International Journal of Engineering & Technology, 7 (4.19) (2018) 83-88

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A New Scheme of Representing a Secret Message Using Second Quotient Remainder Theorem in Text Steganography

Baharudin Osman[1*], Azman Yasin[2], Mohd Nizam Omar[3]

[1,2,3] School of Computing, College of Arts and Sciences, Universiti Utara Malaysia, 06010 UUM Sintok, Malaysia
**Corresponding Author Email: [1]bahaosman@uum.edu.my*

Abstract

Steganography is a method of concealing a secret message in a cover medium in such a way that the intruder is not able to detect the existence of the secret message. Due to extensive use of Internet and other communications media, the confidentiality and data integrity need a special means to protect it against unauthorized access. Several schemes were introduced by researchers to represent the value of a hidden message. The first scheme converts a secret message to ASCII code and is represented into *x,y* form which raises a redundancy representation for a repeating character of the secret message. The second scheme is using the octal value of secret message and converts it into a similar form. Another scheme modifies the previous scheme by adding a "+" and "-"symbols. These techniques are resulted in the limited range of the *x,y* values, repeating representation and a difficulty on generating a stegotext. To overcome the problems, the proposed scheme represents a secret message with a various representation and converts it into *x, y, z* form. A Second Quotient Remainder Theorem is used to converts the value to *x, y, z* forms. As a result, the range of the new representation is increased.

Keywords: *Text Steganography, Secret Message, Second Quotient Remainder Theorem.*

1. Introduction

The growth of a modern communication in present day requires a particular security especially on computer network. The volume of communication via Internet has increased significantly, and data traffic has raised security concerns of the data being transmitted [1]. Due to extensive use of internet and other communication channels, the confidentiality and data integrity need a special means to protect it against an unauthorized access [2], [3]. Researchers have been proposing various security methods such as cryptography, steganography, and watermarking in order to make sure the data exchange occurs in a secured way. Steganography is an ancient technique of hiding messages behind a covered object in such manner that the presence of the existing message cannot be detected by eavesdroppers. On the other hand, cryptography encrypts or scrambles a plain text by altering it into a cipher text using a secret code. According to [4], the existence of an enciphered message of cryptography can be visible to anyone and still attracts the consciousness of potential attackers. As such, many governments prohibit, completely create a law or limit the use of cryptosystem in their organization. Unfortunately, this restriction raises a new problem for their daily communication. This is where steganography comes in. Hence, both cryptography and steganography are the key techniques used to secure data transfer over the Internet [4]. However, in recent years, steganography has attracted more attention than others. It performed a communication in a healthy manner to avoid an attraction against eavesdroppers. This policy encourages a steganographic technique to be adapted in sensitive organizations instead of cryptography for a better communication [5].

2. Text Steganography Categories

Text, image, audio, video and network protocol are various media used as a carrier to hide information in steganography. Steganography can be divided into five different categories as shown in Fig. 1.

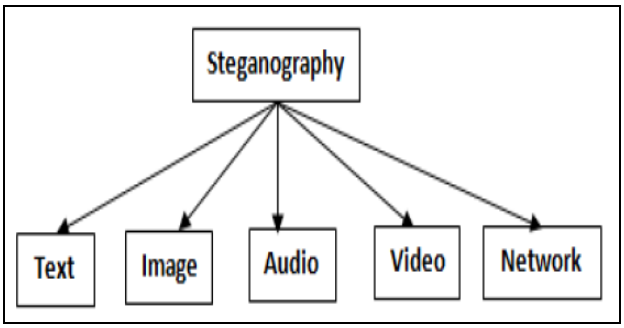


Fig. 1: Type of Steganography

Most of steganography works on image, video and audio as a cover medium since the large amount of bits can be embedded into it. Nonetheless, text steganography is considered to be the hardest, as the amount of redundant information presents in documents is low compared to the other cover media. However, the wide use of text

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

document [6], faster transmission over network, low processing power, smaller file size and using low bandwidth [7] insist text document to be a better choice as a cover medium by researchers. Hence, researches on text steganography are continuously attracting more researchers as text data are popularly used by people world-wide for daily work [1]. Text steganography can be classified into three different types of categories as shown in Fig. 2.

