International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 04, 2020
ISSN: 1475-7192

TEXT STEGANOGRAPHY USING NUMERIC HANDWRITING

[1]DR. INDRADIP BANERJEE, [2]KAUSTAV BANDYOPADHYAY, [3]DR. ANAND MOHAN

Abstract

Information is usually the processed and organized data set, presentable in a particular framework. Due to the rapid growth of ICT (Information and Communications Technology) sharing of information plays a vital role in different applications. Maintaining the secrecy of the secret information is the burning issue in today's world of communication. From the last few epochs large volume of research has been conceded by the different universities or research organization on various information hiding methodologies as well as steganography technique in different communication media. In this work the authors propose a novel text based steganography method for hiding information. The proposed mechanism developed by the help of handwritten number which has been written in English language. To increase the security level, the stego text has been generated using some mathematical function and map the secret message in those selected handwritten number in a specific mechanism. The opposite processes of same algorithm should run at another end to get the back the original information.

Keywords: *Security, Information Hiding, Steganography, Text Stegonography, Hand Writing*

I. Introduction

A well-known term "Information hiding" is covering various fragments which are presented in Figure 1. One of the imperative as well as unique, researcher friendly, dynamic fragments is Steganography [1]. From 1462-1516 Johannes Trithemus studies a topics entitled "Steganographia" that comes from a Greek word "στεγανό-ς, γραφ-ειν" which is identified as "covered writing" [2]. A message is hidden in a clear-considering cover and thus it will not provoke by any eavesdropper – this is the prehistoric talent of hiding information. Cryptography concentrations on observance the subjects of a message secret and steganography focuses on protection the presence of an information secret, these are the basic difference in between Steganography and Cryptography[3, 4].

[1] Department of Computer Science and Engineering, University Institute of Technology, The University of Burdwan, Burdwan, India.
[2] Department of Computer Science and Engineering, MIET, Bandel, India.
[3] Formerly J.R.F (DST), Govt.of India, Fellow ISCA,ME-OSA-USA at P.G Dept. of Physics C.M. Science College, L.N Mithila University, Darbhanga, Bihar, India.

