International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 4 Issue: 7 377 – 381

Review of Linguistic Text Steganographic Methods

Sunita Chaudhary
Dept. of CS & IT
Jagannath University, Jaipur
er.sunita03@gmail.com

Dr. Meenu Dave
Dept. of CS & IT
Jagannath University, Jaipur
meenu.s.dave@gmail.com

Dr. Amit Sanghi
Dept. of CSE
MEC, Bikaner
dr.amitsanghi@gmail.com

Abstract:- Steganography is a method of concealing confidential data in a cover file such that attacker cannot predict about clandestine data. Steganography exploit cover message, for example content, picture, sound, video record and so forth to conceal a mystery or secret message. Text Steganography is one of a procedure to conceal the one kind (text) of content data inside same type of content messages. Linguistic steganography is the language based steganographic scheme which proposes more advanced methods to hide the secret messages in text. Initially linguistic text steganographic techniques are developed only for the English language. But now days different regional languages are also used to hide the information like Hindi. This paper reviews the different linguistic text steganographic methods of hindi and English language.

Keywords:- *Linguistic, Steganography, Text, Random, conceal.*

*****

I. Introduction:- Steganography word is a combination of two technical terms. These are Greek words “stegno” and “graphy”. The later word means concealing under a cover and former word means art, style of writing etc. or we can say Greek state is the place where Steganography word originates from Greek phrasing and that denote "secured composing or writing". Steganography is the act of concealing a classified message in a different non-mystery message into a harmless digital media with the end goal that it hides correspondence or Steganography is a method of concealing confidential data in a cover file such that attacker cannot predict about clandestine data. Steganography incorporates multiple different strategies for concealing a message in a range of media.

II. Cataloging of Steganography in Different Perspective

1. Classification of Steganography:- Figure 1 elaborates the classification of steganographic approach. Steganography is subdivided into two categories, one is technical steganography and another is linguistic steganography.

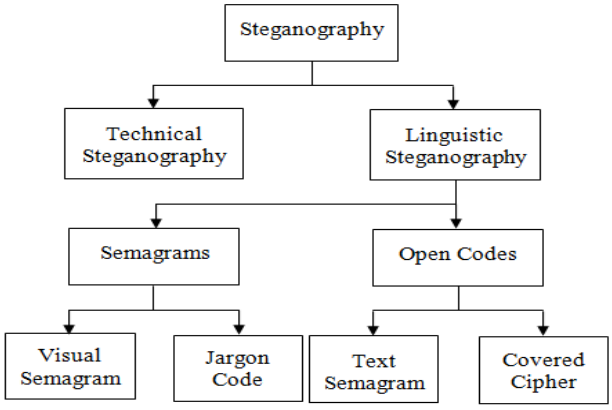

Figure 1. Cataloging of Steganography [1].

Technical steganography deals with the scientific methods like invisible ink and other materialized and size reduction approaches. On the other hand linguistic steganography hide the message by using some non obvious or nontechnical methods. It further can be sub grouped in semagrams and open codes methods. As the name referred semagrams use different symbol and sign of the particular language to hide the message. Visual semagrams generally use the visual effects which attract the innocent people most like; doodles etc. Text semagrams performs changes in the looking of the cover text to hide the message like it make changes in the font size, color, height and width of letters, add extra space, add extra letter etc.[1].

To hide the secret message, open codes use carrier message in such a way that it is not commonly visible to the innocent persons. Open code is categorized in two ways; jargon codes and covered cipher. Jargon code use properties of the particular language to hide the message, so that it can be understood only by the people who are familiar with that particular language. In covered cipher secret message is openly embedded into the carrier message, so it can be recovered by any person who is aware about the method of embedding.[1].

2.Classification of Steganography according to key:- In steganographic system use of key is optional. In the earlier techniques keys are not used, but in current techniques keys are used due to increase in security. As use of key take the security at one level up. So, according to the use of key steganographic system can be categorized in three types as shown in figure 2.