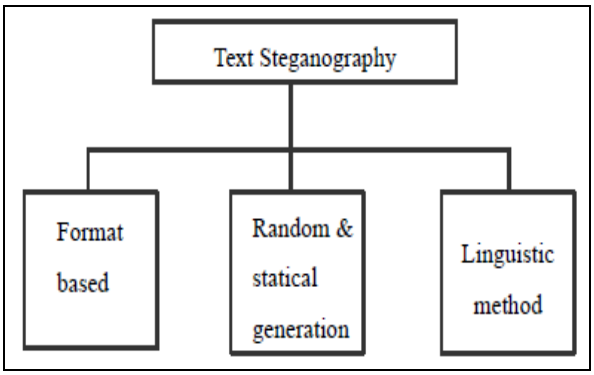


Fig. 2: Basic Categories of Text Steganography

2.1. Format based Method

This method normally modifies the existence of the cover text to hide data. The words or sentence remain as original but there might be some changes to formatting. Adding extra white space between words, changing font styles (bold, italic, underline, strikethrough etc.) and font colors, changing letter size, adding extra space between paragraphs, word and line shifting are some examples of using this method. Normally, this method is used to avoid any harm and to retain the 'value' of the cover-text.

2.2. Random and Statistical Generation

This method is used to generate cover-text automatically according to the statistical properties of language such as using probabilistic context-free grammar (PCFG) which generates word sequence according to the secret message to be hidden. The quality of the generated stego text depends directly on the quality of the grammar used. Another way to deal with this kind of technique is to create words that have similar factual properties like word length and letter frequency of a word in the cover text. The words are frequently produced without any lexical value.

2.3. Linguistic Method

This method specifically considers the linguistic properties of text to generate a stego text by using the syntax and semantics of a language, or the combination of them. Semantics is implemented by introducing a change in the meaning of the text. The synonym of word from dictionary is substituted to represent bits of the secret message. The synonyms convey the same meaning so they can be used in a better way to hide a message. For example: primary and secondary meanings of a word can be used in a text to hide a message. This will prevent an attacker from knowing that he is reading a cover text. It should be noted that both the sender and the receiver must have a complete list of words and their respective synonyms for encoding and decoding purposes. Another example is by using different spellings of a word such as American English and British English. The attraction of the attacker will not be attracted if syntactical methods are used in an appropriate manner. However, this method will alter the meaning of the cover text [8]. Semantic method is the one which does not destroys the hidden information even if an Optical Character Recognition (OCR) technique is used.

Syntactical method focuses on the syntax of the text to ensure that the structures of modifying a sentence are syntactically correct. The idea behind this technique is to embed a hidden bit in a punctuation mark at proper places in a cover text such as comma (,), semi colon (;), full stop (.) and etc. This technique can hide a small amount of hidden message due to limited punctuation marks. However, the inconsistency of using the punctuation marks is noticeable by readers [9]. As cited in [10], two disadvantages of using this methods are firstly, an intruder who has good knowledge of English may intercept the hidden message because he or she knows that the exact may be positioned in such marks in a text document, and secondly, this method has a low embedding capacity due to limited punctuations marks.

3. Related Works

The distribution probability of characters in English text documents can be classified into three categories as high, average and medium [11]. High occurrence probability characters are E,T,A,O,I,S with 5.0% to 9.6% occurrence frequency followed by an average probability characters that are H,N,D,R,L,U,W,M with occurrence frequency is 2.0% to 4.9%. The rest of the characters are the lowest occurrence probability with 0.04% to 1.69% occurrence frequency. However, space characters are the highest occurrence probability in text documents with 20.2% occurrence frequency.

The position of characters in text documents has been used by [12] to hide the bits of a secret message in certain characters. This technique only selects a few characters which are a, c, f, h, i, j, p, q and x to hide 2 bits of a secret message for each selected characters. The selection of characters f,j,x,p and q may cause the embedding capacity to decrease due to the probability occurrence of this characters in text document is low.