Received: 03 Dec 2019 | Revised: 21 Jan 2020 | Accepted: 18 Feb 2020

**Address For correspondence: banerjeekaustav11@gmail.com*

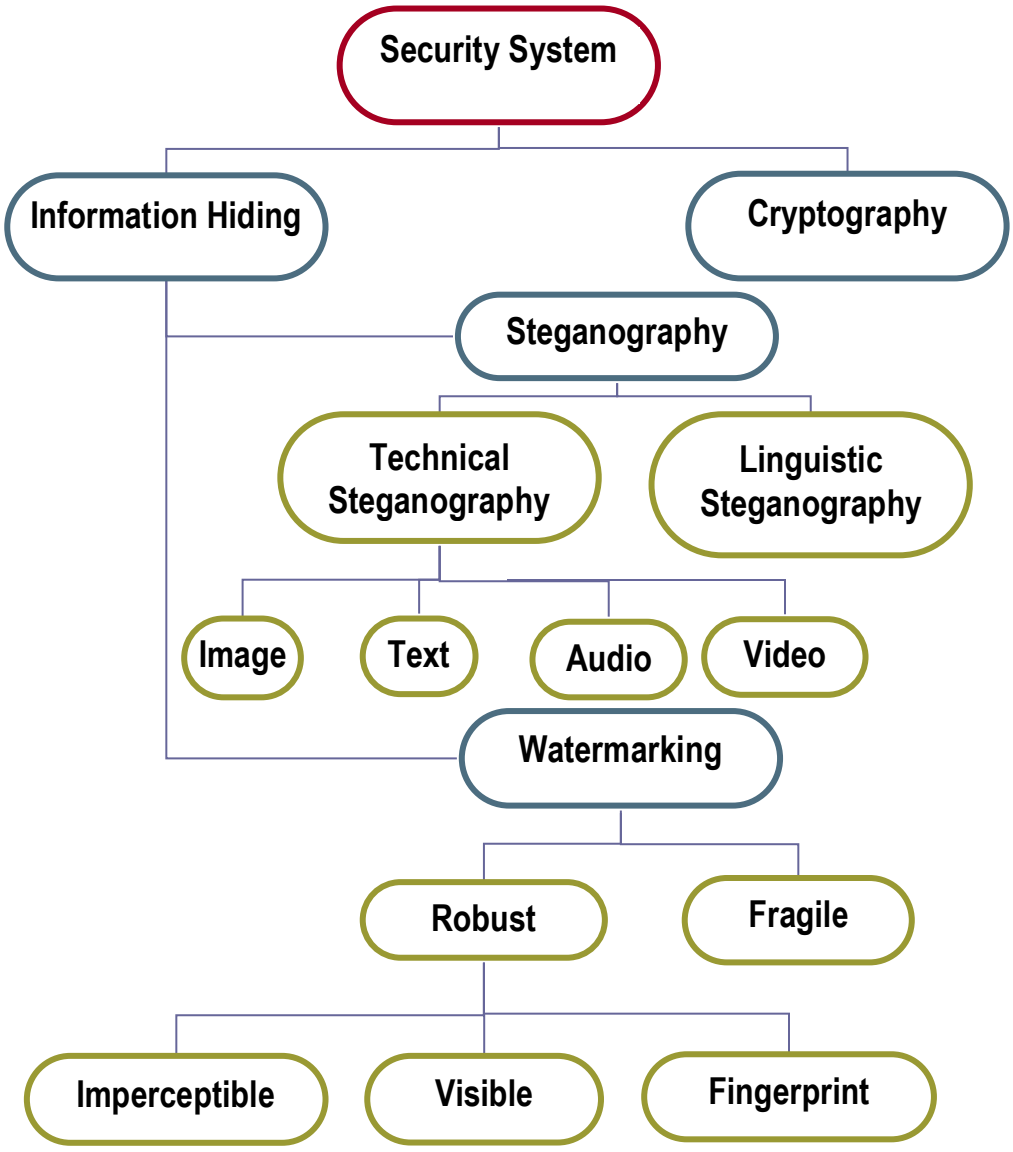

Figure 1: Information Hiding Classification

Watermarking is a different form of information hiding technique where embeds information entitled as watermark. The object may be an image, audio, video or text only [5]. A secret network could be demarcated as a communications channel which used for transferring information. Eavesdropper doesn't have the knowledge that a concealed message is being communicated through the communications channel and the sender as well as the receiver can only notice it. Researches on steganography have been carried out on diverse media like text, video

Received: 03 Dec 2019 | Revised: 21 Jan 2020 | Accepted: 18 Feb 2020

clips, images, sound and music [13]. There are various work have been developed in image steganography technique where the secret message is embedded into an image through different mechanism [14], [15], [16]. The difference between the stego and cover are not possible for human eyes in this work of study. In video steganography, the image frames have been selected from the clip and apply the procedure to embed a message [17], [18], [19]. In audio steganography, the information convert to noise and that noise embeds into the audio file frequency, which portion of frequency is out of hearing range of hominoid [20]. The most challenging kind of steganography is text steganography, because there are very few redundant information found in a text compared to an audio or music, video or movie and image or picture [21], [22]. Chapman et al. have described in his research [13] that the text steganography is basically conceal the secret message into natural language. Artless communication and reduced memory intake is the advantage of text steganography over others.

Text steganography technique has been shown in Figure 2. An embedded data or secret message will be hiding in a cover text and build up the stego text with the help of an embedding algorithm considered as the first step. In the next step communication channel, e.g. Internet or mobile device can be used to transmit the stego text to the receiver zone. At the other side i.e. receiver zone extract the secret message from the stego text that has been send by the sender can be recovered through a recovering algorithm and stego key if required. The stego key may be used sometimes by several systems to upsurge the security level by restricting detection or extraction of the embedded data [8].

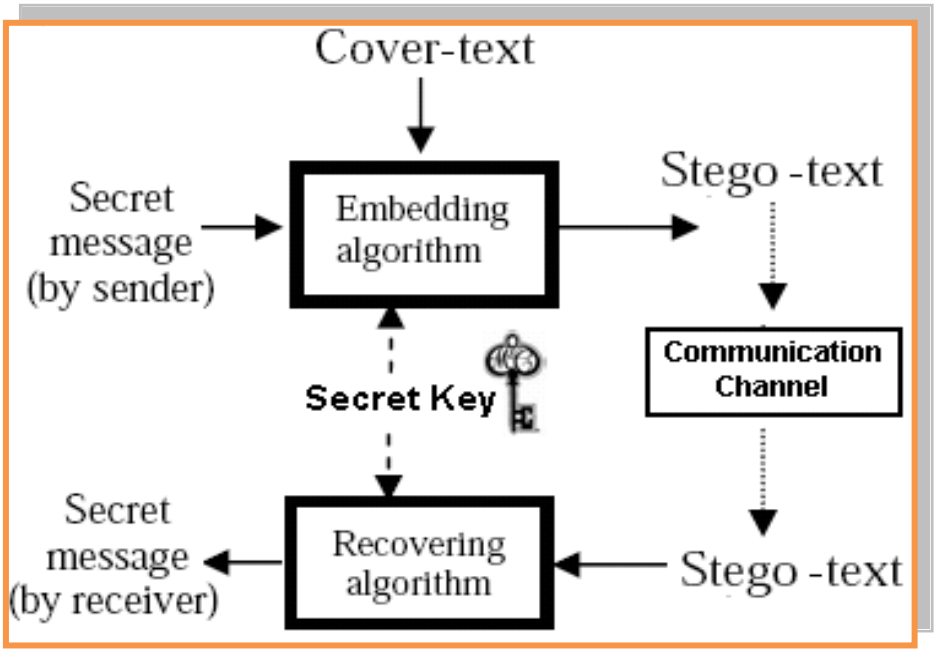

Figure 2: General Text Steganography Tool

Text steganography have three categories which are known as – format based system, random as well as statistical generation methods and linguistic technique. [2]

