International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

# Text Steganography using Helping Verbs

**Rahul Sanwal[1], Rishabh Jain[2]**

[1, 2]Bachelor of Technology, Department of Information Technology, Galgotias College of Engineering and Technology Greater Noida, affiliated to Uttar Pradesh Technical University, Lucknow, India

**Abstract:** ***Steganography is the art and science of hiding a message inside a cover message. This message can only be detected by the intended recipient of the message. Steganography can be classified into image, text, audio and video steganography, this classification is based on the type of cover media used to hide the message. Text steganography can involve anything from changing the formatting of an existing text, to changing words within a text, to generating random character sequences or using context-free grammars to generate readable texts. The problem with text steganography is that if slight change has been done to the document then it will become visible to the third party or attacker. The key to this problem is that to alter the document in such a way that it is simply not visible to the human eye yet it is possible to decode it with computer. In our proposed approach we hide the message in the cover text using the concept of apostrophe and helping verbs. Firstly, the message is encrypted using one time pad technique and then converted into tertiary code. After this the cover text is searched for any helping verbs which would match the ones given in the table which stores the various helping verbs and their apostrophe form. If we want to store '0' we don't modify the cover text by keeping the helping verb in its original form, to hide '1' in the cover text we convert the helping verb in the cover text into its apostrophe form and to hide '2' The helping verb is converted into its apostrophe form and concatenated with the previous word but with a space in between. This process goes on again and again until all the message bits are hidden in the cover text.***

**Keywords:** Steganography, Helping Verbs, Apostrophes, Cover Text.

## 1. Introduction

Steganography is the art and science of hiding messages in such a way that no one, apart from the sender and intended recipient, suspects the existence of the message. The word steganography is of Greek origin and means "concealed writing." It combines the Greek words steganos (στεγανός), meaning "covered or protected," and graphei (γραυή) meaning "writing." The first recorded use of the term was in 1499 by Johannes Trithemius in his Steganographia. In steganography, the hidden messages are embedded in something else: images, articles, shopping lists, or some other cover text. For example, the hidden message may be in invisible ink between the visible lines of a private letter.

The main advantage of steganography over cryptography is that the secret message does not attract attention to itself. Plainly visible encrypted messages attract unwanted attention of everyone viewing it. Thus, cryptography is used to protect the contents of the secret message whereas steganography on the other hand, conceals the fact that a secret message is being transferred altogether.

In our paper we would be proposing a new technique in text steganography. We would be hiding message in text using helping verbs and their apostrophe form. For increasing the security of the transmitted message we would be using one time pad technique to encrypt the message. This cipher text is further converted into tertiary code. This tertiary code will be hid in the cover text with the help of the helping verbs and their apostrophe forms. The extraction process involves first procuring the hidden tertiary code using the stored table of helping verbs and there full forms and converting the tertiary code into cipher text. This Cipher text obtained is decrypted using the key.

## 2. Related Work

### A. Format Based Steganography

As mentioned by K. Bennett in[6], format-based text steganography methods modify the given cover text to hide the secret message. These modifications include deliberate insertion of spaces, misspellings throughout the text, changing font sizes of text etc. Each of these methods have their own advantages and disadvantages like- deliberate insertion of spaces might be missed by human eye but a computer will be able to easily detect it, on the other hand, font resizing would not be considered a problem by the computer but human eye will be able to detect discrepancies in size of given text.

### B. Text Steganography by Changing Words Spelling

Shirali-Shahreza in[4] present a new text steganography method for hiding data in English texts. This method uses difference in US-English and UK-English spellings of certain words to hide the information in cover text. In English some words have different spelling in UK and US. For example "dialog" has different terms in UK (dialogue) and US (dialog). So data can be hidden in the text by substituting these words.

### C. Text Steganography in SMS

Mohammad Shirali-Shahreza and M. Hassan Shirali-Shahreza in[5] have suggested the substitution of words with their abbreviations or viza viz to hide bits of secret message.

