TEXT STEGANOGRAPHY: REVIEW

Nitesh Rao M[1]
(M.Tech), Dept. of CSE
MCET, Osmania University.
Hyderabad, Telangana, India.
niteshkick@gmail.com

Lavanya Pamulaparty[2]
Associate Professor and HOD, Dept. of CSE
MCET, Osmania University.
Hyderabad, Telangana, India.
lavanya.post@gmail.com

***Abstract*—In the current era of Information Technology, unlawful replicating and illegal distribution accompany the adoption of far reaching electronic spreading of copyrighted material. This is the fundamental motivation behind why individuals consider how to secure their work and how to stop such unlawful activities. For this reason different strategies including cryptography, steganography, coding thus on have been utilized. Steganography is the most appropriate strategy that permit client to hide a message in another message (cover file). The greater part of steganography is it can utilize cover media as pictures, video clips and audio. However, text steganography is not ordinarily favored because of the trouble in finding excess bits in text document. To hide data inside a document its characteristics ought to be changed. These characteristics can be either the text organization or attributes of the character. In any case, the issue is that if slight change has been done to the document then it will get to be noticeable to the outsider or intruder. The way to this issue is that to change the document in a manner that it is just not noticeable to the human eye yet it is conceivable to unhide it with PC. For this reason different strategies for text based steganography have been proposed like line shifting, word shifting, feature coding, white space handling and so forth. In this paper, we exhibit an outline of the steganography, with a specific focus on text-based steganography.**

Keywords-component; formatting; style; styling; insert (key words)

I. INTRODUCTION

Security has dependably been a vital part of human. We are encompassed by a universe of secure communication, where individuals of different types are transmitting data such as credit card number to an online store than and as cunning as a terrorist plot to hijackers. The strategies that make secure communication practicable are not new. There has dependably been a need of securing the messages that are sensitive in nature. Such messages if presented to a few intruders may represent a risk to country's security or organization's basic choices. Therefore, such data must be secured at any expense and to fill the need to encrypt or hide the data. Cryptography (derived from Greek work 'kryptos' meaning hidden and 'graphein' meaning to write) [1] is utilized to encode the content to make it understandable. Cryptography may draw the suspicion of the intruder or third party towards the content that is in encoded. Steganography is the craftsmanship and exploration of composing concealed messages in such a way that nobody, aside from the sender and expected beneficiary, suspects the presence of the message, a type of security without knowledge of its presence. The word Steganography is of Greek origin and means "concealed writing" from the Greek words steganos (στεγανός) meaning "covered or protected", and graphein (γράφειν) meaning "to write" [2]. Steganography can be categorized based on the kind of media it utilizes to hide the data.

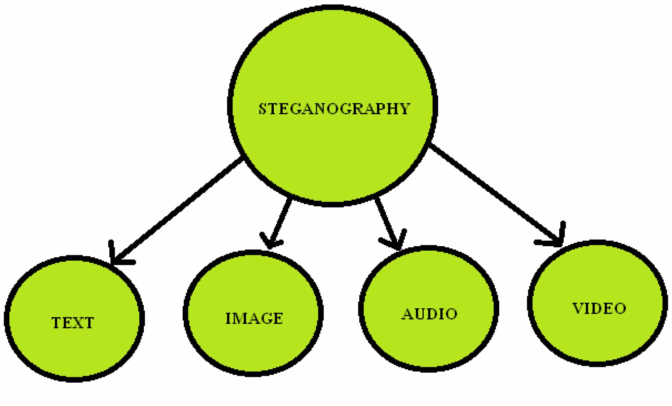

Fig.1. Types of Steganography

Text Steganography: It conceals the text behind some other text file. It is toughest type of steganography as the repetitive measure of text to hide the secret message is rare in text files.

Image Steganography: This type hides text or an image inside another text. It is the most frequently used strategy due to the restriction of the Human Eye.

Audio Steganography: Audio Steganography is a method used to transmit hidden data by adjusting a sound sign in an undetectable way. It is the science of concealing some secret text or audio data in a host message [3]. The host message before applying steganography and stego message after steganography have the same attributes.