 ISSN: 2321-8169
377 – 381

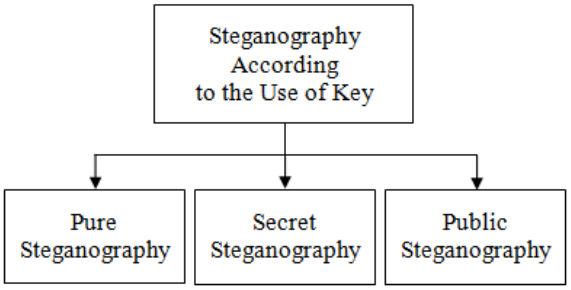

Fig. 2. Categorization of Steganographic System according to the key

1. Pure Steganography:-

It is the method, which follow the concept of no key or pure steganography is the steganography which do not require any prior exchange of data before sending the actual secret message. Means the security of the system is depends on its own strength of secret method only and there is no need of any communication regarding key between the sender and receiver before starting the session.

2. Secret Key Steganography:-

In this method one secret key is used to embed the data into the cover text. This secret should be known to both the sender and receiver, so that after embedding by the sender, receiver can extract the original message with the key. Thus, we can say it like a symmetric key cipher method and used secret key also travel with the embedded data during communication or should be prior known to the sender.

3. Public Key Steganography:-

Public key steganographic methods use two keys. One is public key, which is known to both the sender and receiver and second is secret (Private) key, which is only known to receiver. Sender send the message by embedding the secret text into the cover text using public key of receiver and at the receiver's end receiver extract the secret text from stego text by using its own secret or private key.

3.Classification of Steganography According to Carrier file:-

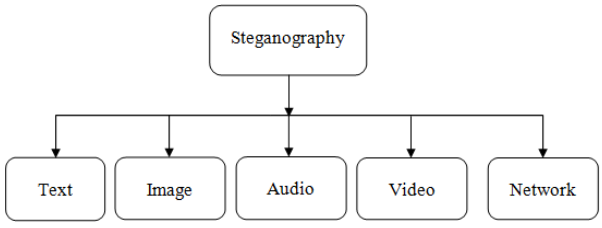

Fig. 3.Classification of Steganography According to the carrier file.

1. Text steganography:- Text Steganography utilizes the properties of text and other text language features for transferring the secure messages over the digital medium

2.Image Steganography:- Image Steganography use image object as the cover media to hide information. Generally, in this type of steganography, pixel intensities are used to cover up the information. File extension such as PNG, BMP, GIF, JPG uses as cover media in image steganography.

3. Audio Steganography

If we hide secret information by using some audible file or music records as a cover medium than it is known as audio steganography. The host audio file used for cover media has same characteristics before and after applying steganography. The cover files have an extension such as WAVE, MIDI, and MPEG.

4. Video Steganography

Video steganography uses some frames of video files to embed a confidential message or secret file. It uses Mp4, AVI, MPEG or other video formats for information hiding. There are multiple methods available for video steganography such that substitution technique, transform domain technique, spread spectrum technique, statistical technique and distortion techniques etc.

5. Network or Protocol Steganography

Concealing of information is done using network protocols as CP, TCP, UDP, ICMP, IP etc. The Base is OSI layer network model in which covert channels are present those can be used for network steganography. For instance, there are some empty or least used fields available in TCP/IP header, these fields can be used to send secret message.

III. Linguistic Steganography

Initially linguistic text steganographic techniques are developed only for the English language. But now days different regional languages are also used to hide the information, which increase the security in term of difficulty to get the original message, as all the persons are not aware about all the language. Chinese, Urdu, Arabic, Parisian language are popular for the steganography as they have many letters with the dot symbols. Even hindi language is a language which became a common and beautiful way to communicate. Even persons from outside India are also interested to learn and speak this language. Hindi text steganographic approaches can be divided into two categories, as described in figure 4.