### D. The Word Shift Coding

As described by Richard Popa in[13], this method modifies the cover text by shifting the words horizontally. This technique is feasible only when the spaces between adjacent words are different. Generally, variable word spacing is used to distribute white space when justifying a document. As variable spacing is used in this technique, the decoder needs the original document or a specification about word spacing in the original document.

## 3. Proposed Approach

This approach is based on the apostrophe forms of helping verbs used in English language. Initially, one time pad cipher encryption technique has been used to encrypt the message. The original message and the encrypted message can consist of alphabet only . Now, this encrypted message is converted into tertiary code starting from “a” taken as “000” till “z” taken as “221”(Refer to table 3 for the tertiary code of all letters). The result is a string of numbers which would represent the encrypted message. This string of numbers is used to modify the cover text wherever helping verbs occur.

Each helping verb in the cover text is used to represent a number from the string of numbers and the helping verb is modified accordingly. Each helping verb can be modified in 3 ways according to the number in the string which it represents-
0- The helping verb remains unchanged.
1- The helping verb is converted into its apostrophe form and concatenated with the previous word.
2- The helping verb is converted into its apostrophe form and concatenated with the previous word but with a space in between.(Refer to table 2)

Thus after modifying the helping verbs present in the cover text according to the string of numbers we obtain the stego text.

### A. Hiding Algorithm

- Take the message to be hidden and encrypt it using one time pad encryption technique.
- Convert the cipher text into tertiary code using the codes of letters as given in table 3.
- Now search the cover text for any apostrophe form of helping verbs and convert them back into their original form (without apostrophe).
- Modify each occurrence of helping verb with respect to corresponding number in the string of numbers which we would obtain after step 2.
- Helping verbs can be modified in 3 ways according to number which they represent. (Refer to table 2).
- The obtained text after modifying helping verbs is the stego text.

### B. Units

- Search the stego text created after the hiding process for occurrences of an apostrophe or any of the helping verbs given in the table 1.
- For each such occurrence compute the tertiary value associated with it(Refer to table 2)
- Convert the obtained string of numbers to cipher text by taking block of 3 numbers at a time (Refer to table 3).
- Use the one time pad key to decrypt it into the original message.

**Table 1:** Helping verbs and their apostrophe forms

| Normal Form | Apostrophe Form |
|---|---|
| is | „s |
| Us | „s |
| will | „ll |
| have | „ve |
| had | „d |
| shall | „ll |
| would | „d |
| should | „d |
| not | „t |
| am | „m |
| are | „re |
| could | „d |

**Table 2:** Modification of helping verbs w.r.t Tertiary code

| 0 (Original Form) | 1 (Apostrophe form without space) | 2 (Apostrophe form with space) |
|---|---|---|
| “Have” | “He‟ve” | “He „ve” |
| “Had” | “He‟d” | “He „d” |
| “Will” | “It‟ll” | “It „ll” |
| “Should” | “He‟d” | “He „d” |
| “Not” | “Doesn‟t” | “Doesn „t” |
| “Is” | “He‟s” | “He „s” |
| “Am” | “I‟m” | “I „m” |
| “Are” | “You‟re” | “You „re” |
| “Shall” | “He‟ll” | “He „ll” |

**Table 3:** Tertiary code of alphabets.

| Alphabets | Number Code used |
|---|---|
| A , a | 000 |
| B, b | 001 |
| C, c | 002 |
| D, d | 010 |
| E, e | 011 |
| F, f | 012 |
| G, g | 020 |
| H, h | 021 |
| I ,i | 022 |
| J, j | 100 |
| K, k | 101 |
| L, l | 102 |
| M, m | 110 |
| N, n | 111 |
| O, o | 112 |
| P, p | 120 |
| Q, q | 121 |
| R, r | 122 |
| S, s | 200 |
| T, t | 201 |
| U, u | 202 |
| V, v | 210 |
| W, w | 211 |
| X, x | 212 |
| Y, y | 220 |
| Z, z | 221 |