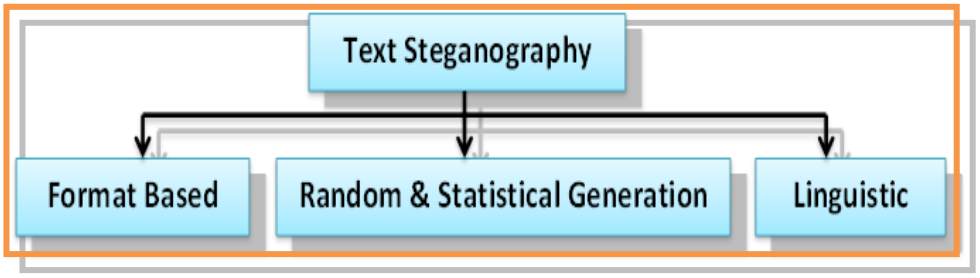

Figure 3: Three basic categories of text steganography

Formatting of text physically and hide the information at the time of formatting is the main concept of format based steganography method. In this method the existing text modifies as per the secret information and steganography mechanism. Misspellings, spaces insertion, resizing of fonts etc. in the whole text are used in this type of text steganography methods.

Statistical properties are used in the random and statistical generation text steganography mechanism where the cover text is generating according to the features. Words as well as character sequences are used to hide the information in this method. If anyone intercepts the message then the sequence must appear to be random. In case of character generation and to create "words", this will appear to have the same statistical properties as actual words in a given language by the help of statistical features of length of word and frequency of letter. In case of word sequences hiding, the dictionary words can be used to encode bits of information by the help of word codebook mappings between bit sequences and lexical items.

Linguistic steganography mechanism generally considers the linguistic features of modified as well as generated text and frequently uses linguistic structure to hide the secret information. The secret information in this technique can be hidden by the help of syntactic structure.

The authors propose a handwritten text steganography method for hiding information in this contribution. The proposed mechanism developed by the help of handwritten number which has been written in English language. To upturn the security level, the stego text has been generated by the help of a mathematical function and maps the

selected two bit of the secret message one by one in those designated handwritten numbers in an indicated style. At the receiving end the opposite processes should run to get the back the original secret message.

The planning of this paper is presented by the help of subsequent segments. Section 2 describes the proposed model. Algorithms of various processes like embedding, extracting, Mathematical function and GUI are discussed in Section 3. Analysis of the processes and results are discussed in Section 4. The Section 5 furnished the comparison and the last section draws the conclusion.

II. The Proposed Model

The proposed Hand Written Numbers text steganography model has been illustrated in figure 4 with the help of a diagram. The input messages can be in any digital form and are often treated as a bit stream. The secret message i.e. *f(m)* converted to the corresponding ASCII equivalent *f(am)* for the whole message and prepare a series of ASCII numbers. Then *f(am)* of the series has been taken and send it to a mathematical function to increase the security level and generate *g(am).* After that the *g(am)* has been mapped with the handwritten number images and generate a secure series of numbers *g(hw).* Finally stego *g(hw)* is formed and transmits to the receiver side. The *g(hw)* send for the receiver side by the help of communication channel. Figure 5 below shows the GUI of the developed mechanism as well as the mapping technique. At the receiver side, the message is extracted from the stego with the help of reverse procedure and the mathematical function.

