A Review on Regional Language Text Steganography Techniques in India

[1]Sreenath.M, [2] Sumathi.D

[1]PG Scholar, [2]Assistant Professor,

[1,2] Computer Science and Engineering, PPG Institute of Technology Coimbatore,641035, India

***Abstract*— The rapid development in increase of communication over internet had made the attention towards the confidentiality of messages send over unsecured communication medium. Transmitting encrypted information frequently will bring the attention of third parties, i.e. crackers and hackers, possibly causing efforts to infringe and disclose the secret messages. In order to safeguard the secret data from attack, a number of information masking techniques was introduced. In a digital world, steganography is brought to conceal the existence of the communication by hiding a secret message within another indubitable message. Steganography is frequently being used together with cryptography and offers privacy and security in an acceptable way over the communication link. This paper focuses on overview of text steganography and review of regional language text Steganography methods.**

***Index Terms*— Steganography, information hiding, text steganography, regional language text steganography, information security**

I. INTRODUCTION

The principle of information hiding is first proposed and documented in "On the Criteria to be Used in Decomposing Systems Into Modules" in 1972 [1].Researchers are trying to implement new concepts in information hiding. There are three features that information hiding systems should hold: robustness, capacity, security [2]. Capacity belongs to the amount of information that is to be masked in the medium, whereas security is important constraint where a secret message is kept to be concealed and undetectable by hackers. Robustness refers to the amount of modification in the stego-medium that can confront before an antagonist can devastate secret information. Fig. 1: shows diagram for classification of information hiding techniques.

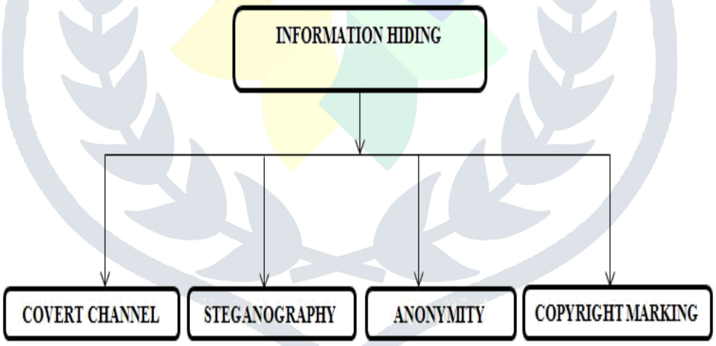

Fig.1: Information hiding techniques

.

II. STEGANOGRAPHY

Steganography is the art and science of sending surreptitious messages over public media so that its existence is only known by the sender and receiver. The term steganography is coined by Johannes Trithemius in his book "Polygraphia and Steganographia". The technical term itself is reaped from Greek words steganos and graphia which means covered and writing [3].Steganography has been practised for thousands of years, but for the last few decades it has been introduced to digital media. The history of steganography can be traced back from 440 BC.The use of steganography was first reported by Herodotus, historian, who mentions that in ancient Greek, secret text was written on wax tablets. Aeneas mentioned a lot of other steganography methods in his documents .Letters can be concealed in the women's ear rings or messengers' shoe soles, and the pigeons can be used to carry secret notes. Another bright idea was to shave the head of a messenger and to paint the secret letters on the messenger's head. A very common steganography scheme in olden days was the use of chemical substances to conceal information. Organic substances like urine, milk can be used to make invisible ink. The confidential message can be written among the lines with the veiled ink. The letters written with the invisible ink will become visible if the document is heated up gently. During World War II, the documents were sent which can conceal a secret message through the use of null ciphers (unencrypted message), which perfectly cover the real message in an ordinary letter. Very famous message sent by German spy during World War II [4]:"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.". The hidden message can be revealed by taking the second letter from each word results in the sentence: "Pershing sails from NY June 1."Steganography shows difference from cryptography that the steganography focuses on keeping the existence of the information secluded, cryptography focuses on keeping the contents of a message secret [5,6].

III. Types of Steganography

Figure 2: shows different types of stenography techniques based on different conveyance media like audio, text, images, or video [7]. The image and audio files are more preferred as transmission media because of their high degree of the inclusion of information than necessary for communication [8].In Image Steganography method the hidden message is embedded into an image as noise to it, which is nearly impossible to be detected by human eyes [6, 9].In video steganography, same technique may be used to embed a message [10, 11]. Audio steganography hides the message in cover audio file as noise at a frequency out of human hearing range [12].Text steganography is the most difficult on because of the lack of verbose information in a text compared to an image or audio. Chapman et al. [13] stated text steganography uses written natural language to hide a secret message.