## 4. Example

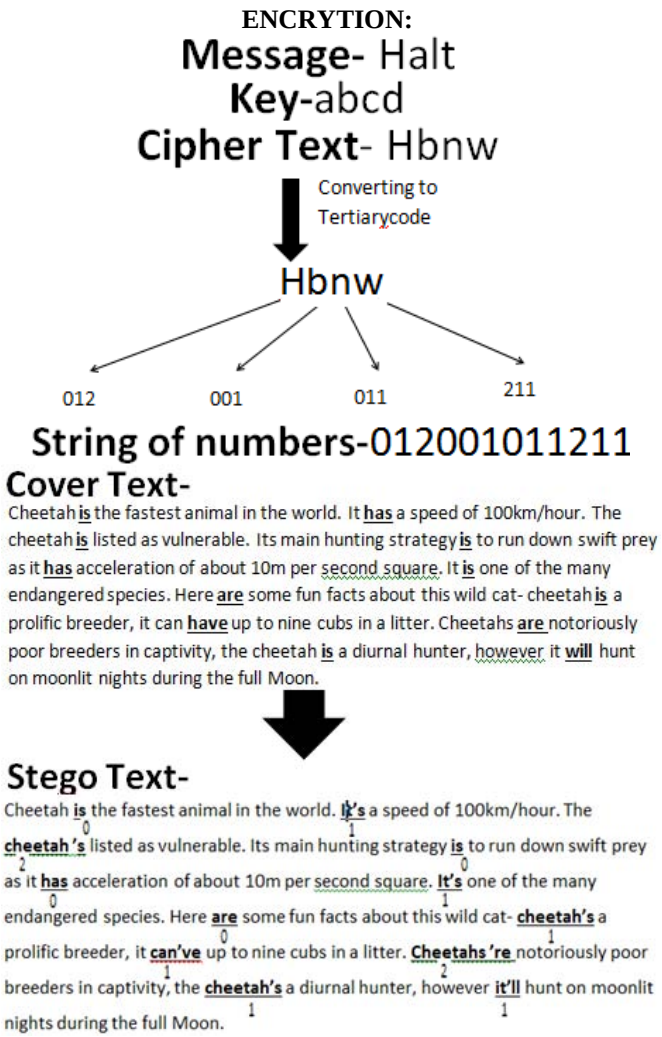
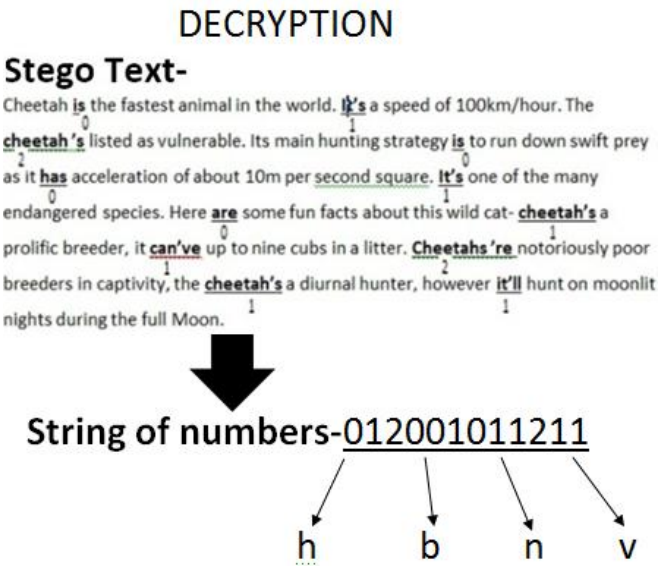



**Figure 1:** Hiding Process



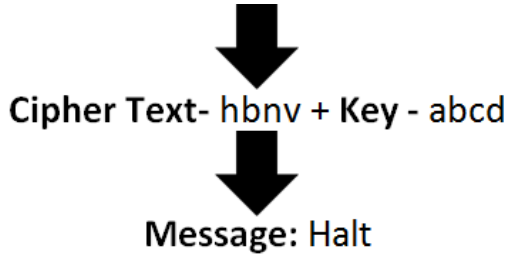




**Figure 2:** Unhiding process

## 5. Results

We have developed software to implement this steganography technique. It counts the number of helping verbs and apostrophe in the cover text before starting the hiding process which enables us to determine if the size of cover text is sufficient to hide the given message. The sources of the cover text are various articles on google news.