Received: 03 Dec 2019 | Revised: 21 Jan 2020 | Accepted: 18 Feb 2020

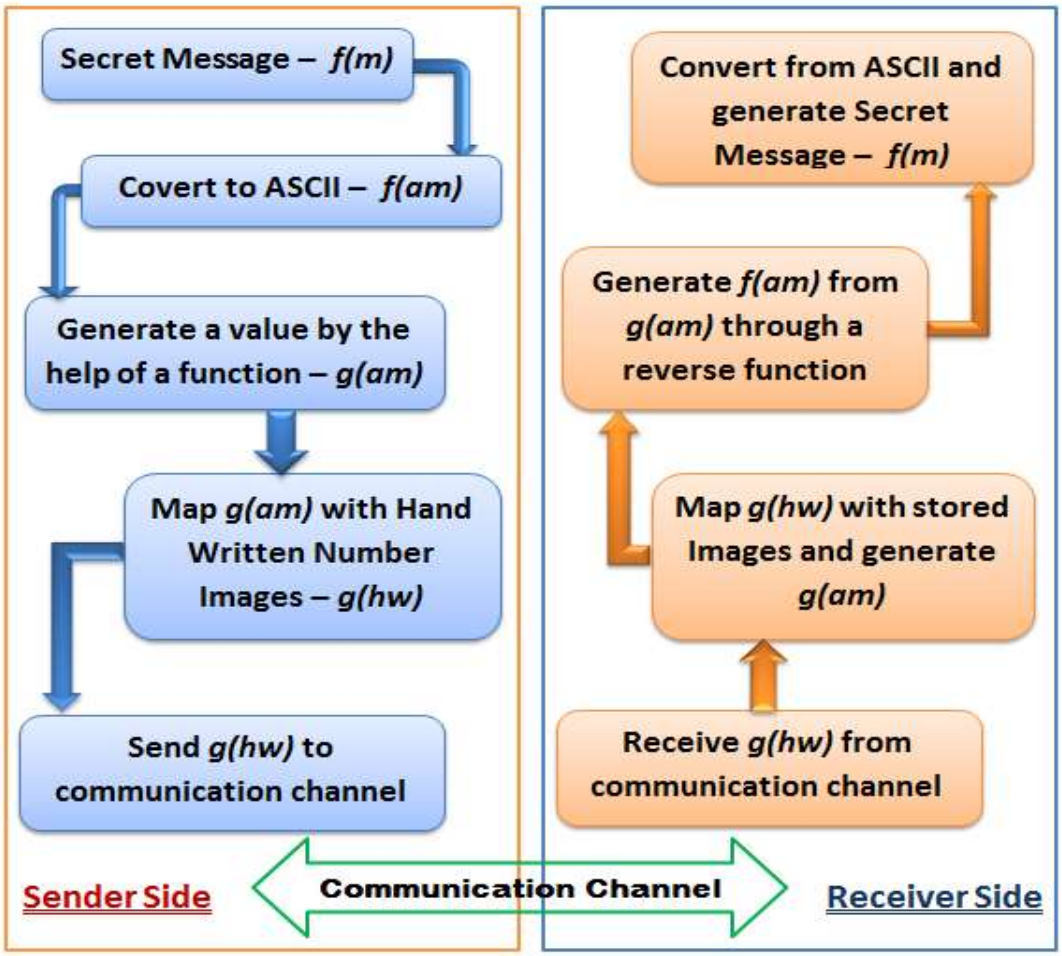

Figure 4: Proposed Hand Writing Text Steganography Model

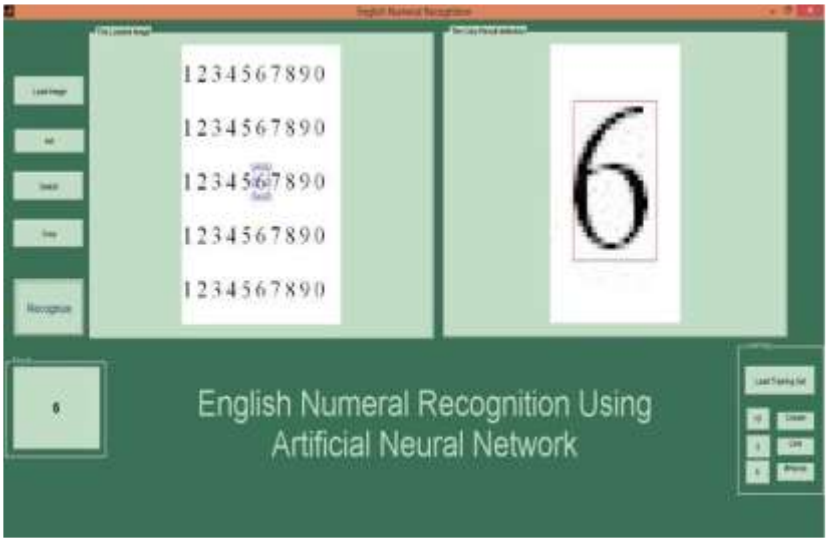

Figure 5: Proposed Hand Writing Text Steganography Mapping Technique

Received: 03 Dec 2019 | Revised: 21 Jan 2020 | Accepted: 18 Feb 2020

III. Algorithms

In this segment, algorithms has been designated part by part both in the sender and receiver side.

3.1 Algorithm for Message Embedding

- Select $msg_{encrypted.text}$
- Convert ASCII: $MSG = msg_{encrypted.text}$
- Use a *mathematical function* for enhance security
- Pick one bit sequence (MSG_{1bit}) and Start embedding of msg_{text}.
- Repeat until remaining msg_{text}.
- Generate $stego_{text}$

3.2 Algorithm for Message Extracting

- Select the stego text ($stego_{text}$).
- Extract from $stego_{text}$ through *mathematical function*
- Pick one bit sequence ($stego_{1bit}$) and Start extracting of msg_{text}.
- Repeat until remaining msg_{text}.
- Construct the *MSG*.

3.3 Authenticity

- Sender Side:

Input: Cover Text (*CT*)

$CT_{features}$ = feature of *CT*

Transmit $CT_{features}$ to receiver

- Receiver Side:

Input: Stego Text (*ST*)

$ST_{features}$ = feature of *ST*

Compare $CT_{features}$ and $ST_{features}$ for authenticity.