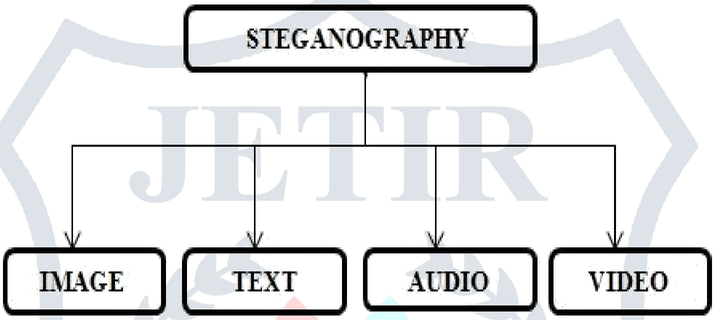

Fig. 2: Steganography techniques

IV. Text Steganography

Fig. 3: shows the basic text steganography mechanism. Initially, a hidden message will be covered in a cover-text by applying an embedding algorithm to produce a stego-text. The stego-text will then be transferred over a communication medium, e.g. Internet or mobile device to an intended recipient. At the receiver end the original message gets recovered from stego-text by applying stego key to the recovery algorithm. A stego-key is used to control the hiding process so as to restrict the detection and/or recovery of the secret data to intended receiver is controlled by usage of stego-key [5].

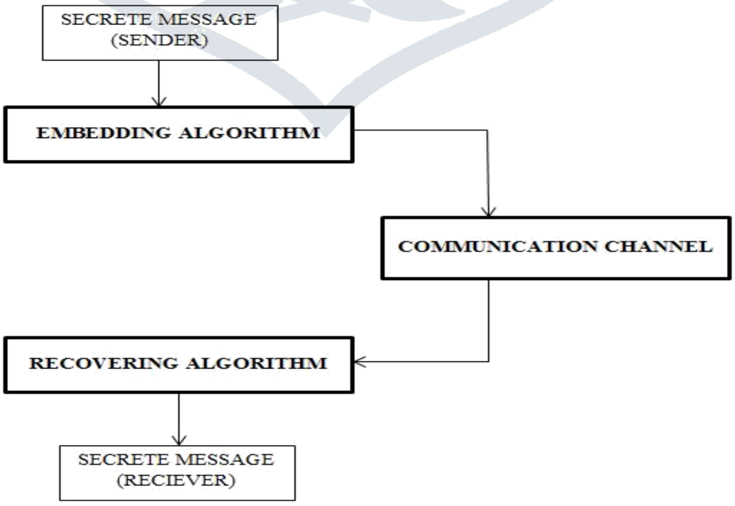

Fig. 3: Mechanism of text steganography

Text steganography can be grouped into three broad categories [14]

a) Format based

b) Random and statistical
c) Linguistic
And it is graphically represented in Fig. 4.

Format-based: Format based methods used to change the formatting of the cover text to conceal the data. They do not make any change of words or Sentences, so that it does not harm the originality of the cover-text. Open space method [15] is one of the format based technique where hidden message is added by entering extra white spaces in to the cover text. Two consecutive spaces are interpreted as ‘1’and a single space is interpreted as ‘0’. This technique can be applied to almost all types of text without revealing the existence of secret data. Line Shifting and word shifting are the well-known format based methods. In word shifting technique, the horizontal alignments of some words are shifted by changing distances between Words to hide the secret data [4].In line shifting method, vertical alignments of some lines of the text are shifted to create a concealed shape to embed a message in it [14].

Random and statistical generation: In this method generating cover text is according to the statistical properties. Character sequences and words sequences are considered in this method. The masking of information inside character arrangements is by embedding the information in random sequence of characters. This arrangement appears to be random to anyone who intercepts the message. A second technique to character production is to take the statistical features of word length and letter repetition in order to create words (without lexical value) which will appear to have the same statistical features as actual words in the given language.

Linguistic method: The linguistic method considers the linguistic properties of the text to make changes in it. In order to conceal the information the linguistic Structure of the message is utilized. Another method is Syntactic method in which punctuation signs like full-stop (.) and comma (,) are placed in proper places in the document to hide data. The proper identification of places where to insert the signs is needed in this method. Another linguistic steganography method is Semantic method where the synonyms of words for some preselected words are used. The words are superseded with their Synonyms to hide information in it.

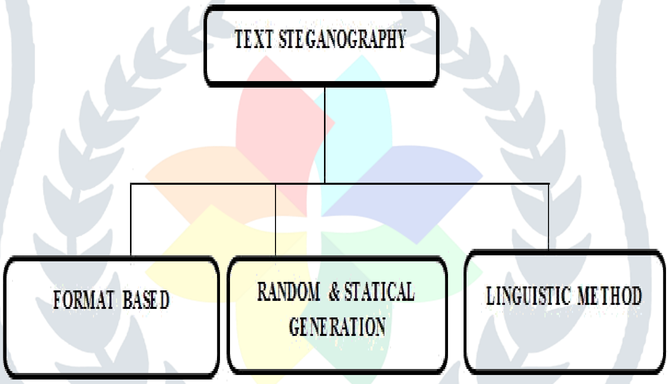

Fig.4: Text steganography methods

V. Regional Language Steganography Methods in India

Steganography techniques exist in Greek, Chinese, Persian, Arabic, Sindhi and other regional languages in different countries. Different Steganography methods are introduced in Oriya, Bengali, Telugu, Gujarati, Hindi, Punjabi etc. languages.