Video Steganography: Video Steganography is the procedure of concealing some secret data inside a video. The expansion of this data to the video is not conspicuous by the human eye as the change of a pixel colour is negligible [3].

II. TEXT STEGANOGRAPHY

Text steganography utilizes text as the medium to hide data. It is the toughest sort of steganography; this is because of the relative absence of repetitive data in a text file as differentiated with a image or audio file. The structure of text documents is indistinguishable with what we observer, while in different sorts of documents, for example, in image, the structure of document is unique in relation to what we observe. Thus, in such documents, we can hide data by presenting changes in the structure of the document without forming a notable change in the required output.

Text steganography is broadly classified into the three categories; Format based Random and Statistical generation, Linguistic methods [3].

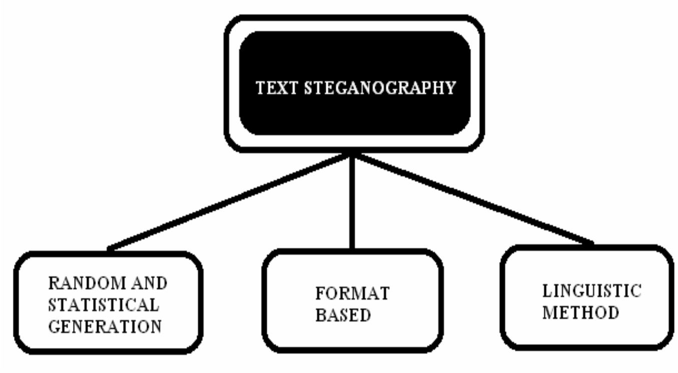

Fig. 2.Types of Text Steganography

2.1 Random and Statistical Generation

Keeping in mind to overcome the problem of comparison with a known plaintext, steganographers frequently produce their own particular cover texts [4]. The techniques used are - hiding data in irregular looking group of characters, the numerical properties of word length and letter frequencies are utilized as a part to make words which will seem to have same numerical properties as genuine words.

2.2 Format Based Methods

Format Based Methods [4] include adjusting the physical organization of text to hide the data. The drawback of this technique is, if the stego file is opened with a word processor, incorrect spellings and additional white spaces will get distinguished. Changed textual styles sizes can excite suspicion to an intruder. What's more, if the first plaintext is accessible, contrasting this plaintext and the suspected steganographic text would make controlled parts of the text entirely visible.

2.3 Linguistic Methods

Linguistic steganography [5] particularly considers the semantic properties of produced and adjusted text, and much of the time, utilizes etymological structure as the space in which messages are hidden. CFG make tree structure which can be utilized for hiding the bits where left branch speaks to "0" and right branch relates to '1'. A language structure in GNF can likewise be utilized where the main decision as a part of a generation speaks to bit 0 and the second decision speaks to bit 1. A few disadvantages of utilizing this technique are - a little sentence structure will prompt part of text redundancy, despite the fact that the text is grammatically impeccable, yet there is an absence of semantic structure. The outcome is a series of sentences which have no connection to each other.

III. DIFFERENT TECHNIQUES OF TEXT STEGANOGRAPHY

3.1 Line Shift

This technique [4,6] changes a document by vertically moving the position of text lines and might be connected to both the page picture and the file group. The codeword reassigned for a specific document determines the text lines that will be moved in that document. We may have a "0" for a line moved up and a "1" for a line moved down. Be that as it may, likewise we may have a "- 1" for a line moved up, a "0" for an unaffected line and a "+1" for a line moved down. This strategy utilizes the differential encoding method for accomplishing performance. Determination of whether the line has been moved up or down is done by measuring the separation of the centroid of the marked line and its control lines. The defect of this technique is if the text is retyped or if a OCR program is used, the hidden data would get destroyed.

