Scientific Journal of Impact Factor (SJIF): 4.14
e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 3, Issue 4, April -2016

Practical Evaluation and Comparative Study of Text Steganography Algorithms

Mr. Sailesh .S. Iyer [1],. Dr. Kamaljit Lakhtaria[2]

[1] S.K Patel Institute of Management and Computer Studies, Gandhinagar.
[2]Rolwala Computer Centre, Gujarat University, Ahmedabad.

***Abstract*—*Text Steganography is an emerging technique in the field of Information Exchange. Many Text Steganography Algorithms have been proposed. This paper provides a practical evaluation of Text Steganography methods/algorithms like Mixed Case Font, Text Rotation and Font Type with respect to primary parameters like Security, Size, Robustness and Embedding Capacity. Similarity comparison between cover text i.e. original text or message and Stego text through Jaro Winkler distance also gives an indication of whether the Stego text can actually embed the text. Text Steganography Tools are also analyzed and studied with the intention of improving upon the existing tools*.**

***Keywords*—*Text Steganography, Practical Evaluation, Algorithms, Mixed Case Font, Text Rotation, Font Type, Security, Size, Robustness, Embedding , Stego Text, Jaro Winkler Distance*.**

I. INTRODUCTION (INFORMATION HIDING & *STEGANOGRAPHY*)

Information Hiding has been practiced in various forms. Steganography, one of the major forms of Information Hiding, has been prevalent since centuries. Steganography can be classified as given below:

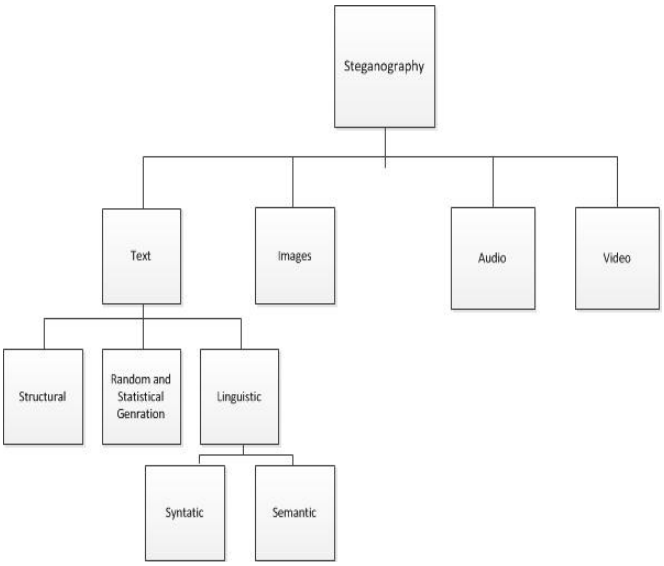

Fig.1 Classification of Steganography.

Natural language cover texts must not only pass the statistical test of automatic analysis, but also the minds of human readers. Linguistic Steganography is the least focused areas as major focus lies on Digital Watermarking and techniques of Audio, Video and Image Steganography. Steganography consists of components like secret message to be hidden, hiding image or text (stego), method to hide text in stego and secret key to decrypt the hidden text.

Text Steganography has much scope as most of the work in this area is still at its infancy. Text Steganography deals with hiding text in text. The primitive work done include shifting letters within the sentence, applying synonyms, including spaces in text etc.. Many techniques or algorithms have been suggested and research papers published on Text Steganography. However most of these papers are not able to prove the practical quantum as the major parameters of Text Steganography are not satisfactory. The parameters like Security, Size, Robustness and Embedding capacity etc. are major criteria on which each algorithm should adhere to.

- **Security**: Security is the property in which a person should be unable to distinguish the original and the stego-image.
- **Size:** A major factor to be considered is that the size of the Text/Image should not increase or decrease i.e. the original image/text size and the image/text should not have a large difference.
- **Embedding Capacity:** The large message should be embedded in such a way that the quality of image is intact and the size of the Image/Text is not considerably changing.

