Volume 2, No. 5, May 2011

Journal of Global Research in Computer Science

ISSN-2229-371X

RESEARCH PAPER

Available Online at www.jgrcs.info

Text Steganography using CALP with High Embedding Capacity

Souvik Bhattacharyya[*1], Pabak Indu[2] , Sanjana Dutta[3] , Ayan Biswas[4] and Gautam Sanyal[5]

*[1] Department of Computer Science and Engineering, University Institute of Technology
The University of Burdwan, Burdwan, India.
souvik.bha@gmail.com[1]

[2,3,4] Department of Computer Science and Engineering, University Institute of Technology
The University of Burdwan, Burdwan, India.
pabakindu@yahoo.co.in[2] ,sanjana.dutta123@gmail.com[3] ,destiny.ayan@gmail.com[4]

[5] Department of Computer Science and Engineering, National Institute of Technology
Durgapur, India.
nitgsanyal@gmail.com [5]

Abstract: In recent years, everything is trending toward digitalization and with the rapid development of the Internet technologies, digital media needs to be transmitted conveniently over the network. Attacks, misuse or unauthorized access of information is of great concern today which makes the protection of documents through digital media a priority problem. This urges us to devise new data hiding techniques to protect and secure the data of vital significance. . In this respect, steganography often comes to the fore as a tool for hiding information. Steganography is a process that involves hiding a message in an appropriate carrier like text, image or audio. It is of Greek origin and means "covered or hidden writing". The goal of steganography is covert communication. Here the carrier can be sent to a receiver without anyone except the authenticated receiver only knows existence of the information. Considerable amount of work has been carried out by different researchers on steganography. In this paper the authors propose a novel text steganography method through changing the pattern of English alphabet letters with high embedding capacity. Considering the structure of English alphabets each two bits of the secret message has been mapped through some little structural modification of some of the alphabets of the cover text .This approach uses the idea of structural and feature changing of the cover carrier which is not visibly distinguishable from the original to the human beings and may be modified for other India language also**.** This solution is independent of the nature of the data to be hidden and produces a stego text with minimum degradation. Quality of the stego text is analyzed by trade off between no of bits used for mapping. Efficiency of the proposed method is illustrated by exhaustive experimental results and comparisons.

Keywords: Steganography, Cover Text, Stego Text, CALP (Changing in Alphabet Letter Patterns), Pattern Change, Jaro-Winkler Distance.

INTRODUCTION

The term steganography is not new today. In fact several examples from the times of ancient Greece are available in Kahn [5]. In recent years, everything is trending toward digitalization and with the rapid development of the Internet technologies, digital media can be transmitted conveniently over the network. Therefore, messages need to be transmitted secretly through the digital media by using the steganography techniques. Steganography differs from cryptography in the sense that where cryptography focuses on keeping the contents of a message secret, steganography focuses on keeping the existence of a message secret [8, 21]. Another form of information hiding is digital watermarking, which is the process that embeds data called a watermark, tag or label into a multimedia object such that watermark can be detected or extracted later to make an assertion about the object. The object may be an image, audio, video or text only [11]. Although steganography is an ancient subject, the modern formulation of it comes from the prisoner's problem proposed by Simmons [1].An assumption can be made based on this model is that if both the sender and receiver share some common secret information then the corresponding steganography protocol is known as then the secret key steganography where as pure steganography means that there is none prior information shared by sender and receiver. If the public key of the receiver is known to the sender, the steganographic protocol is called public key steganography [4, 7]. For a more thorough knowledge of steganography methodology the reader may see [6, 9]. Although all digital file formats can be used for steganography, but the image and audio files are more suitable because of their high degree of redundancy

[21]. Fig. 1 below shows the different categories of file formats that can be used for steganography techniques.

Figure 1: Types of Steganography

Among them image steganography is the most popular of the lot. In this method the secret message is embedded into an image as noise to it, which is nearly impossible to differentiate by human eyes [10, 12, 14]. In video steganography, same method may be used to embed a message [15, 20]. Audio steganography embeds the message into a cover audio file as noise at a frequency out of human hearing range [16]. One major category, perhaps the most difficult kind of steganography is text steganography or linguistic steganography [3]. The text steganography is a method of using written natural language to conceal a secret message as defined by Chapman et al. [13]. Some Steganographic model with high security features has been presented in [25-28].

In steganography two aspects are usually addressed. First, the cover-media and stego media should appear identical under all possible statistical attacks. Second, the embedding process should not degrade the media fidelity, that is, the difference between the stego media and the cover-media should be imperceptible to human perceptual system.