A researcher in [13] proposed a scheme to represent a character using mathematical model of numbering system. An ASCII value is represented in (x,y) form using the following formulae.

ASCII value => $(x,y) = (x*(x+1)/2) + y$

By using this formulae, a secret message "DYNAMITE" is represented by the following (x, y) values as shown in Table 1.

Table 1: Representation of (x,y) by [13]

Character	ASCII Value	(x,y) representation
D	68	(11,2)
Y	89	(12,11)
N	78	(11,12)
A	65	(10,10)
M	77	(11,11)
I	73	(11,7)
T	84	(12,6)
E	69	(11,3)

This method conceals a hidden message by plotting the value of (x,y) using a directed graph. The cost value is used to link from one node to another. The minimum cost represents the movement of the actual node.

A study by[14] has implemented the scheme by representing the value of (x,y) into a date format (DD/MM). The characters A to Z are converted to (x,y) form where the value of x and y represent DD and MM respectively as shown in Table 2.

Table 2: Representation of characters by [14]

Character	ASCII Value	(x,y)	Character	ASCII Value	(x,y)
A	65	(10,10)	N	78	(11,12)

B	66	(10,11)	O	79	(12,01)
C	67	(11,01)	P	80	(12,02)
D	68	(11,02)	Q	81	(12,03)
E	69	(11,03)	R	81	(12,04)
F	70	(11,04)	S	83	(12,05)
G	71	(11,05)	T	84	(12,06)
H	72	(11,06)	U	85	(12,07)
I	73	(11,07)	V	86	(12,08)
J	74	(11,08)	W	87	(12,09)
K	75	(11,09)	X	88	(12,10)
L	76	(11,10)	Y	89	(12,11)
M	77	(11,11)	Z	90	(12,12)

Table 2 shows that, the value of *x* and *y* is in the range of 10 to 12 and 1 to 12 respectively. A study by [15] mapped the value of (*x*,*y*) with a date format (DD/MM) to hide a secret message and then generate a stego text. Fig. 3 shows the secret message "SOLARIS" is hidden by generating a suitable sentence to produce a stego text using this implementation. The value of year (YYYY) can be replaced with any suitable number. The disadvantage of this technique is the difficulty to generate a suitable sentence related to the particular DD/MM value. Other limitations of this technique are the repeating representation of the same characters and a limited range of *x* and *y* value.

"Narayan Gopal Guruwacharya was a prominent popular singer of nepali music, who was died on 12/05/1990. Swami Vivekananda, a famous Indian hindu monk was born on 12/01/1893. The famous Indian Bollywood actor, Amitabh Bachchan was born on 11/10/1942. A famous writer R. K. Narayan was born on 10/10/1942. A famous Indian scientist Srinivasa Ramanujan was born on 12/04/1920. Indian railway minister, Suresh Prabhu was born on 11/07/1953. International Nurse's Day will be held on every year 12th May."

Fig. 3: Stego Text using [15]

Next study is done by [16] used an octal value to represent characters A to Z. The study used the same formulae to improve the range of x and y values. This octal representation of an actual text can be applied on any combination of characters A to Z, a to z, 0 to 9 or other special characters [15]. In this technique, the value of spaces is represented by 0408 and numbers 0 to 9 is represented by 0008 to 0118 respectively. Table 3 shows the sample of a hidden message "Ami 1 Ghanta" using this technique.

Table 3: Representation of (x,y) using [16]

Character	Octal value	(*x*,*y*) value
A	101	(13,10)
m	155	(17,02)
i	151	(16,15)
space	040	040
1	001	001
space	040	040
G	107	(14,02)
h	150	(16,14)
a	141	(16,05)
n	156	(17,03)
t	164	(17,11)
a	141	(16,05)

The x and y value is used to embed a hidden message and to generate a stego text. Fig. 4 shows the sample of the generated stego text using the same hidden message.

"Dear customer blue print of your bungalow is ready. The specified bedroom size is (13,10) lobby (17,02) second bedroom (16,15) chairs are (040) (001) (040) storeroom (14,02) bedroom (16,14) two bathrooms (16,05) and (17,03) bedroom (17,11) (16,05) fourth chair (040) ……"

Fig. 4: Stego Text using [16]

This technique generates a stego text which contains various numbers which are difficult to generate a stego text especially for a long hidden message, and it tends to be suspicious.