  Capacity Ratio= (Amount of Hidden bytes)

Size of Cover Text

- **Robustness:** Refers to the degree of difficulty required to destroy embedded information without destroying the cover image.
- **Similarity Measure:** Detection of similarity of Cover text and Stego text can be found by measuring Jaro Winkler distance. If Jaro Winkler distance is 1, then Cover and Stego text are similar else if distance is 0, then both text are dissimilar. Jaro Score can be calculated as:

  1/3 ***** (mc/ length (cs) + mc / length (ss) + (mc - t) / mc)

  Where, mc is the number of matching characters, cs is the first string, ss is the second string, t is the number of transpositions.

  Jaro-Winkler Distance [21] is:

  Jaro_score + (L ***** p ***** (1 - Jaro_score)) [22]

  where L is the length of the common prefix at the start of the string up to a maximum of 4, P is the constant scaling factor (usually 0.1 and not more than 0.25)

II. Related Paper Summary:

2.1 Frequency of Letters:

In [1], a technique based on combination of Vedic Numeric Code and frequency of letters in English alphabet is used. Frequency of letters in English vocabulary is used as basis of assigning numbers to letters in English alphabet.

The challenges are on two dimensions:

1. To hide a large message, larger number of words is required. e.g. 25 words are required to hide a 13 letter word.
2. The size of the file changes considerably.

2.2 Word Shift/Line Shift:

In Line Shift method, vertical shifting of Line by a particular angle to hide a 0 or 1 bit. To hide bit 0, a line is shifted up and to hide bit 1, the line is shifted down.

In Word Shift method, the secret message is hidden by shifting the words horizontally. This method is difficult to detect as normally it is quite natural to have unequal spacing or distance between words.

2.3 Semantic technique involves substituting synonyms [27] i.e. similar meaning words comprising nouns, adjectives, verbs etc.. An adjective or a group of adjectives approach [6] containing a particular secret message which will only be decoded by the intended receiver. At both the sender and the intended receiver they will have a database or a table mapping a particular adjective or a group of adjectives to a unique secret message that need to be sent over the unsecured channel.

The greatest challenge in this method is that generating such a database of adjectives is a tedious task.

2.4 In Text Steganography by Hiding Information in Specific Character of Words [17] approach, specific characters from some particular words are selected to hide the information. e.g. The first character of every alternative word hides the secret message.

2.5 Inter Word spacing method [2] was proposed to achieve the objective of reducing the size of objects created using Steganography. Experiments suggest that size reduction of objects is achieved but if the spaces are deleted then the message will be lost. One of the drawbacks of this method is that any alteration in spacing would leave this method ineffective.

2.6 Linguistic Steganography by context based substitution has been suggested [18]. Experiment results of the blind test demonstrate that the substitution can hardly generate syntax errors or unsuitable words, which implies the concealment of Steganography. This method has a relatively low embedding bit rate. Work can be extended to include imperfect filters, incomplete vocabulary, small scale dictionaries and limited training materials.

2.7 The proposed new Steganography method that uses a statistical compression technique called 'arithmetic coding'. This method has not been tested for long messages. This method has been tested for short messages which are proved by results of the experiments. Out of range messages are produced for large messages.

2.8 Text Rotation method is used in Excel to rotate the cell content by 1° if text and -1° if numeric. The cell text length is the key as if the cell length is 4 or less than 4, detection becomes very hard. If the text length is greater than 4 then it becomes easier to detect change in Stego text from Cover text.

Algorithm	Pros	Cons
Word Spelling	Hidden data is not destroyed	Spelling mistake will convey wrong meaning.
Word Shift	Difficult to crack as distance between words may not be uniform	• Prior knowledge of correlation or distances method can expose this way of hiding text. • Retyping text can remove hidden message.
Synonym Substitution	Exact meanings or similar words are difficult to find out.	Replacing words requires time and effort.
White Steg	Spacing between words etc. cannot be easily detected as it common to have one or two additional spaces.	Too much use of spacing can lead to suspicion.