**Table 4:** Tabular representation of results

| Number of words in cover text | Message | Number of bits to be hidden |
|---|---|---|
| 309 | george | 18 |
| 349 | george | 18 |
| 394 | george | 18 |

According to results shown in above table(Table 4) the average number of words required in cover text to hide 6 letter word is 351 words.

## 6. Conclusion

This paper presents a new text steganography method for hiding data in English text. This method modifies the cover text by changing the helping words into their apostrophe form wherever needed in the cover text according to the message to be hidden. In English, helping verbs can be written using the apostrophe techniques. The frequency of occurrence of helping verbs is very high. Therefore less amount of cover text would be required to hide the message as observed in results provided above. We have provided higher level of security by using an encryption technique to encrypt the message before hiding it.

## 7. Acknowledgment

We owe deep gratitude to the ones who have contributed greatly in completion of this paper. At the outset, we would like to express my gratitude to our guide **Mr. Manoj Kumar Beniwal, Assistant Professor, Department of Information Technology**. As our supervisor, his constant inspiration helped us to remain focused on achieving our goal. His observations and comments helped us to establish the overall direction of our project research and to move forward with investigation in depth.

## References

[1] Chen Zhi-li, Huang Liu-sheng, Yu Zhen-shan, Zhao Xin-xin and Zheng Xue-ling, "Effective Linguistic Steganography Detection" , in IEEE 8th International

Conference on Computer and Information Technology Workshops.
[2] Hitesh Singh, Pradeep Kumar Singh and Kriti Saroha, " A Survey on Text Based Steganography " , in Proceedings of the 3rd National Conference; INDIACom-2009.
[3] KHAN FARHAN RAFAT, "Enhanced Text Steganography in SMS", in IEEE transaction, 2008.
[4] Mohammad Shirali-Shahreza, "Text Steganography by Changing Words Spelling", in IEEE journal Feb. 17-20, 2008 ICACT 2008.
[5] Mohammad Shirali-Shahreza, "Text Steganography in SMS", 2007 International Conference on Convergence Information Technology.
[6] K.Bennet, "Linguistic Steganography: Survey, Analysis, and Robustness Concerns for Hiding Information in Text", Purdue University, CERIAS Tech Report 2004-13
[7] Monika Agarwal," TEXT STEGANOGRAPHIC APPROACHES: A COMPARISON", in International Journal of Network Security & Its Applications (IJNSA), Vol.5, No.1, January 2013.
[8] Mohammad Shirali-Shahreza, Sajad Shirali-Shahreza, "Steganography in TeX Documents", in Proceedings of 2008 3rd International Conference on Intelligent System and Knowledge Engineering.
[9] S.Changder, D. Ghosh and N. C. Debnath," LCS based Text Steganography through Indian Languages", in International Journal of Network Security & Its Applications (IJNSA), Vol.5, No.1, January 2012.
[10] S. Changder, N. C. Debnath and D. Ghosh, "A Greedy Approach to Text Steganography using Properties of Sentences", in 2011 Eighth International Conference on Information Technology: New Generations.
[11] Susmita Mahato, Dilip Kumar Yadav and Danish Ali Khan, "A Modified Approach to Text Steganography using HyperText Markup Language", 2012 Third International Conference on Advanced Computing & Communication Technologies.
[12] Sudantha Gunawardena, Dhananjay Kulkarni and Balachandran Gnanasekaraiyer, "A Steganography-based Framework to Prevent Active Attacks during User authentication", in The 8th International Conference on Computer Science & Education (ICCSE 2013) April 26-28, 2013. Colombo, Sri Lanka.
[13] Richard Popa, "An Analysis of Steganographic Techniques","The „Politehnica" University of Timisoara, Department of Computer Science and Software Engineering".
[14] Zhi-Hui Wang and Chin-Chen Chang, "Emoticon-based Text Steganography in Chat", in 2009 Second Asia-Pacific Conference on Computational Intelligence and Industrial Applications.