3.2 Evaluation of Steganography for Urdu/Arabic Text

This technique proposed in [7] utilized for Arabic/Urdu watermarking. One of the attributes of Arabic dialect is the utilization of Araabs i.e. (Fatah, Kasra, and Damma). Where Fatha is slash like symbol and is composed over the character, while Kasra is likewise a slash like image however is utilized underneath the character and Damma is number nine like symbol which too is put over the character. These araabs are appropriate on each single character of the Arabic dialect. In the technique proposed they have utilized fatha in reverse order to hide the secret characters in the text. Fatha in reverse order is known as reverse fatha. For applying this technique on any Urdu text most importantly make a secret message which is most likely of one line barely 5 words containing 10-15 characters. Presently let us select an article of 4 to 5 paragraph or as the quantity of paragraphs is increased the security is also expanded. At that point we put araabs on the complete text or take an article having araabs already. After this read the secret message character by character and match it in the article, and we need to put reverse fatha where the secret characters exist consecutively attempt to utilize the reverse fatha not on the same line however on various lines. Presently the article to the receiver is prepared to be sent as a letter.

At the point when the receiver gets the letter the receiver will need to read the article carefully and tries to locate the reverse fatha in the letter and extract the characters accurately underneath the reverse fatha and gather them. At last when letter is totally read then concatenate the characters. At last when the characters are concatenated, it will give the secret message.

3.3 Arabic Text Steganography Method Using Letter Points and Extensions

In a technique proposed in [2007] is used to hide data in letters. They use the pointed letters with extension to hide secret bit "1" and the un-pointed letters with extension to hide secret bit '0'. Note that letter extension doesn't have any effect to the content. It has a standard character hexadecimal code: 0640 in the Unicode framework. Indeed, this Arabic extension character in electronic typing is considered as a redundant character only for arrangement and format purposes. The main advantage in utilizing the extension is that not all letters can be extended with this extension character because of their position in words and Arabic composition nature. The extension must be included between connected letters of Arabic text; i.e. extensions can't be put after letters at end of words or before letter at starting. The proposed steganography hypothesis is that at whatever point a letter can't have an extension or found without extension it is considered not holding any secret bits.

This proposed steganography technique can have the alternative of including extensions before or after the letters. To be steady, be that as it may, the area of the extensions ought to be the same all through the complete steganography document.

3.4 White Steg

This procedure utilizes white spaces for concealing a secret message [9]. There are three techniques for concealing information utilizing white spaces - Inter Sentence Spacing, End of Line Spaces, Inter Word Spacing procedure. In Inter Sentence Spacing, to conceal a bit 0, we put single space and to conceal bit 1 we put two spaces at the end of each ending character. In End of Line Spaces, settled number of spaces is embedded toward the end of every line. For instance, two spaces to encode one bit per line, four spaces to encode two bits etc. In Inter Word Spacing system, bit 0 is represented by one space after

Identify applicable sponsor/s here. *(sponsors)*

a word and bit 1 is represented by two spaces after a word. But, inconsistent utilization of white space is not straightforward.

3.5 Text Steganography Using Hindi Letters and Its Diacritics

Hindi Language obviously has a blend of letters and letter diacritics. It additionally has compound letters. Basically data is moved as bit streams. We can hide these bit streams Using Hindi Letters. These bit streams are 1's and O's. This technique encodes the bit O with vowel and consonant letters. In the same way bit 1 is hided by letter diacritics and compound letters. So absolutely here we have just two classifications [10].

1. Letters (which are pure vowels and consonants not vowel diacritics and consonant diacritic)
2. Letter Diacritics and Compound Letter (All Letters that are in Vowels and Consonants)

3.6 Word Shift

In this technique [4], by moving words horizontally and by changing distance, data is hidden up in the text, i.e. left or right to hide bit 0 and 1 respectively. This technique is worthy for texts where the distance between words is varying. Words shift can be identified using correlation techniques. This technique can be recognized less, because of the fact that change of distance between words to fill a line is very normal.

3.7 Text Steganography Technique Based on Html Documents

This technique [11] uses the html tags and their attributes to hide the secret message. In this technique essentially has three components viz. key file generation, hiding process and extracting process. The important part of this technique is the generation of key file. The key file a collection of key combinations stored in the form of rows and columns. These combinations are generated by scanning through the html documents. The attributes combinations used in the html tags are used to generate a key file.

The key file contains two types of attributes, corresponding to two columns:

1. Primary Attribute
2. Secondary Attribute

The hiding process scans each attribute of each html tag, and checks to see whether that attribute exists in the primary attribute field of the key file [11]. If **yes**, its corresponding secondary attribute is searched in the corresponding html tag. If found, then the combination of attribute is used to hide a bit. If not, skip the attribute. The hiding of a bit is determined by the order of the attributes in the attribute combination. If primary attribute is followed by a secondary attribute, it can hide a bit 1; else it can hide a bit 0.