Table-1. Pros and Cons of Text Steganography Methods/Approaches

III. PERFORMANCE ANALYSIS

STEGANOGRAPHY METHODS	**ATTRIBUTES**				
	DOMAIN	***EMBEDDING CAPACITY***	***ROBUSTNESS***	***IMPERCEPTIBLY***	***INTEGRITY***
Null Spaces	**Text**	**Low**	**High**	**Low**	**Low**
Synonym Substitution	**Text**	**High**	**Low**	**High**	**Low**
Context Based Equivalent Substitution	**Text**	**Low**	**Low**	**High**	**Low**
Frequency of Letters	**Text**	**Low**	**Low**	**Low**	**Low**
White Steg	**Text**	**High**	**Low**	**High**	**Low**
Reflection Symmetry	**Text**	**High**	**High**	**Low**	**Low**
Text Rotation Techniques	**Text**	**Very High**	**High**	**High**	**High**
Mixed Case Font	**Text**	**Very High**	**High**	**Medium**	**High**
Font Type	**Text**	**Very High**	**High**	**Very High**	**High**

Table-2. Comparative Analysis of Text Steganography

On comparison of above Text Steganography methods, it was found that:

i) Null Spaces was easy to detect and has less embedding capacity. The Jaro Winkler measure was close to 0.8 but since all the other parameters, this method does not qualify and hence this method cannot be considered effective method for text Steganography based on the parameters mentioned above.

ii) Synonym Substitution has good embedding capacity but can be detected making it less robust.

iii) Context based Equivalent Substitution also has low acceptability on most of the parameters.

iv) Selection of specific characters also does not meet all the required parameters.

v) White Steg has high embedding capacity but has low robustness and Integrity.

vi) Reflection Symmetry has high Embedding capacity but can be detected quite easily.

Our main comparison is based on the last three methods namely Text Rotation, Mixed Case Font and Font type.

Text Rotation Algorithm [23**]:**
Embedding Algorithm:
Input: MS Excel document, Limit p, secret bits
Output: stego-text
Body:
For each non-empty cell G do.
Get the selected cell's length n.
If n < p and secret bit is 1 then
If the type of G is text then Rotate the angle of G to 1°

Else If the type of G is numeric then Rotate the angle of G to -1°.
Output the embedded document.

Extracting Algorithm:
Input: stego-text, Limit p
Output: secret bits
For each non-empty cell G do
Get the selected cell's length n.
If n < p then
If the angle of selected cell is 1o or -1o angle then Secret bits = 1.
Else If the angle of selected cell is 0o angle then Secret bits = 0.
Convert the secret bits into ASCII value and Output the secret message.

Criteria	Text Rotation
Embedding Capacity	Short length text in cells increases the Embedding Capacity.
Invisibility	1° rotation doesn't make any difference so it is hard to detect through Human vision. HIGH
Undetectability	Hidden effect was relatively poor when limit increases. But if Limit is less than equal to 4 it gives high Hidden effect. HIGH
Robustness	High

Table-3. Detailed Performance Analysis of Text Rotation

Rotation in text is very difficult to detect but as the size of text increases, display of stego-text becomes harder.

Mixed Case Font Algorithm [25]:
Embedding Algorithm:
Input file: Text file T, Secret Message M.
Output file: Stego Text S.
Choose a text file T.
Divide T into letters, T={T1, T2,----- Tn}.
Get the secret message M.
Convert secret messge M into stream of bits b.
Divide b into bits, b={b1, b2,----bn}.
Select Ti from T and bi from b.
IF the bi is 'one' then change T; case into capital else change Ti case into small.
Repeat step 6, 7 till the whole b is hidden.
The resultant file will be the stego text S.

Extracting Algorithm:
Get the stego text T, and an empty array S for to store secret bits.
Divide T into letters, T = { T1, T2,----- Tn}.
If Ti case is in capital letter then include bit 1 in Sith index else if Ti case is in small letter then include bit 0 in Sith index.
Convert the secret bit S into ASCII value to get the secret message.
The resultant message will be the secret message.

Criteria	Mixed Case Font
Embedding Capacity	This approach will insert one character within each 8 letters. So the hiding capacity will be very high compared to other text Steganography methods.
Invisibility	Alteration depends on number of 1 bit in secret bits stream. - MEDIUM
Undetectability	The stego text will attract no attention because it will look like the "cool fonts" used in chat rooms and presentations
Robustness	Hidden information cannot be destroyed when stego text is enlarged or reduced - HIGH

Table-4. Performance Analysis of Mixed Case Font

A large volume of information can be embedded in text with comparison to other methods.