 ISSN: 2321-8169
377 – 381

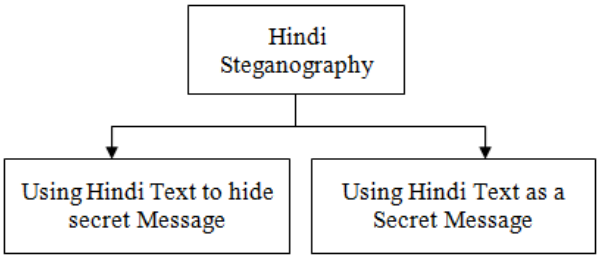

Fig.4.Classification of Hindi Text Steganography [1].

One in which, Hindi text is used as a cover text to hide the secret text of Hindi language itself or any other language. Second in which Hindi text is used as a secret message and embedded or encoded into the other language. There are many methods available which uses Hindi as a carrier text but limited methods are exists for hiding the Hindi text.[1].

IV. Techniques to hide English text

1. Line shift:- To hide the original data the line of text is shifted vertically by some degree as 3/10. To hide a 0 line can be shifted upwards and to hide a 1 we shift downwards or vice versa. We can also use no shifting for 0 and shifting up or down for 1.The problem with the method is the information is destroyed in case of retyping the text and also distances of the line shifting can be measured by instruments as by an OCR[2,3,4]. The main advantage of shifting algorithm is minimalism in executing, on the contrary hidden data ratio is very less compared to other algorithms of text steganography [4].

2. Word Shift:- In the scheme the confidential data is kept hidden by placing the words horizontally i. e. by increasing the length of the word in left or right. The method is less noticeable than the line shift because it gives illusion that the text is justified. But it is also having same problem as line shift i. e information is lost in case of an OCR machine reading or retyping[5,3].Same problem is exist in this method like line shift that is has very low hidden ratio.

3. Syntactic method:-The method take use of punctuation marks as comma (,), full stop (.) etc. These marks serves as a basis of hiding 0 or 1.The method is very good but requires a lot of care. An intruder having good knowledge of English can intercept because he or she knows that what the exact may position of such marks in a text document and has low hidden ratio [6,2].

4. White tag:- In this method white space or blank space serves as basis of concealing the information [6]. The method can be used in three different ways. One is Inter sentence case, in this scheme space at the end of a sentence is used to hide the secret or confidential data. Second is Inter word case, this method takes use of space available between words. Third is End of the line, this method takes use of the space available at the end of a line. The difficulty with these schemes is that incorrect use or retyping again makes the hidden data noticeable to an attacker [6].

5 Spam text:- In the scheme bits are hided in the tags of the markup language file as HTML or XML file. HTML starts and end tags are case insensitive as well as they can occur more than once. Space is also not a considerable thing while writing tags and all of these features can serve as a basis of concealing [4].

6 SMS –Texting :- People take use of short forms of words while messaging each other. These short forms are called abbreviated words. In this scheme a full word can hide a 1 and the abbreviated word can hide a 0 or vice versa [7].

7 Feature Coding :- In the method we can alter the features of the text one or more and the altered feature can serve as a basis of steganography. The feature can be style, shape and size of writing a letter or text. As for example size of the dot used in the small English alphabets i and j can be altered to hide a 0 or 1. [4, 8].

8 Secret Stenographic code for embedding :- The method take use of the article of the English language a, an, to hide 0 and 1[9]. For example to hide a 0 we use article “a” and to hide a 1 we use “an”.

9 MS word document :-In this method some parts of a text documents are relapsed using mimicking and further the relapsed or mimicked parts are used to hide a 0 or 1[10].

10 Cricket Score Board :- Cricket score board serves as the basis of hiding. As a senseless zero before a number can be used for concealing a 1 and the number without a leading zero can be kept as it is to hide a 0. [11].