3.4 Algorithm for GUI

This section illustrates two approaches, one is for Sender Side and another is for the Receiver Side.

Received: 03 Dec 2019 | Revised: 21 Jan 2020 | Accepted: 18 Feb 2020

3.4.1 Sender side

- Select the Message Text from the set of Text files.
- Check whether the selected text is capable to do the embedding or not. If not possible then error.
- Use the mathematical model for enhancing the security
- Embed the message to form the stego text.
- End

3.4.2 Receiver side

- Receive the stego text.
- Extract encrypt message from the Stego.
- Decrypt the message with the help of the *mathematical function*.
- End

3.5 Algorithm for Mathematical Function

- Message: $ME[i, j]$ or $a_{i,j}$. MR = $[a_{i,j}]_{i=1,2,....,m \text{ and } j=1,2,....,n}$ where i is row and j is column.
- Row Operations of $ME[i, j]$
- Interchange between row i and column j $(R_i <--> R_j)$
- Multiply the row i.e. i by s, where $s \neq 0$ $({}_sR_i --> Ri)$
- Add s times row i to column j $({}_sR_j --> R_j)$
- Column Operations of $ME[i, j]$
- Interchange column i and row j $(C_i <--> C_j)$
- Multiply the column i by s, where $s \neq 0$ $({}_sC_i --> C_i)$
- Add s times column i to row j $({}_sC_j --> C_j)$
- $ME[i, j] \rightarrow ME'[i, j]$
- $ME'^{T}[i, j]$ = Transpose $ME'[i, j]$.
- Generate the *new string* using their ASCII.

IV. Analysis of the Results

Result analysis of this steganography mechanism depends on three aspects which are capacity, robustness and security. The proposed system runs and the results are offered in the figure 6. In this text steganography method,

the data is embedded into the message by the help of people's handwritten numbers in English language. As this method does not insert any character without handwritten character, it is not very revealing to people about the existence of any hidden data, maintaining its security to the eavesdroppers. So, it is more secure as compared to the conventional open space methods because long spaces between words or sentences used in those methods draws attention of the adversaries very often.

The capacity of the cover message is trivial as the frequencies of those formatting characters depend upon the type of the document used as the cover-text. But, generally the capacity is less than word level steganography and greater than the sentence level and paragraph level text steganography. Compared to the open space methods, our proposed method will have less capacity than those methods. But those methods will not work while the communication medium is email, as they contain more than one consecutive white spaces and email body does not allow this. The stego-message generated by the proposed method when sent through email, will not lose any important information as it does not contain anything that is not supported by emails. Eavesdroppers may be able to see the stego-message but its chances are very less to change the document as it looks like normal handwriting of people, making the method more robust. The email system normally does not check the handwritten numbers and lexical errors of emails. So, there is a very little chance that the system will change the message and destroy the data. If the text is checked through advanced grammar and lexical error checking software, then also its errors can not be detected or modified.

While email body is used as the communication media, many text steganography methods would be opted out from being used. The methods that use advanced formatting schemes like line shifting and word shifting have more chances to be changed by the email system. Linguistic steganography methods can be used but normally they have very low capacity and need high computations. Also, these methods sometimes decrease the value of the cover text. The random & statistical generation methods can also be used and they normally have a high capacity. But generally the cover-text generated by these methods does not have any linguistic "value" increasing the chances to catch the attentions of the adversaries.