Ref. [16] takes Revised SSCE code and passkey which is used for concealing the data in Gujarati characters which accepts secret data in digital form. The steganography method [17] in Hindi proposes a numerical code to Hindi letters which is built on the basis of 4-bit binary value. Later each 4-bit code is replaced with a different word which starts with the respective letters that are assigned in the scheme. The method [18] for steganography in Telugu is achieved by using the linguistic properties of the Telugu language. This method is implemented by shifting the talakattu between the free spaces available to move in a Telungu scripts. A mechanism which takes input messages in any digital format and make use of the quantum truth table in the mapping method which utilizes the Oriya text as the cover text. Mapped value from the truth table is used for embedding process [19]. Ref. [20] Proposes LCS based Text Steganography in Indian Languages. Considering the reach of more characters and flexible grammar sentences of Indian Languages this method conceal the confidential message in the stego-text by creating meaningful sentences after finding the longest common subsequence of two binary strings among which one is the hidden message and another may be any binary string. The collection of the sentences created will be used as the cover media for the steganography approach. The hidden message will be recovered from stego-text by reverse method that is after acquiring the longest common subsequence of sentences from the stego-text and replacing the matched character by the bits of another binary string. Ref. [21] is based on the fact that Arab’s uses majority of letters of Urdu alphabets. On this basis, information was concealed in text by placing the Reverse Fatah. This method can be used in hidden exchange of message through text documents and text watermarking.

VI. CONCLUSION

This paper renders ideas on Steganography, predominantly text steganography. This paper explores the steganography techniques employed in regional Indian languages.

REFERENCES

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

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

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

[4] N. F. Johnson, S. Jajodia, "Exploring Steganography: Seeing the Unseen," IEEE Computer, pp.26-34, 1998.

[5] Fabien A. P. Petit colas, Ross J. Anderson & Markus G. Kuhn(1999), "Information Hiding A Survey," Proceedings of the IEEE-Special Issue on Protection of Multimedia Content, Vol.87, No. 7, Pp. 1062-1078.

[6] Kahn, The Code breakers - the comprehensive history of secret communication from ancient times to the Internet, 1996.

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

[8] T Markel, JHP Eloff and MS Olivier. " • An Overview of Image Steganography," proceedings of the fifth annual Information Security, 2005.

[9] Z.Duric N, F.Johnson and S. Jajodia Information hiding: Steganography and digital watermarking attacks and countermeasures, 2001.

[10] G. Doerr and J.L. Dugelay. "A guide tour of video watermarking," Signal Processing: Image Communication. Vol. 18, pp.263-282, 2003.

[11] G. Doerr and J.L. Dugelay. "Security pitfalls of frame by frame approaches to video watermarking," IEEE Transactions on Signal Processing, Vol.52, pp.2955-2964, 2004.

[12] S. Low N.F. Maxemchuk J.T. Brassil and L.O.Gorman, "Electronic marking and identification techniques to discourage document copying," IEEE Journal on Selected Areas in Communications, pp. 95-1504, 1995.

[13] Kran Bailey, Kevin Curran. An evaluation of image based steganography methods. 1999.

[14] Souvik Bhattacharyya, Indradip Banerjee and Gautam Sanyal, " A Novel Approach of Secure Text Based Steganography Model using Word Mapping Method (WMM)," International Journal of Computer and Information Engineering,2010.

[15] W. Sweldens. "The lifting scheme. A construction of second generation wavelets," SIAM J. Math. Anal., Vol.29, pp.511-546, 1997.

[16] Indradip Banerjee, Souvik Bhattacharyya, Gautam Sanyal, "Text Steganography Using Quantum Approach in Regional Language with Revised SSCE," I. J. Computer Network and Information Security, 2012.

[17] Megha Pathak,"A New Approach for Text Steganography Using Hindi Numerical Code," International Journal of Computer Applications (0975-8887), Vol.1,No.8 .

[18] Sravani Alameti et al "A new approach to Telungu text steganography by shifting inherent vowel signs," International Journal of Engineering Science and Technology ,Vol. 2 No.12 , pp.7203-7214, 2010.

[19] Indradip Banerjee, Souvik Bhattacharyya and Gautam Sanyal,"A Procedure of Text Steganography Using Indian Regional Language," I. J. Computer Network and Information Security,Vol 8, pp.65-73, 2012.

[20] S.Changder, D. Ghosh and N. C. Debnath,"LCS based Text Steganography through Indian Languages", Proceedings of the 3rd IEEE International Conference on Computer Science and Information Technology, 2010.

[21] Jibran Ahmed Memon, Kamran Khowaja, Hameedullah Kazi, "evaluation of steganography for Urdu/Arabic text," Journal of Theoretical and Applied Information Technology, 2005