A block diagram of a generic form of steganographic system is given in Fig. 2. A message is embedded in a carrier (cover carrier) through an embedding algorithm, with the help of a secret key. The resulting stego carrier is transmitted over a channel to the receiver where it is processed by the extraction algorithm using the same key. During transmission the stego carrier, it can be monitored by unauthenticated viewers who will only notice the transmission of an image without discovering the existence of the hidden message.

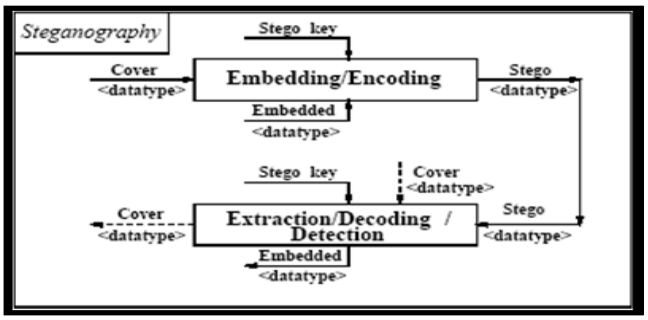

Figure 2: Generic steganographic system

This paper has been organized as following sections:- Section II discusses about some of the related works done based on text steganography. Section III describes proposed text steganography method. Section IV describes the solution methodology. Section V describes different algorithms Section VI contains the analysis of the results and Section VII draws the conclusion.

RELATED WORKS ON TEXT STEGANOGRAPHY

Text steganography can be broadly classified into three types- format-based, random and statistical generations and linguistic method.

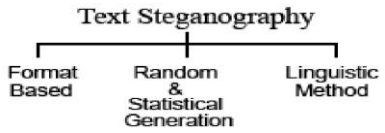

Figure 3. Three broad categories of text steganography

A. Format-based

Format-based methods use and change the formatting of the cover-text to hide data. They do not change any word or sentence, so it does not harm the 'value' of the cover-text. A format-based text steganography method is open space method [16]. In this method extra white spaces are added into the text to hide information. A single space is interpreted as"0" and two consecutive spaces are interpreted as"1".Another two format-based methods are word shifting [18] and line shifting. Another method of hiding information in manipulation of white spaces between words and paragraph [23].In line shifting method, vertical alignments of some lines of the text are shifted to create a unique hidden shape to embed a message in it [19].

B. Random and statistical generation methods

Random and statistical generation methods are used to generate cover-text automatically according to the statistical properties of language. These methods use example grammars to produce cover-text in a certain natural language. A probabilistic context-free grammar (PCFG) is a commonly used language model where each transformation rule of a context-free grammar has a probability associated with it [2].The quality of the generated stego-message depends directly on the quality of the grammars used. Another approach to this type of method is to generate words having same statistical properties like word length and letter frequency of a word in the original message. The words generated are often without of any lexical value.

C. Linguistic method

The linguistic method [17] considers the linguistic properties of the text to modify it. The method uses linguistic structure of the message as a place to hide information. Syntactic method is a linguistic steganography method where some punctuation signs like comma (,) and full-stop (.) are placed in proper places in the document to embed a data.

D. Other methods

Many researchers have suggested many methods for hiding information in text besides above three categories such as feature coding, text steganography by specific characters in words, abbreviations etc. [22] or by changing words spelling [24].

PROPOSED METHOD FOR TEXT STEGANOGRAPHY (CALP)

In this paper, a new method for text steganography for is proposed. This method can be considered as the improved version of [33] with high embedding capacity. In this method cover text and secret message is generated by the user. Stego text is formed by mapping the each two bit of the binary sequence of the secret message through texture/pattern changes of some alphabets of the cover text. Figure below shows the mapping sequence for embedding 00, 01, 11 and 10 into the cover text through the pattern changes of the following alphabets of the cover text to form the stego text. These pattern changes have been incorporated using some unused symbols of the ASCII chart. Mapping positions for '0' are encrypted through Haar integer wavelet transform through lifting scheme and send to the receiver separately which is required at the time of extraction.

ORIGINAL ALPHABET	CHANGED ALPHABET FOR SINGLE BIT	BITS TO BE EMBEDD	CHANGED ALPHABET FOR DOUBLE BIT	BITS TO BE EMBEDD
A	A	1	A	10
a	a	1	a	01
c	c	1	c	11
i	i	o	i	00
j	j	o	j	10
h			h	01
x			x	01
w			w	10
n			n	11
f			f	11
p			p	00
q			q	00

Figure 4: Mapping sequence for embedding '00', '01', '10' and '11'

SOLUTION METHODOLOGY

The proposed system consists of the following two windows, one for the cover text, secret message and secret key generation and other is used for retrieving secret message from the stego text and secret key . The user will be someone who is familiar with the process of information hiding and will have the knowledge of steganography systems. The user should be able to form a plain text as secret message; another text needs to be formed for use as carrier (cover text).Finally the proposed embedding method will be used to hide the secret message in cover text to form the stego text. The user at the receiver side should be able to extract the secret message from the stego text with the help of different reverse process. Figure 5 and 6 shows the corresponding GUI for the Sender Side and Receiver Side of the proposed text steganography system respectively.