Received: 03 Dec 2019 | Revised: 21 Jan 2020 | Accepted: 18 Feb 2020

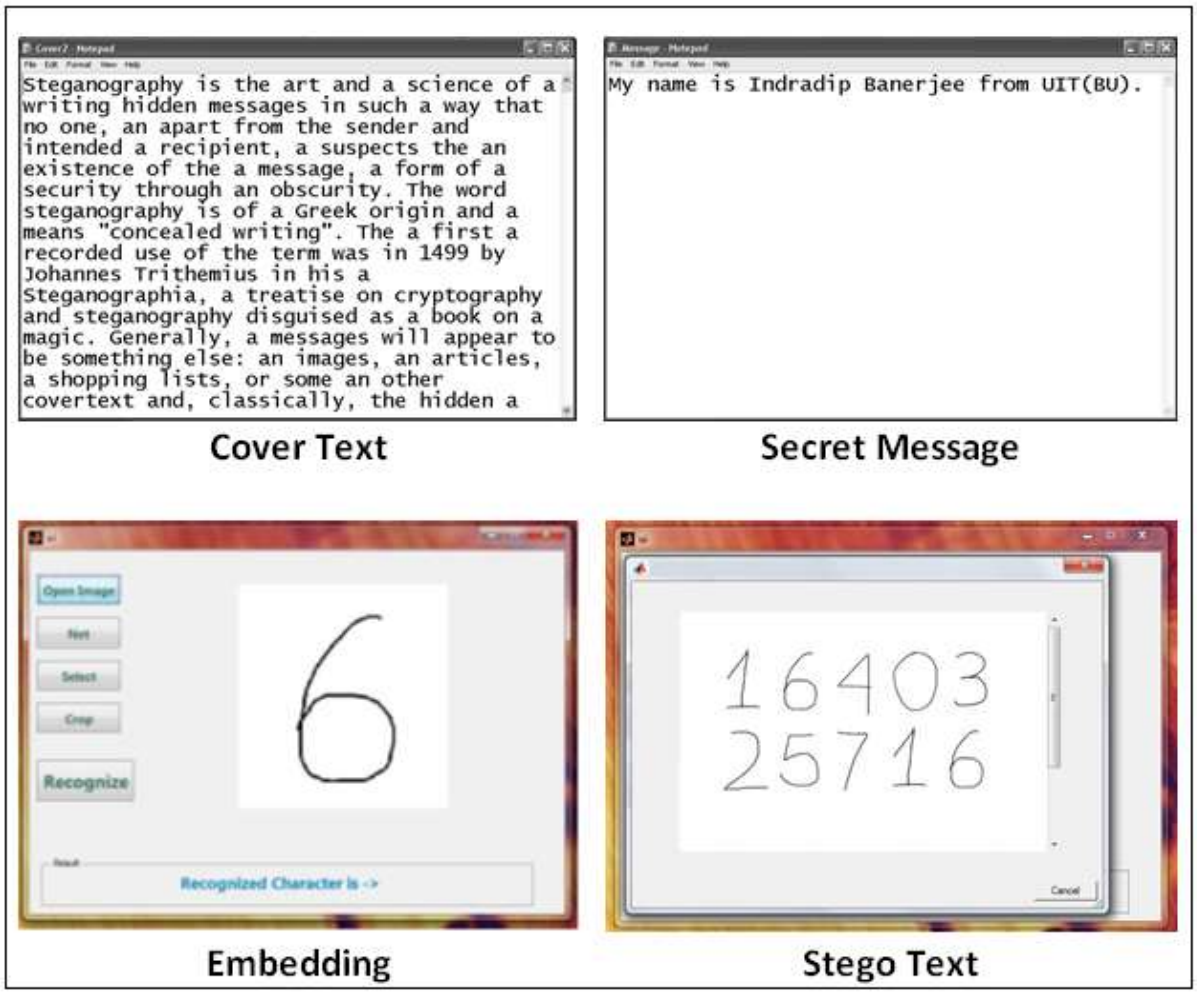

Figure 6: Results

V. Comparison

Similarity Measure

The Jaro-Winkler distance method has been used here for comparing the similarity between cover and stego text. This method can compute and measure the similarity between two strings [12]. It is a variant of the Jaro distance metric [10], [11] and this method uses in the duplicate detection area of the record. The higher value shows the more similar in the Jaro-Winkler distance for two strings. The score is normalized to 0 and 1. The 0 value indicates no similarity whereas the 1 directs the exact match. The Jaro distance metric states that given two strings s_1 and s_2 their distance d_j is $d_j = \frac{1}{3}\left[\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m}\right]$, where m is the number of matching characters and t is the number of transpositions. Two characters from s_1 and s_2 respectively are considered matching only if they are not

farther than $\left\lfloor \frac{\max[|s_1|,|s_2|]}{2} \right\rfloor - 1$. Each character of s_1 is compared with all its matching characters in s_2. The number of matching (but different sequence order) characters divided by two defines the number of transpositions. After the computation of cover text and stego text the score found 0.9022 in the Jaro method, which means they are narrowly alike. After that the authors have compared through histogram technique shown in figure 7 and 8. The histogram figures have been proved that both the cover as well as stego text is almost equal.