Font Type Algorithm [24]:

Embedding Algorithm:

Open cover document, find its type of font.
Scan cover document to find capitals English letters,
compute number of capitals English letters to check the capability of embedding.
For each symbol in secret message.
Retrieve its code.
Change font type of three capitals letters by resembling font array according to its code.

Extracting Algorithm:

Each three capitals letters; determine the code of one hiding symbol.
Open Stego document.
For each three capitals letters.
Determine the code.
If the code is (0, 0, 0), then the end of secret.
Message was reached.
Else find corresponding secret symbol, using code table.
First find the font name of the cover text and then similar font.

In **Text Rotation** method, there is no change in cover text which means that the Cover and Stego Text are the same. In this case the Jaro Winkler distance is **1**. The embedding capacity in Text Rotation method is **3.25%.**

In **Mixed Case Font Method**, Jaro Winkler Distance is **0.75.** Hence this method there is minor difference between Cover Text and Stego Text. The embedding capacity of Mixed Case Font Method is **10.3%.**

In the **Font type method**, Jaro Winkler distance is **1** as there is no change in the Cover and Stego text. The embedding capacity percentage of Mixed Case Font Method is app. **2.95%.**

Table-5. Performance Analysis of Font Type Method.

Criteria	Font Type
Embedding Capacity	Three characters to hide one character of secret message (one symbol in three capitals letters). This method has good perceptual transparency based on font type's resemblance. It has VERY HIGH capacity.
Invisibility	Extremely High.
Undetectability	Hidden effect is HIGH since resemblance fonts are used to hide secret message
Robustness	The stego document will not change during compression, copying and paste between computer programs, the data hidden in texts remains intact during these operations. Hence HIGH

IV. CONCLUSION

This comparative study has been performed based on more than 27 papers in addition to various other hybrid methods involving a combination of these papers.

Text Rotation Method, Mixed Case Font Method and Font type Method have been considered for detailed performance analysis. On comparison, the results indicate that Text Rotation and Font type Method are very good as far as similarity between Jaro Winkler distance is 1 which means that the Cover text and Stego text are similar. However the Embedding Capacity is approximately around 3.25% and 2.95% which means that large text cannot be embedded.

Mixed Case Font Method has great embedding capacity at 10.3% which means large text can be embedded but Jaro Winkler distance is 0.75 which means that Cover text and Stego text is not fully similar.

Out of all the text steganography methods, practical experiments have revealed that no algorithm satisfies all the criteria stated above.

This paper has not taken into consideration Genetic Algorithms, SMS text based algorithms, online credit card based algorithms and particular regional language based algorithms like Arabic, Tamil, Telegu etc..

V. FUTURE WORK:

Text Steganography Algorithms which satisfy all the above parameters should be suggested and practically implemented supported by tools like MATLAB, SCILAB etc.. The results obtained should be compared with existing algorithms.

Many new dimensions towards Text Steganography and Steganalysis can be obtained through combination or improvement of existing algorithms and rigorous testing on parameters like Robustness, Size, Capacity, Integrity and Vulnerability.

REFERENCES

[1] Souvik Roy & P. Venkateswaran G., "A Text based Steganography Technique with Indian Root", International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013 published in Procedia Science Direct.

[2] Vishal Ranjan and Shailja Bagdi, "An approach to reduce the size of secret text and a secure text Steganography", International Conference on Electrical, Electronics and Computer Engineering 2013.

[3] Lakhtaria Kamaljit .I. "Protecting computer network with Encryption technique: A Study, "Ubiquitous Computing and Multimedia Applications. Springer Berlin Heidelberg 2011, 381-390..

[4] Lakhtaria Kamaljit .I., ed. Next Generation Wireless Network Security and Privacy, IGI Global 2015..

[5] Nitin Kaul and Mrinal Chandra,"A Proposed Algorithm for Text in Image Steganography based on Character Pairing and Positioning", International Journal of Computer Applications (0975 – 8887) Volume 126 – No.3, September 2015.