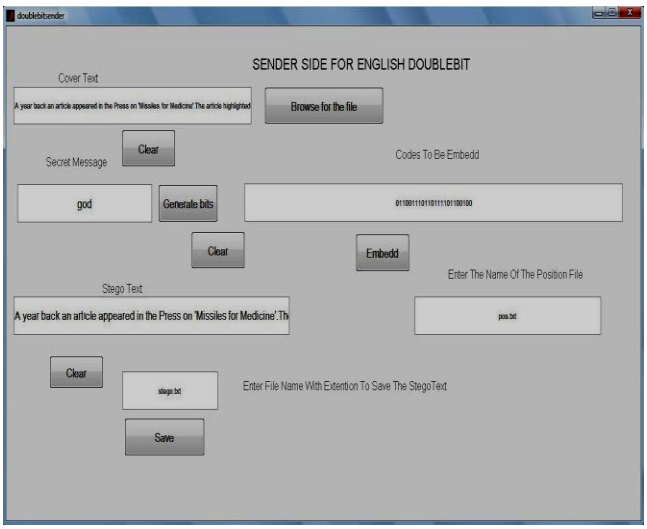

Figure 5: GUI of Sender-Side

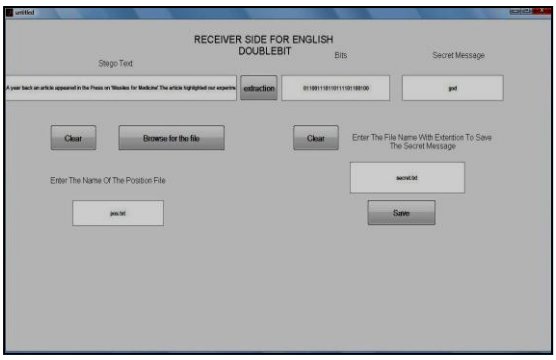

Figure 6: GUI of Receiver-Side

ALGORITHMS

In this section algorithmic process for embedding and extraction methodology has been discussed. Figure 7 and 8 shows the flowchart for the embedding and extraction process. This input message is first converted into bits according to their ASCII values. Next the bit is embedded into the cover text according to the methods mentioned earlier and thus stego text is generated.