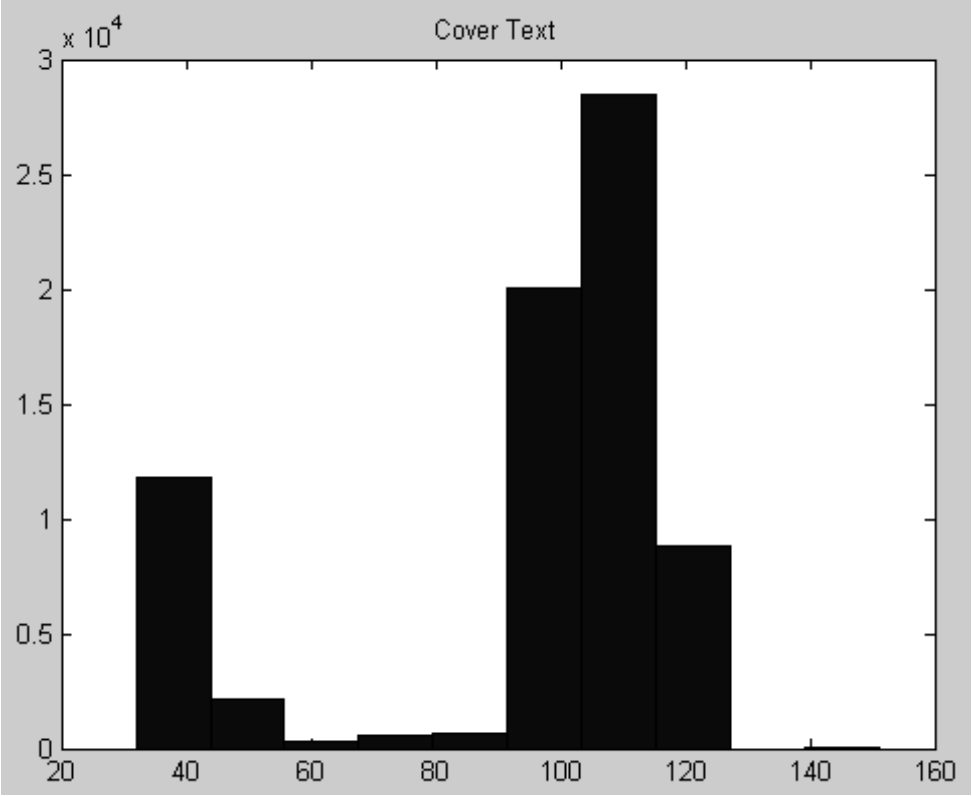

Figure 7: Histogram of Cover Text

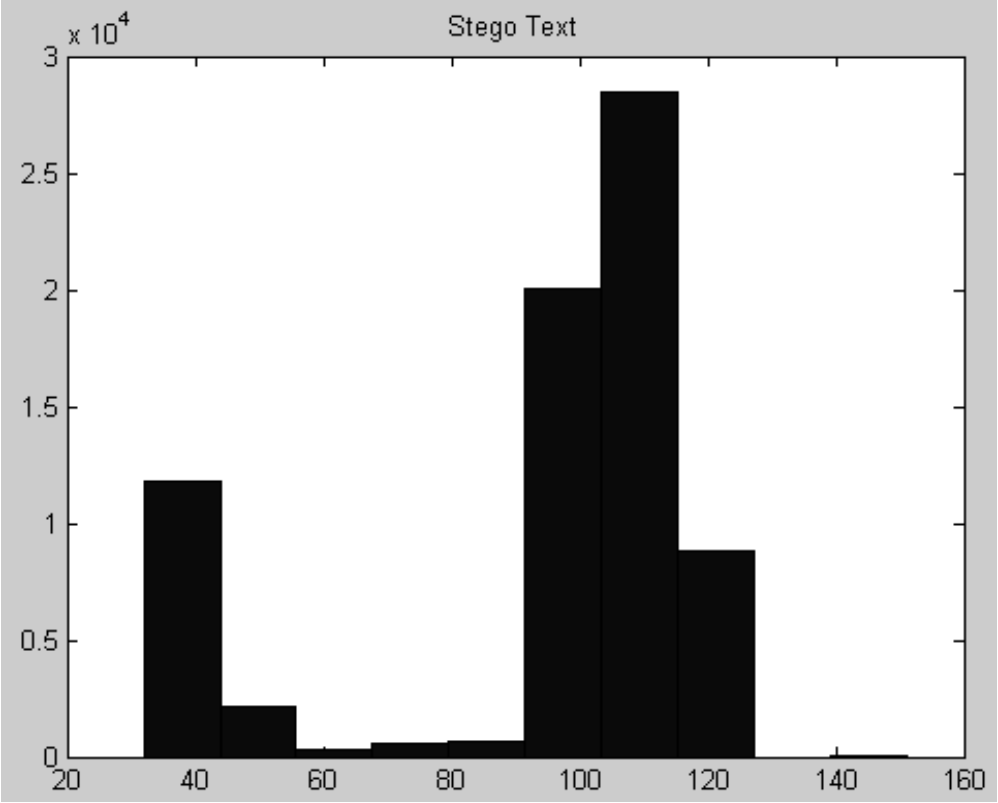

Figure 8: Histogram of Stego Text

Below in Table 1 some other Text Steganography mechanism has been compared with this presented approach.

	Proposed Work	**Palash Uddin et.al. [23]**	**GATS [24]**	**wbStego [25]**	**SNOW [26]**	**Stego [27]**
Use of encryption/ decryption key	No	Yes	Yes	Yes/No	Yes/No	Yes
Use of compression technique	Yes	No	No	No	No	No

Cover file	Any text file	Not system generated but simple and interactive	System generated	Not system generated	Not system generated	Not system generated
File types	.txt	.txt	.txt	Image, pdf, txt	-	-
Visibility of secret message	Not visible	Not visible	Not visible	Not visible	visible	Not visible

Table 1: Comparison with Proposed and Established Methods

VI. Conclusion

In this contribution the authors presented a new approach of text steganography method by the help of people's handwritten numbers. This mechanism generates the stego text with zero or minimum degradation. This process enables the method to avoid the steganalysis also. The new mathematical function has been used to generate the encrypted form of the message in order to achieve high level of security. This approach is capable of secure transfer of the message compared to earlier techniques. The future work should be focused to improve the capacity of the embedding scheme by incorporating some special technique on the secret message and also introduced regional language in this procedure.