[6] Tatwadarshi P. Nagarhalli, "A New Approach to Text Steganography Using Adjectives", International Journal of Innovative Research in Science,Engineering and Technology Vol. 4, Issue 5, May 2015.

[7] K. Aditya Kumar, Dr. Suresh Pabboju and Neela Megha Desai,"Advance Text Steganography Algorithms: An overview", International Journal of Research and Applications Jan-March 2014.

[8] Abhishek Kolugiri, Sheikh Ghouse and Dr. P. Bhaskara Reddy, "Text Steganography Methods and its tools", International Journal of Advanced Scientific and Technical Research, March- April 2014.

[9] Ammar Odeh, Khaled Elleithy and Miad Faezipour, "Steganography in Text by using MS Word Symbols", Proceedings of 2014 Zone 1 Conference of American Society for Engineering Education.

[10] Reihane Saniei, Karim Faez, "The Capacity of Arithmetic Compression Based Text Steganography Method", 2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP).

[11] Baharudin Osman, Roshidi Din, Tuan Zalizam Tuan Muda, Mohd. Nizam Omar, "A Performance of Embedding Process for Text Steganography Method", Recent Advances in Computer Science.

[12] Esra Satir, Hakan Isik, "A compression-based text steganography method", The Journal of Systems and Software 85 (2012) 2385– 2394.

[13] Prem Singh, Rajat Chaudhary and Ambika Agarwal, "A Novel Approach of Text Steganography based on null spaces", IOSR Journal of Computer Engineering (IOSRJCE) Volume 3, Issue 4 (July-Aug. 2012), PP 11-17.

[14] Indrajit Banerjee, Souvik Bhattacharya and Gautam Sanyal, "A Procedure of Text Steganography Using Indian Regional Language", I. J. Computer Network and Information Security, 2012, 8, 65-73.

[15] Sharon Rose Govada, Bonu Satish Kumar , Manjula Devarakonda and Meka James Stephen, "Text Steganography with Multi level Shielding, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3, July 2012.

[16] Anandaprova Majumder, Suvamoy Changder, "A Novel Approach for Text Steganography: Generating Text Summary using Reflection Symmetry", International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013.
[17] T. Moerland "Steganography and Steganalysis", May 15, 2003, www.liacs.nllhome/tmoerlaniprivtech. pdf.
[18] Fei Wang, Liusheng Huang, Zhili Chen, Wei Yang, Haibo Miao, "A Novel Text Steganography by Context-Based Equivalent Substitution", IEEE 2013.
[19] Sahil Kataria, Kavita Singh, Tarun Kumar, Mahendra Singh Nehra, "ECR(Encryption with Cover Text and Reordering)
based Text Steganography", IEEE 2nd Intl. Conf. on IIP 2013.
[20] Reihane Saniei, Karim Fiez, "The Capacity of Arithmetic Compression Based Text Steganography Method", 8th Iranian Conference on Machine Vision and Image Processing (MVIP).
[21] Jaro-Winkler distance. 2015 Jan. Available from: http:// en.wikipedia.org/wiki/Jaro%E2%80%93Winkler_distance
[22] Approximate String Matching. 2015 Jan. Availaible from: http://biostat.mc.vanderbilt.edu/wiki/Main/ApproximateString Matching.
[23] Yang B, Sun X, Xiang L, Ruan Z, Wu R. Steganography in
Ms Excel Document using Text-rotation Technique. Information Technology Journal. 2011.
[24] Bhaya W, Rahma AM, Al-Nasrawi D. Text Steganography based on Font Type in Ms-Word Documents. Journal of Computer Science. 2013.
[25] Ali AA, Al–Hussien S. New Text Steganography Technique by using Mixed-Case Font. OJCSIT. 2013; 3(2):138–41.
[26] Monika Agarwal, "Text Steganographic Approaches: A Comparison", International Journal of Network Security & Its Applications.
[27] M. Hassan Shirali- Shahreza and Mohammad Shirali-Shahreza, "New Synonym Text Steganography", International Conference on Intelligent Information Hiding and Multimedia Signal Processing