11 Cascading style sheet :- The scheme comes under embedded cryptography and steganography approach discussed above. For the cryptographic part of the scheme RSA is applied and resultant cipher text is embedded with a CSS by using end of line white space scheme explained earlier in this paper [12].

V. Techniques for Hindi Text Steganography

1. **HHK, Matraye and Core categorization based method:-** In this researcher explains three approaches to Hindi text based steganography. One is based on matras or modifying character, the second approach take use of special letters, no bar, bar etc. And third approach is by using HHK Scheme.Paper explains Matraye of Hindi language can divide into three types as top, core and bottom. The top and bottom modifier can easily identify by Hindi OCRs. According to the feature of Hindi letters they are classified

as open, bar, special and no bar characters. It uses the Hexadecimal value to encode the vowel and constant of Hindi Language.[2].

2. **Using Punctuation Marks:-** In this method, punctuation marks are used for hiding the secret data. The secret message is hiding in these punctuation marks. In this punctuation marks are divided in four groups and according to the groups these punctuation marks hide binary bits 00, 01, 10 and 11 respectively.[13].

3. Synonym Based:- This method is based on word substitution which uses synonyms of Hindi language. A Hindi dictionary contains synonyms of Hindi word is used for word substitution. Secret message in binary form is then hide in Hindi sentence created by using the dictionary word and secret word bit stream. The same dictionary is used to decode the same dictionary is used to decode the message from Hindi sentence. Only words in dictionary are used for purpose of encoding.[13].

4. Sanskrit Classification Based:- The proposed scheme uses dictionary to hide the secret bit. The dictionary is of two levels. The Tatbhav words are placed at level zero and encode with bit value 0. The Tatsam words are placed at level one and encoded with bit value 1.[13].

5. Using Hindi Letters and its Diacritics:- Hindi Letters without diacritics are encoded as '0' and letter with diacritic are encoded as '1'.[8].

6.Hindi Numerical Code:- The vowel and consonants are assigned numerical code based on frequency of their occurring. Letter with least frequency occurring are assigned numeric code 0. Highest frequently occurring letter have high numeric value.[14].

7.Karak Kriyaye:- The Hindi karak Kriya Vibhakti can classify into eight types. These eight karak are arranged in four groups and every group has assigned a code or bits. Now, according to the hidden secret bits cover text is generated and these karak symbols are used to hide the bit sequence.[15].

8.Shifting Matra:- The paper [16] present the work on Hindi text steganography based on feature coding method. The Hindi text in cover file shifts Matra of Hindi letter left or right to hide the secret message, and the matra will remain unchanged for hiding 0 bit [16].

9.Use of Text Colour:- This paper exposes a new technique in feature encoding technique to hide secret message. The text colour is changed to hide data. For example a Devanagari letter '**v**' coloured with RGB values '000' and '111' which shown very similar coloured letter '**v**'. These values are used to hide data and encoding is done to provide security. These changed features are undetectable to human eye and software as well [17].

VI. Conclusion:- we have outlined a list of existing text Steganography techniques to hide bit-level information or used Hindi and English script to hide information. The advantages and the problems of the existing techniques have been analyzed. Extensive research has been carried out in the field of Hindi and English text steganography. And all the techniques proposed by the different researchers have, in their unique way, proven to be very useful. The main hindrance and problem that remains is that, In linguistic steganography it is very tedious job to take care of syntax and grammar means if we use any method other than random character as a carrier file will, it is hiding very small amount of data and need a very large number of English and Hindi words are used. Further research needs to be undertaken to tackle this drawback. Also, it has been seen that all the techniques that have been proposed pertains to hiding binary bits into Hindi words [1]. So, further research is also necessary in the field of Hindi text steganography where Hindi words can be hidden in some other form or medium.[1].

VII. References:-

[1] Tatwadarshi P. Nagarhalli, Dr. J. W. Bakal, Neha Jain, "A survey of Hindi Text Steganography", International Journal of Scientific & Engineering Research, vol.- 7, no.- 3, pp.55-61, Mar-2016.