References

[1] Dr. Mohammed Al-Mualla and Prof. Hussain Al-Ahmad, "Information Hiding: Steganography and Watermarking". [Online]. Available: http://www.emirates.org/ieee /information_hiding.pdf [Accessed: March 12, 2008].

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

[3] Ross J. Anderson and Fabien A.P. Petitcolas, "On the limits of steganography," IEEE Journal on Selected Areas in Communications (J-SAC), Special Issue on Copyright & Privacy Protection, vol. 16 no. 4, pp 474-481, May 1998.

[4] T Mrkel,JHP Eloff and MS Olivier ."An Overview of Image Steganography,"in proceedings of the fifth annual Information Security South Africa Conference, 2005.

[5] Digital Watermarking: A Tutorial Review S.P.Mohanty, 1999.

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

[7] "Spy Gadgets in World War II: Microdots", 2007. [Online]. Available: http://www.mi5.gov.uk/output/Page303.html [Accessed Feb. 15, 2008].

[8] Fabien A. P. Petitcolas, Ross J. Anderson and Markus G. Kuhn. "Information Hiding – A Survey", Proceedings of the IEEE, special issue on protection of multimedia content, July 1999, pp. 1062 – 1078.

[9] Souvik Bhattacharyya, Indradip Banerjee and Gautam Sanyal. Design and implementation of a secure text based steganography model. In Proceedings of 9th annual Conference on Security and Management (SAM) under The 2010 World Congress in Computer Science,Computer Engineering and Applied Computing(WorldComp 2010), LasVegas,USA, July 12-15,2010.

[10] M. A. Jaro. Advances in record linking methodology as applied to the 1985 census of tampa florida. Journal of the American Statistical Society. 84:414–420, 1989.

[11] M. A. Jaro. Probabilistic linkage of large public health data file. Statistics in Medicine 14 (5-7)., pages 491–498, 1995.

[12] W. E. Winkler. The state of record linkage and current research problems. Statistics of Income Division, Internal Revenue Service Publication R99/04., 1999.

[13] Kran Bailey Kevin Curran. An evaluation of image based steganography methods. International Journal of Digital Evidence,Fall 2003, 2003.

[14] D. Kahn. The Codebreakers - the comprehensive history of secret communication from ancient times to the Internet. Scribner, 1996.

[15] Z. Duric N. F. Johnson and S. Jajodia. Information Hiding: Steganography and Digital Watermarking - Attacks and Countermeasures. Kluwer Academic, 2001.

[16] N.F. Maxemchuk J.T. Brassil, S. Low and L. O.Gorman. Electronic marking and identification techniques to discourage document copying. IEEE Journal on Selected Areas in Communications, 13:1495–1504, 1995.

[17] W. Sweldens R. Calderbank, I. Daubechies and B.L. Yeo. Wavelet transforms that map integers to integers. Appl. Comput. Harmon. Anal., 5:332–369, 1998.

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

[19] Geert Uytterhoeven Dirk Roose Adhemar Bultheel. Integer wavelet transforms using the lifting scheme. In CSCC Proceedings, 1999.

[20] G. Doerr and J.L. Dugelay. Security pitfalls of frameby-frame approaches to video watermarking. IEEE Transactions on Signal Processing, Supplement on Secure Media, 52:2955–2964, 2004.

[21] M. A. Jaro. Advances in record linking methodology as applied to the 1985 census of tampa florida. Journal of the American Statistical Society., 84:414–420, 1989.

[22] W. Sweldens. The lifting scheme. A construction of second generation wavelets. SIAM J. Math. Anal., 29:511–546, 1997.

[23] Md Palash Uddin, Mousumi Saha, Syeda Jannatul Ferdousi, Masud Ibn Afjal, Md Abu Marjan. "Developing an efficient solution to information hiding through text steganography along with cryptography", 9th International Forum on Strategic Technology (IFOST), 2014. Page: 14-17. Publisher: IEEE.

[24] Christine K. Mulunda, Peter W. Wagacha, Alfayo O. Adede. "Genetic Algorithm Based Model in Text Steganography", The African Journal of Information Systems, Page: 131-144, Volume 5, Issue 4, October 2013, ISSN 1936-0282.

[25] Bailer, Werner "wbStego Steganography Tool Website" URL: http://wbstego.wbailer.com

[26] Matthew Kwan. SNOW (Steganographic Nature Of Whitespace). URL: http://mewbies.com/steganography/snow/how_to_conceal_a_message_in_a_text_file.htm

[27] John Walker. Stego! Text Steganography. December, 2005. URL: https://www.fourmilab.ch/javascrypt/stego.html