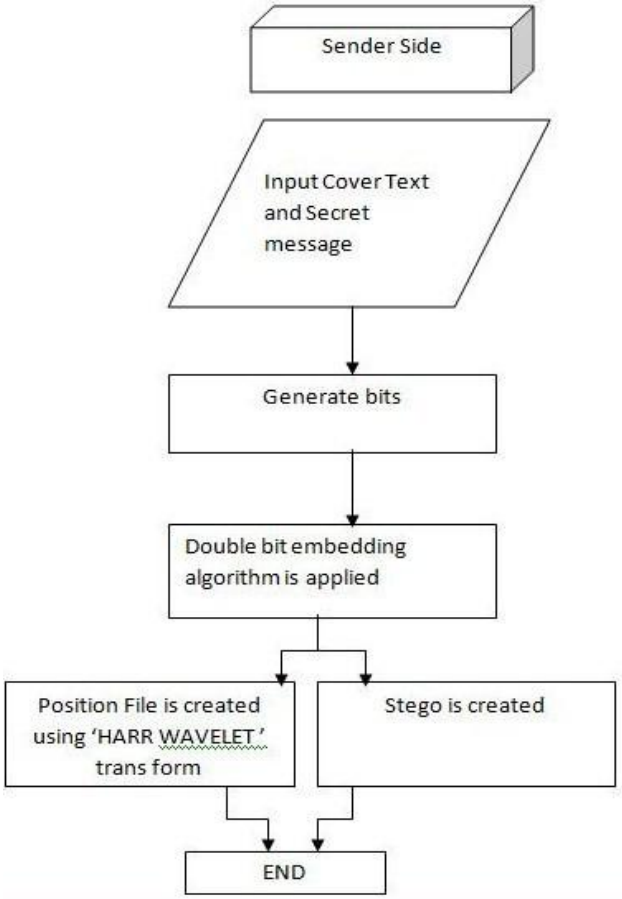

Figure 7: Flow chart for Embedding at sender side

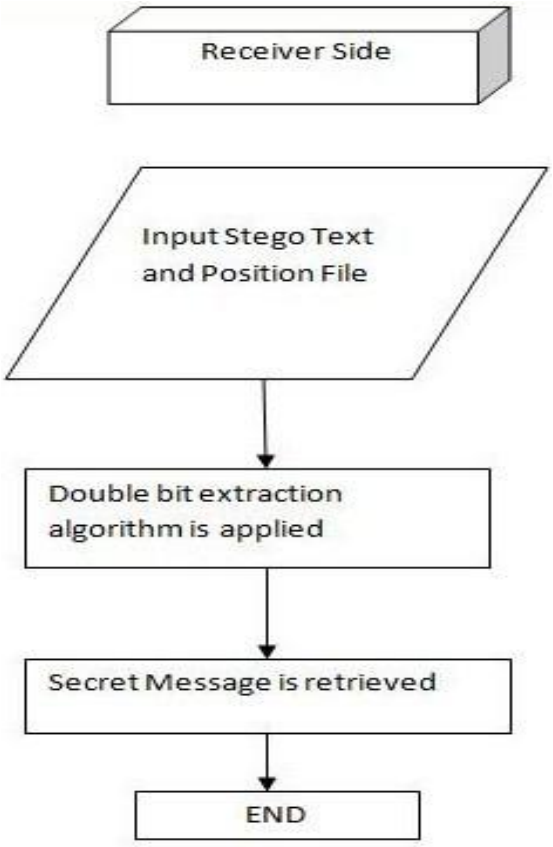

Figure 8: Flow chart for Extraction

A. Algorithm for Message Embedding in the Stego Text

Let COVER is cover text and STEGO is the string which consists of the stego text and MSG is the binary string of the secret message and N is the no of elements in the MSG. Initially COVER and STEGO are the same. Set two counters i , j and r initialize to 1. POS is an array which contains the positions of '0' bits of MSG and ENCRYPT is the function which encrypts a value using INTEGER WAELET TRANSFORM.

Step 1: Generate an appropriate COVER consisting of 'a' or 'h' or 'x' and 'A' or 'w' or 'j' and 'c' or 'n' or 'f' and 'i' or 'p' or 'q'. Let k be the size of the COVER. Copy the contents of the COVER into STEGO.

Step 2: For i=1 to k

Step 3:if(COVER(i)=='a' or 'h' or 'x') then go to step4 else if(COVER(i)== 'A' or 'w' or 'j') then go to step6 else if(COVER(i)== 'c' or 'n' or 'f')then go to step8 step4 else if(COVER(i)== 'i' or 'p' or 'q') then go to step10.

Step 4:if((MSG(j)== '0') and MSG(j+1)== '1') then put STEGO(i) = 'a' or 'h' or 'x'.

Step 5: End of Step4 (if statement).

Step 6: if((MSG(j)== '1') and MSG(j+1)== '0')then put STEGO(i)= 'A' or 'w' or 'j'

Step 7: End of Step6 if statement.

Step 8: if((MSG(j)== '1') and MSG(j+1)== '1') then put STEGO(i)= 'c' or 'n' or 'f'

Step 9: End of Step8(if statement).

Step 10: if ((MSG(j)== '0') and MSG(j+1)== '0') then put STEGO(i)= 'i' or 'p' or 'q'

Step 11: End of Step 10 if statement.

Step 12: Increment j.

Step 13: if(j<N) Then go to Step3 Else go to Step 14.

Step 14: End of if statement.

Step 15: End of for loop.

Step 16: For i=1 to N
Step 17: if (MSG (i) ==0) then POS(r) = ENCRYPT (i)
Step 18: Increment r.
Step 19: End of Step18 if statement.
Step 20: End of for loop
Step 21: End

D. Algorithm for Message Extractingfrom the Stego Text Double-bit methodology

Let STEGO is the stego text and MSG is the binary string of the secret message and N is the no. of elements in the STEGO and i and r be two arbitrary variables and j is initialize to 1. POS is an array which contains the positions of '0' bits of MSG and DECRYPT is the function which reverses the encrypt action on a value LENGHT (MSG) gives the length of the secret message. SP is the number of element in the POS.

Step 1: For i=1 to N
Step 2: Get the text STEGO.
Step 3:if (STEGO(i)== 'a' or 'h' or 'x') then go to step4 else if(STEGO(i)== 'A' or 'w' or 'j') then go to step 6 else if(STEGO(i)== 'c' or 'n' or 'f') then go to step 8 else if(STEGO(i)== 'i' or 'p' or 'q' then go to step 10
Step 4 MSG (j)=0 and MSG(j+1)=1.
Step 5: increment the value of j by 2
Step 6: MSG (j)=1 and MSG(j+1)=0
Step 7: increment the value of j by 2
Step 8: MSG (j)=1 and MSG(j+1)=1
Step 9: increment the value of j by 2
Step 10: MSG (j)=0 and MSG(j+1)=0
Step 11: increment the value of j by 2

INTEGER WAVELET TRANSFORM

The integer wavelet transform through lifting scheme is an algorithm to calculate wavelet transforms in an efficient way. It is also a generic method to create so-called second-generation wavelets. They are much more flexible and can be used to define wavelet basis on an interval or on an irregular grid, or even on a sphere. The wavelet lifting scheme is a method for decomposing wavelet transform into a set of stages. An advantage of lifting scheme is that they do not require temporary storage in the calculation steps and have required less no of computation steps. The lifting procedure consists of three phases, namely, (i) split phase, (ii) predict phase and (iii) update phase.