The extractor program [11] extracts the message from stego text by first identifying the attribute combinations that hides a bit and then finding the bit corresponding to the order of those attributes. If primary attribute is followed by secondary attribute, a bit 1 is detected, else a 0 is detected.

3.8 Hiding Text in Images Using Steganography

In this technique [12] we hide text inside an image. This technique is a modified version of LSB techniques. LSB is very vulnerable and can easily give out the hidden message in it if the original image is found. To overcome that problem, this technique uses a unique but reliant strategy. We hide each character inside each pixel. In the hiding process we choose an image to hide the text. We convert the image to its RGB image and then

1. Replace red component of the first pixel with first character.
2. Replace green component of the second pixel with second character.
3. Replace blue component of the third pixel with third character.
4. And repeat iterations until pixels get exhaust.

The hidden message is extracted at the receiver side using the stego key.

3.9 Missing letter puzzle

Missing letter puzzle [13] is a puzzle comprising of a collection of words with one or more letters missing in each word. A letter, at certain position in a word, is missed by replacing it with a question mark. The puzzle is solved by replacing each question mark by an appropriate letter in each word so as to make the words meaningful. Words in a puzzle can be of different length and different domains, i.e., they can be terminologies of various fields or can be proper nouns or a combination of both. Each character of secret message is hidden in a word of certain length by missing one or two letters in it. Hint is also given for some words. Since the length of words depends on the (ASCII) value of characters to be hidden, the words are dynamically generated. There is no pre-determined cover file. The technique is designed in such a way that the stego file is formed without the cover file.

3.10 Hiding Data in Paragraphs

This technique [13] makes use of a pre-determined cover file. This technique works by hiding a message using start and end letter of the words of a cover file. This technique works on the binary value of a character. After converting the cipher text to a stream of bits, each bit is hidden by picking a word from the cover file and using either the start or the end letter of that word depending on the bit to be hidden. Bit 0 or 1 is hidden by reading a word, sequentially, from the cover file and including the starting letter or the end letter, respectively, of the word in the stego key. A word having same start and end letter is skipped. Since no change is made to the cover, the cover file and its corresponding stego file are exactly the same.

3.11 Mixed-Case Font

In this technique, the data can be hidden in English text utilizing the letters as bearers. This methodology will embed one character inside every 7 letters [14]. So the capacity limit will be high compared with other text steganography techniques.

3.12 Text Rotation Techniques in MS Excel

In this technique [15] we first convert the secret message into binary format and then choose each bit from the binary format sequentially. We then check if the bit is 1 or 0. If the bit is 1, we rotate a cell of the MS Excel file by 1 degree else we keep it to 0 degree. This process is repeated until all the bits are hidden. The choosing of the cell is also done sequentially. Once the hiding process is done the stego file is formed and is sent to the receiver. The receiver then checks for rotation of cells in the stego file sequentially. If the cell rotation is 1 degree the secret bit is 1 else the secret bit is 0. These 1 and 0 are put in another file and is then converted into character format to get back original secret message.

3.13 Font Type in MS Word

Before starting this technique[15] we create a resemble font array which contains a table of cover document font and their resembling fonts for assumption 15 type of cover document fonts and their resembling font. Create a code table that contains coding of each symbol in secret message represented by three types of fonts, thus, 27 characters (English alphabets with space) can be hidden in 3 letters of cover using 3 different fonts, for example: similar font array of Century font is : Century = {Century 751BT, Century Old Style, Century Expd BT}. Secret message is embedded in Capital letters of the cover text document. First step is to find the font of the cover document to get it's resemble fonts array. Secondly, scan the capital letters in cover text i.e., needed three capital letters to hide one symbol. Finally, choose the corresponding font type of character in secret message from code table. The final step repeats till all the secret message characters are hidden in the cover file to get the stego file. The stego file is then sent to the receiver where he applies the reverse process to get back the secret message.