[2] Kalavathi Alla., "A New Approach to Hindi Text Steganography Using Matraye, Core Classification and HHK Scheme", Seventh International Conference on Information Technology New Generations, vol. - 3, pp. 1223-1224, Apr 2010.

[3] 3.Chaudhary S, Dave M, Sanghi A, Manocha J., "An elucidation on steganography and cryptography", ACM Second international conference on information and communication technology for competitive strategies, vol.-9, no.-43, pp. 1-6, Mar 2016.

[4] Pramod P. Sukhadeve, S. K. Diwedi, "Enlargement of Clinical Stemmer in Hindi Language for Homoeopathy Province", Lecture Notes of the Institute for Computer Sciences Social Informatics and Telecommunications Engineering, vol.-13, no.- 3, pp. 239-248, Jan 2012.

[5] L. O'Gorman., "Electronic marking and identification techniques to discourage document copying", Conference on Computer Communications INFCOM-94, vol.- 13, no.-8, pp. 1495-1504, Oct 1994.

[6] Ray, Rishav, Jeeyan Sanyal, Debanjan Das, and Asoke Nath, "A New Challenge of Hiding any Encrypted Secret Message inside any Text/ASCII File or in MS Word File: RJDA Algorithm", International Conference on Communication Systems and Network Technologies, vol. – 6, pp. 889-893, May 2012.

[7] Mohammad Shirali-Shahreza, "Text Steganography by

Changing Words Spelling", 10th International Conference on Advanced Communication Technology, vol.- 3, pp. 1912-1913, Feb 2008

[8] Kalavathi Alla., "An Evolution of Hindi Text Steganography", Sixth International Conference on Information Technology New Generations, vol.- 8, pp. 1577- 1578, Apr 2009.

[9] Banerjee, Indradip. "An Approach of Quantum Steganography through Special SSCE Code", International journal of computer, electrical, automation control and information engineering, Engineering & Technology, vol.- 5, no.- 8, pp. 901-908, Apr 2011.

[10] Bensaad, Mohammed Lahcen, and Mohammed Bachir Yagoubi., "Boosting the Capacity of Diacritics-Based Methods for Information Hiding in Arabic Text", Arabian Journal for Science and Engineering, vol.- 3, pp. 2035-2042, Mar 2016.

[11] Xiaoliang Wang., "A Steganography Scheme in P2P Network", 2008 International Conference on Intelligent Information Hiding and Multimedia Signal Processing, vol.- 3, pp. 20-25, Aug 2008

[12] Odeh, Ammar, Khaled Elleithy, and Miad Faezipour, "Steganography in text by using MS word symbols", Conference of the American Society for Engineering Education, vol.- 1, pp. 1-5, May 2014.

[13] Mayank Srivastava, Mohd. Qasim Rafiq, and Rajesh Kumar Tiwari, "A Novel Approach to Hindi Text Steganography", International Conference on Advances in Communication, Network and Computing, vol.- 142, pp.295-298, June 2011.

[14] Megha Pathak, "A New Approach for Text Steganography Using Hindi Numerical Code", International Journal of Computer Applications, vol.- 1, no.- 8, pp. 56-59, Mar 2010

[15] Mishra, Rina, and Praveen Bhanodiya., "A review on steganography and cryptography", 2015 International Conference on Advances in Computer Engineering and Applications, vol.- 1, pp. 119-122, Apr 2015

[16] S. Changder, N. C. Debnath and D. Ghosh, "A New Approach to Hindi Text Steganography by Shifting Matra", International Conference on Advances in Recent Technologies in Communication and Computing, vol.-7, pp. 199-202, July 2009.

[17] Hitesh Singh, Anirudra Diwakar and Shailja Upadhyaya, "A Novel Approach to Text Steganography", 1st International Congress on Computer, Electronics, Electrical, and Communication Engineering, Singapore, vol.- 59, no.- 2, pp.7-12, Feb 2014.