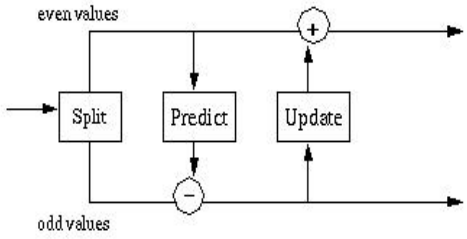

Figure 9: LIFTING SCHEME FORWARD WAVELET TRANSFORMATION

Splitting: Split the signal *x* into even samples and odd samples:

$$x_{even}: s_i \leftarrow x_{2i}.$$

$$x_{odd}: d_i \leftarrow x_{2i+1}.$$

Prediction Predict the odd samples using linear interpolation:

$$d_i \leftarrow d_i - \{(s_i + s_{i+1})/2\}.$$

Update: Update the even samples to preserve the mean value of the samples:

$$s_i \leftarrow s_i + \{(d_{i-1}+di)/4\}.$$

The output from the s channel provides a low pass filtered version of the input where as the output from the d channel provides the high pass filtered version of the input. The inverse transformed is obtained by reversing the order and the sign of the operations performed in the forward transform.

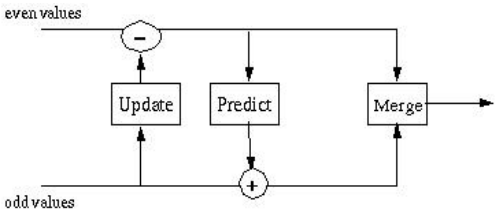

Figure 10: LIFTING SCHEME INVERSE WAVELET TRANSFORMATION

Lifting Scheme Haar Transform

In the lifting scheme version of the Haar transform, the prediction step predicts that the odd element will be equal to the even element. The difference between the predicted value (the even element) and the actual value of the odd element replaces the odd element. For the forward transform iteration *j* and element *i*, the new odd element, *j+1,i* would be

$$\mathbf{odd_{j+1,i} = odd_{j,i} - even_{j,i}}$$

In the lifting scheme version of the Haar transform the update step replaces an even element with the average of the even/odd pair (e.g., the even element s_i and its odd successor, s_{i+1}):

$$even_{j+1,i} = \frac{even_{j,i} + odd_{j,i}}{2}$$

The original value of the $odd_{j,i}$ element has been replaced by the difference between this element and its even predecessor. Simple algebra lets us recover the original value:

$$\mathbf{odd_{j,i} = even_{j,i} + odd_{j+1,i}}$$

Substituting this into the average, we get

$$even_{j+1,i} = \frac{even_{j,i} + even_{j,i} + odd_{j+1,i}}{2}$$

$$even_{j+1,i} = even_{j,i} + \frac{odd_{j+1,i}}{2}$$

ANALYSIS OF THE RESULTS

There are mainly three aspects that should be taken into account when discussing the results of the proposed method of text steganography. They are security, capacity and robustness. The authors simulated the proposed system and the results are shown in the figures 11-14. Where figure 18 shows the cover text, figure 19 is the secret message, figure 20 is the Unicode of the secret message and figure 21 and 22 are the stego texts using Single-bit Method and Double-bit Method. This method satisfies both security aspects and hiding capacity requirements. It generates the stego text with minimum degradation which is not very revealing to people about the existence of any hidden data, maintaining its security to the eavesdroppers. Although the embedding capacity of the proposed method depends upon the cover text structure but the embedding capacity can be maximized by incorporating more no of alphabets through minor pattern changes for mapping '00', '10', '01' or '11'.