3.14 Syntactic Method

This technique [16] utilizes punctuation marks, for example, full stop (.), comma (,), and so on to conceal bits 0 and 1. The issue with this technique is that it requires identification of punctuation at right places.

3.15 Cricket Match Scorecard

In this technique, [16]data is hidden in a cricket match scorecard by concatenating a meaningless zero before a number to represent bit 1 and leaving the number as it is to represent bit 0.

IV. CONCLUSION

In this paper we have given a brief introduction to Steganography and its types. This paper mainly deals with Text Steganographic Techniques. With this review can try to make more standard Text Steganographic Techniques.

REFERENCES

[1] Neha Rani and Jyoti Chaudhary, "*Text Steganography Techniques: A Review*", International Journal of Engineering Trends and Technology (IJETT) – Volume 4 Issue 7- july 2013

[2] Swati Gupta and Deepti Gupta, "*Text -Steganography: Review Study & Comparative Analysis*", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (5), 2011

[3] Chintan Dhanani and Krunal Panchal, "Steganography using web documents as a carrier: A Survey", International Journal of Engineering Development and Research (IJEDR), ISSN: 2321-9939, 2013

[4] S. Low, N.Maxemchuk, J.Brassil, L. O'Gorman, 1995. "*Document marking and identification using both line and word shifting*", Proceedings of the 14th Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM 95.

[5] Krista Bennett, "Linguistic Steganography: Survey, Analysis, and Robustness Concerns for Hiding Information in Text", CERIAS Tech Report 2004-13.

[6] Richard Popa, "*An Analysis of Steganographic Techniques*", The Politehnica University of Timisoara, Faculty of Automatics and Computers, Department of computer science and Software Engineering. 1998.

[7] Jibran Ahmed Memon, Kamran Khowaja, Hameedullah Kazi,"*Evaluation of Steganography for Urdu /Arabic Text*", 2005.

[8] Adnan Abdul-Aziz Gutub, and Manal Mohammad Fattani,"*A Novel Arabic Text Steganography Method Using Letter Points and Extensions*", 2007.

[9] 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.

[10] Mrs. Kalavathi.Alla, Dr. R. Siva Ram Prasad, "*A Novel hindi Text Steganography using Letter Diacritics and its compound words*",2008.

[11] Mohit Garg, "*A Novel Text Steganography Technique Based on Html Documents*", International Journal of Advanced Science and Technology, Vol. 35, October, 2011.

[12] Vipul Sharma and Sunny Kumar, "*A New Approach to Hide Text in Images Using Steganography*", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 4, April 2013.

[13] Monika Agarwal, "*Text Steganographic Approaches: A Comparison*", International Journal of Network Security & Its Applications (IJNSA), Vol.5, No.1, January 2013

[14] Abdelmgeid Amin Ali and Al - Hussien Seddik Saad, "*New Text Steganography Technique by using Mixed-Case Font*", International Journal of Computer Applications (0975 – 8887) Volume 62– No.3, January 2013

[15] Sumathy Kingslin and N. Kavitha, "Evaluative approach towards Text Steganographic Techniques", Indian Journal of Science and Technology, Vol 8(29), 2015.

[16] K. Aditya Kumar , Dr. Suresh Pabboju and Neela Megha Shyam Desai, "Advance Text Steganography Algorithms: An Overview", International Journal of Research and Applications Jan-Mar © 2014 Transactions; 1(1): 31-35

AUTHORS PROFILE

Nitesh Rao M[1],

M.tech(CSE)(Pursuing) , Dept of CSE, MCET, Osmania University, Hyderabad, Obtained his Bachelor's degree in computer science from KITE, Jawaharlal Nehru Technological University, Hyderabad, India. His research interest is Information Security

Lavanya Pamulaparty[2],

Associate Professor and Head of the Dept. Dept of CSE, MCET, Osmania University, Hyderabad, obtained her Bachelor's degree in computer science from Nagpur University of KITS, Nagpur, India, and Masters Degree in Software Engineering from School of Informatics from JNT University Hyderabad, India, and Pursuing the PhD degree in computer science and engineering from JNT University. Her research interests include information storage and retrieval, Web Mining, Clustering technology and computing, performance evaluation and information security. She is a senior member of the ACM, IEEE and ISTE.