R) Core(TM)2 Duo CPU T6570 @ 2.10GHz, 2101 MHz, 2 Core(s), 2 Logical Processor(s), OS Name: Microsoft Windows 7 Ultimate, Version: 6.1.7601 Service Pack 1 Build 7601, Installed Physical Memory (RAM): 3.00 GB.

This code has been tested over the latest available MATLAB software with us i.e. MATLAB 2012b. The omission of the public and private key makes this system more secure, also the use of random functions makes it more reliable and complex as far as security is concerned.

Conclusion

From the above obtained results and the compiled code it can be concluded that the text steganography is an essential factor in transmitting important text information to another end. From the code complied in this thesis and the result studied in the previous work, it can be concluded that to send the data over a channel by scrambling and transforming it into different outlook there will also be a key required to decode it back to its original form, in this thesis this drawback has been removed up to a certain extent by combining the data and the key in a single matrix.

The key generated in this model is random process and cannot be controlled by any individual or by any system or any algorithm. The above formed matrix can be transmitted in the form of a sound signal to confuse or to create a huss at the unauthorized end which is not intended to decode the received data.

Also the transmitted data can be converted into any other format without affecting its original format and without adding any new key to it.

Future Scope: The above designed algorithm can further be developed or improved by finding the method to depict the exact location of the disturbed elements of the transmitted data by calculating its determinant, which needs to be zero if there is no error present in the received matrix.

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, pp. 318-322, 2010.

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

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

[10] 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 1st IEEE/ACIS Int. Workshop on Component-Based Software Engineering, Software Architecture and Reuse, 2006, pp. 310-315.

[11] M. H. S. Shahreza, and M. S. Shahreza, "A new synonym text steganography," Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing, pp. 1524-1526, 2006.

[12] S. H. Low, N. F. Maxemchuk, J. T. Brassil, and L. O. Gorman, "Document marking and identification using both line and word shifting," INFOCOM'95 Proceedings of the Fourteenth Annual Joint Conf. of the IEEE Computer and Communication Societies, pp. 853-860, 1995.

[13] J. Cummins, P. Diskin, S. Lau, and R. Parlett, "Steganography and digital watermarking," School of Computer Science, pp.1-24, 2004.

[14] J. T. Brassil, S. Low, N. F. Maxemchuk, and L. O. Gorman, "Electronic marking and identification techniques to discourage document copying," IEEE Journal on Selected Areas in Communication, vol.1, pp. 1495-1504, 1995.

[15] I. Banerjee, S. Bhattacharyya, and G. Sanyal, "Novel text steganography through special code generation," Int. Conf. on Systemics, Cybernetics and Informatics, pp. 298-303, 2011.

[16] S. Bhattacharyya, I. Banerjee, and G. Sanyal, "A novel approach of secure text based steganography model using word mapping method," Int. Journal of Computer and Information Engineering, vol.4, pp. 96-103, 2010.

[17] T. Y. Liu, and W. H. Tsai, "A new steganographic method for data hiding in Microsoft word documents by a change tracking technique," IEEE Transactions on Information Forensics and Security, vol.2, no.1, pp. 24-30, 2007.

[18] M. Khairullah, "A novel text steganography system in cricket match scorecard", Int. Journal of Computer Applications, vol.21, pp. 43-47, 2011.

[19] H. Kabetta, B. Y. Dwiandiyanta, and Suyoto, "Information hiding in CSS: a secure scheme text steganography using public key cryptosystem," Int. Journal on Cryptography and Information Security, vol.1, pp. 13-22, 2011.

[20] Souvik Roya, P.Venkateswaran, "A Text based Steganography Technique with Indian Root", International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA), vol. Procedia Technology 10, pp. 167 – 171, 2013.

[21] Shivani, Virendra Kumar Yadav, Saumya Batham, "A Novel Approach of Bulk Data Hiding using Text Steganography", 3rd International Conference on Recent Trends in Computing (ICRTC), vol. Procedia Computer Science 57, pp. 1401 – 1410, 2015.

Author's details

[1]M.Tech Scholar, Computer Science Engineering, Swami Keshvanand Institute of Technology, Rajasthan, India, ankitasharma1989@yahoo.com

[2]Senior Lecturer, Computer Science Engineering, Swami Keshvanand Institute of Technology, Rajasthan, India, sangwan.anjana@gmail.com

Copy for Cite this Article- Ankita Sharma and Anjana Sangwan, "Data Transmission and Reception Using Text Steganography with Error Detection', *International Journal of Science, Engineering and Technology,* Volume 3 Issue 6: 2015, pp. 195- 200.