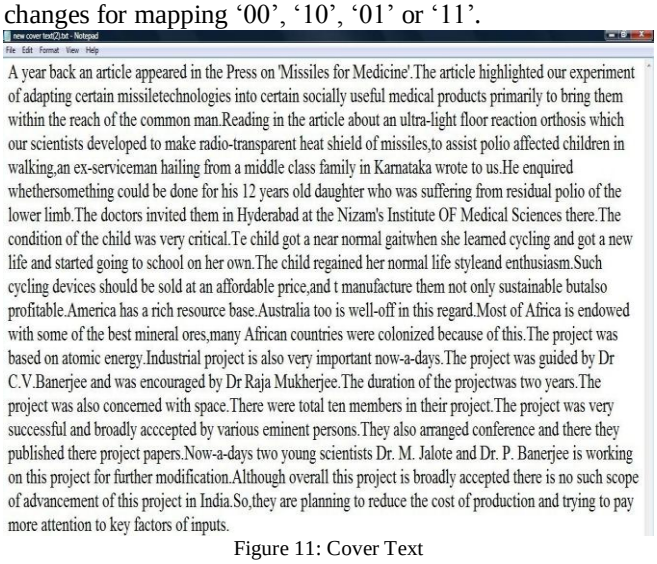

Figure 11: Cover Text

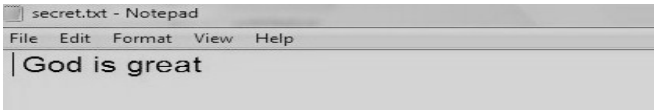

Figure 12: Message to be embedded

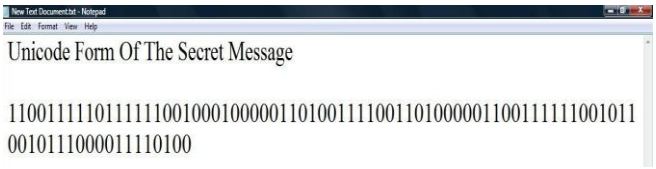

Figure 13: Unicode of the secret message

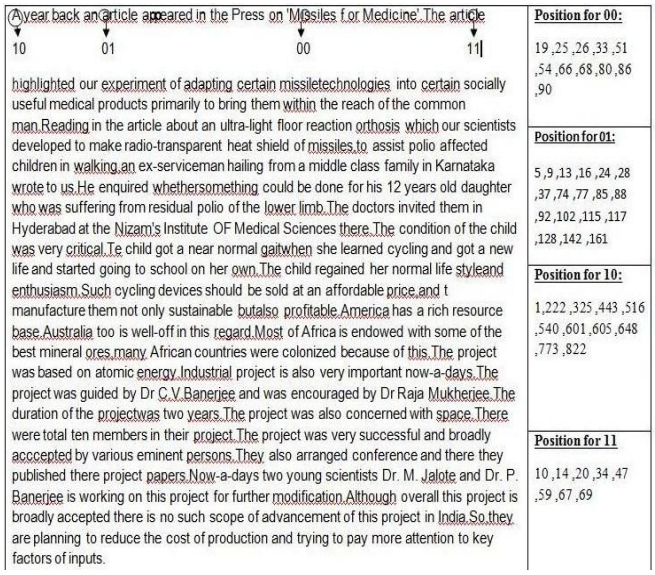

Figure 14: Stego Text with embedding positions

Similarity Measure of the Cover Text and Stego Text through Correlation

The most familiar measure of dependence between two quantities is the Pearson product-moment correlation coefficient [29], or "Pearson's correlation." It is obtained by dividing the covariance of the two variables by the product of their standard deviations. Karl Pearson developed the coefficient from a similar but slightly different idea by Francis Galton. The Pearson correlation is +1 in the case of a perfect positive (increasing) linear relationship (correlation), -1 in the case of a perfect decreasing (negative) linear relationship (anti correlation) , and some value between -1 and 1 in all other cases, indicating the degree of linear dependence between the variables. As it approaches zero there is less of a relationship (closer to uncorrelated). The closer the coefficient is to either -1 or 1, the stronger the correlation between the variables. If the variables are independent, Pearson's correlation coefficient is 0, but the converse is not true because the correlation coefficient detects only linear dependencies between two variables. If we have a series of n measurements of X and Y written as x_i and y_i where i = 1,2,…,n then the sample correlation coefficient can be used in Pearson correlation r between X and Y. The sample correlation coefficient is written as

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$$

where $\bar{x}$ and $\bar{y}$ are the sample means of X and Y, s_x and s_y are the sample standard deviations of X and Y.

Similarity Measure of the Cover Text and Stego Text through Jaro Winkler Distance

For comparing the similarity between cover text and the stego text, the Jaro-Winkler distance for measuring similarity between two strings has been computed. The Jaro-Winkler distance [30] is a measure of similarity between two strings. It is a variant of the Jaro distance metric [31], [32] and mainly used in the area of record linkage (duplicate detection). The higher the Jaro-Winkler distance for two strings is, the more similar the strings are. The score is normalized such that 0 equates to no similarity and 1 is an 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. Figure 15 below shows the Correlation coefficient and Jaro score for various size of cover text along with various size of the secret message of the proposed CALP method.

Comparison of CALP with other Text Steganography Methods.

a. No previous work focuses on keeping the increasing size of the embedding capacity and similarity between cover text and stego text generated based on different message sizes.
b. Besides most of text steganography methods are done based on some features of a specific language and not the universal one.

Proposed method for CALP based text steganography has been designed keeping in mind to overcome the above mentioned short comings.

a) Embedding capacity has been increased by mapping of two bits at a time instead of one.
b) Similarity measure between cover text and stego text has been inducted here through correlation and Jaro-Winkler distance.
c) Although this method has been used here for English text but this method can be used for any other languages which makes this method an universal one.

Secret Message size (in characters)	Cover Text Size (in characters)	Correlation Coefficient	Jaro Score
11	2800	.91	.98
14	2800	.90	.98
16	2800	.88	.97
9	2800	.92	.98
11	3200	.87	.98
14	3200	.87	.98
16	3200	.91	.98
9	3200	.92	.99
11	2400	.92	.96
14	2400	.92	.96
16	2400	.91	.96
9	2400	.90	.97
11	3000	.88	.97
14	3000	.89	.97
16	3000	89	.97
9	3000	.88	.96

Figure 15: CALP parameters

CONCLUDING REMARKS

In this paper the authors presented a novel approach of English text steganography method which is the improved version of the CALP. Stego text is generated by mapping each two bit of the binary sequence of the secret message through small texture/pattern changes of some alphabets of the cover text in order to achieve high level of security. From figure 15 it has been observed that CALP method generates the stego text with minimum or zero degradation as both the Jaro score and Correlation-coefficient value is very high. This property also enables the method to avoid the steganalysis. The proposed steganography technique through texture/pattern changing is a new approach for the English steganography and this methodology can be extended to any Indian language also.

REFERENCES

[1] Gustavus J. Simmons, "The Prisoners' Problem and the Subliminal Channel", in Proceedings of CRYPTO '83, pp 51-67. Plenum Press (1984).

[2] P. Wayner, "*Strong Theoretical Steganography*", Cryptologia, XIX(3), July 1995, pp. 285-299.

[3] J.T. Brassil, S. Low, N.F. Maxemchuk, and L. O'Gorman, "Electronic Marking and Identification Techniques to Discourage Document Copying", *IEEE Journal on Selected Areas in Communications*, vol. 13, Issue. 8, October 1995, pp. 1495-1504.

[4] "Stretching the Limits of Steganography", RJ Anderson, in Information Hiding, Springer Lecture Notes in Computer Science v 1174 (1996) pp 39-48.

[5] Kahn, The Codebreakers - the comprehensive history of secret communication from ancient times to the Internet, Scribner, New York (1996).

[6] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding", *IBM Systems Journal*, vol. 35, Issues 3&4, 1996, pp. 313-336.

[7] Scott Craver, "On Public-key Steganography in the Presence of an Active Warden," in Proceedings of 2nd International Workshop on Information Hiding, April 1998, Portland, Oregon, USA. pp. 355 - 368.

[8] 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.

[9] N. F. Johnson and S. Jajodia, "Steganography: seeing the unseen," IEEE Computer.,Feb., 26-34 (1998).

[10] L. M. Marvel, C. G. Boncelet, Jr. and C. T. Retter, "Spread spectrum image steganography," IEEE Trans. on Image Processing, 8(8), 1075-1083 (1999).

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

[12] Analysis of LSB Based Image Steganography Techniques ,R. Chandramouli, Nasir Memon, Proc. IEEE ICIP, 2001.

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

[14] An Evaluation of Image Based Steganography Methods,Kevin Curran, Kran Bailey, International Journal of Digital Evidence,Fall 2003.

[15] G. Doërr and J.L. Dugelay, "A Guide Tour of Video Watermarking", *Signal Processing: Image Communication*, vol. 18, Issue 4, 2003, pp. 263-282.

[16] K. Gopalan, "Audio steganography using bit modification", *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '03)*, vol. 2, 6-10 April 2003, pp. 421-424.

[17] M. Niimi, S. Minewaki, H. Noda, and E.Kawaguchi, "A Framework of Text-based Steganography Using SD-Form Semantics Model", *Pacific Rim Workshop on Digital Steganography 2003*, Kyushu Institute of Technology, Kitakyushu, Japan, July 3-4, 2003.

[18] Y. Kim, K. Moon, and I. Oh, "A Text Watermarking Algorithm based on Word Classification and Inter-word Space Statistics", *Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR'03),* 2003, pp. 775–779.

[19] A.M. Alattar and O.M. Alattar, "Watermarking electronic text documents containing justified paragraphs and irregular line spacing ", *Proceedings of SPIE - Volume5306, Security, Steganography, and Watermarking of Multimedia Contents VI*, June 2004, pp- 685-695.

[20] G. Doërr and J.L. Dugelay, "Security Pitfalls of Frameby-Frame Approaches to Video Watermarking", *IEEE Transactions on Signal Processing*, Supplement on Secure Media, vol. 52, Issue 10, 2004, pp. 2955-2964.

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

[22] M.H. Shirali-Shahreza and M. Shirali-Shahreza, "Text Steganography in Chat", *Proceedings of the Third IEEE/IFIP International Conference in Central Asia on Internet the Next Generation of Mobile, Wireless and Optical Communications Networks (ICI 2007)*, Tashkent, Uzbekistan, September 26-28, 2007.

[23] L.Y. Por and B. Delina, "Information Hiding: A New Approach in Text Steganography", *7th WSEAS International Conference on Applied Computer & Applied Computational Science*, April 2008, pp- 689-695.

[24] MohammadShirali-Shahreza: "Text Steganography by Changing Words Spelling" at ICACT 2008.

[25] "Study of Secure Steganography model" by Souvik Bhattacharyya and Gautam Sanyal at the proceedings of "International Conference on Advanced Computing & Communication Technologies (ICACCT-2008),Nov, 2008, Panipat, India"

[26] **"**An Image based Steganography model for promoting Global Cyber Security" by Souvik Bhattacharyya and Gautam Sanyal at the proceedings of "International Conference on Systemics,Cybernetics and Informatics (ICSCI-2009),Jan, 09,Hyderabad,India."

[27] **"**Implementation and Design of an Image based Steganographic model" by Souvik Bhattacharyya and Gautam Sanyal at the proceedings of " IEEE International Advance Computing Conference "(IACC-2009)"

[28] A Novel Approach to Develop a Secure Image based Steganographic Model using Integer Wavelet Transform" at the proceedings of International Conference on Recent Trends in Information, Telecommunication and Computing (ITC 2010)" by Souvik Bhattacharyya, Avinash Prasad Kshitij and Gautam Sanyal. (Indexed by IEEE Computer Society).

[29] S. Dowdy and S. Wearden. Statistics for research. Wiley. ISBN 0471086029, page 230, 1983.

[30] 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.

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

[32] W. E. Winkler. The state of record linkage and current researchproblems. Statistics of Income Division, Internal Revenue ServicePublication R99/04., 1999.

[33] Hiding Data in Text Through Changing in Alphabet Letter Patterns (CALP) by Souvik Bhattacharyya, Pabak Indu , Sanjana Dutta , Ayan Biswas and Gautam Sanyal at Journal of Global Research in Computer Science (JGRCS) VOL 2, NO 3 (2011) MARCH-2011.

About the authors

Souvik Bhattacharyya received his B.E. degree in Computer Science and Technology from B.E. College, Shibpur, India, presently known as Bengal Engineering and Science University (BESU) and M.Tech degree in Computer Science and Engineering from National Institute of Technology, Durgapur, India. Currently he is working as an Assistant Professor in Computer Science and Engineering Department at University Institute of Technology, The University of Burdwan. He has a good no of research publication in his credit. His areas of interest are Natural Language Processing, Network Security and Image Processing.

Sanjana Dutta is currently doing her B.E in Information Technology at University Institute Of Technology, The University of Burdwan.She is a final year student of this course and her areas of interest are E-Commerce,Database and Client Server Technology.

Ayan Biswas is currently doing her B.E in Information Technology at University Institute Of Technology, The University of Burdwan .He is a final year student of this course and hirs areas of interest are Database,Web Technology and Computer Network.

Pabak Indu is currently doing her B.E in Information Technology at University Institute Of Technology, The University of Burdwan .He is a final year student of this course and hirs areas of interest are Database, Web Technology and Computer Network.

Gautam Sanyal has received his B.E and M.Tech degree National Institute of Technology (NIT), Durgapur, India. He has received Ph.D (Engg.) from Jadavpur University, Kolkata, India, in the area of Robot Vision. He possesses an experience of more than 25 years in the field of teaching and research. He has published nearly 50 papers in International and National Journals / Conferences. Two Ph.Ds (Engg) have already been awarded under his guidance. At present he is guiding six Ph.Ds scholars in the field of Steganography, Cellular Network, High Performance Computing and Computer Vision. He has guided over 10 PG and 100 UG thesis. His research interests include Natural Language Processing, Stochastic modeling of network traffic, High Performance Computing, Computer Vision. He is presently working as a Professor in the department of Computer Science and Engineering and also holding the post of Dean (Students' Welfare) at National Institute of Technology, Durgapur, India.