In conclusion, these three techniques generate a redundancy value for repeating characters of a hidden message. For instance, Table 4 shows a hidden message "MEETYOUATTEN" as represented by these two techniques.

Table 4: Representation of (x,y) using [15] and [16]

Character	**Researcher** [15]		**Researcher** [16]	
	ASCII	***(x,y)***	**OCTAL**	***(x,y)***
M	65	(11,11)	115	(14,10)
E	69	(11,03)	105	(13,14)
E	69	(11,03)	105	(13,14)
T	84	(12,06)	124	(15,04)
Y	89	(12,11)	131	(15,11)
O	79	(12,01)	117	(14,12)
U	85	(12,07)	125	(15,05)
A	65	(10,10)	101	(13,10)
T	84	(12,06)	124	(15,04)
T	84	(12,06)	124	(15,04)
E	69	(11,03)	105	(13,14)
N	78	(11,12)	116	(14,11)

Table 4 shows in [15], the repeating characters "E" and "T" were represented by a repeating value (11,03) and (12,06) respectively. The same representation of characters "E" and "T" occurs in the following study by the author [16] but with different values of x and y as highlighted in Table 4.

An authors in [14] revamped the above formula by adding a '+' and '–' symbols to the converting ASCII value. A few additional symbols in consecutive manner (++, -+, --, +-) have been added to the values of (x,y) to generate the cover text. For example (11,11), (11,03), (11,03),(12,06) was represented by a new representation which are (+11,+11), (-11,+03), (-11,-03),(+12,-06) . The following table shows the new representation of (x,y) using this scheme.

Table 5: Representation of (x,y) using [14]

Character	ASCII Value	Using (x (x + 1)) / 2 + y) formula	Using additional conservative symbols (++,-+,--,+-)
M	65	(11,11)	(+11,+11)
E	69	(11,03)	(-11,+03)
E	69	(11,03)	(-11,-03)
T	84	(12,06)	(+12,-06)
Y	89	(12,11)	(+12,+11)
O	79	(12,01)	(-12,+01)
U	85	(12,07)	(-12,-07)
A	65	(10,10)	(+10,-10)
T	84	(12,06)	(+12,+06)
T	84	(12,06)	(-12,+06)
E	69	(11,03)	(-11,-03)
N	78	(11,12)	(+11,-12)

It clearly shows that the repeating characters, "E" and "T" were represented by different representations. However, the second and the third character of "E" still have a similar representation of (-11,-03). The disadvantage of this scheme is, the possibility of repeating characters that has a similar representation is high when the size of the hidden message increases.

A Quotient Remainder (QR) technique is used by [17] by converting an encrypted hidden message value to a binary number. A group of 4

bits of number was hidden in html space by mapping this bit with a 10 fixed space code. An average of embedding capacity for this technique is 99.9% with 11.428Kb of cover text file. Although the embedding capacity is high, this technique required a huge cover text file for the embedding process. In future, the researcher may need to extend further to improve the robustness of this technique.

4. Proposed Method

According to [18], hiding information in a consecutive location will allow an intruder to extract the secret message from a stego text. Hence, a random number can play a significant role which has been used in encryption for various network security applications. Random Number Generators (RNG) are divided into three types; the first type is the True Random Number Generators (TRNGs), in which their output cannot be reproduced [19]. TRNGs are based on a physical experiment such as coin flipped 100 times and the result recorded as binary bit. So, it is impossible to generate the same bit again by using the same way. The second type is Pseudorandom Number Generators (PRNG) which generates sequences that are computed from an initial seed, and produces a sequence of output bits using a deterministic algorithm. Typically, PRNG can work by a feedback path. PRNG uses the following formulae

$S[i + 1] = S[i] * A + B \bmod m$; $i = 0,1,2,3 \ldots$.

where

$S[i], A, B \in \{ 0, 1, 2, 3 \ldots\ldots m - 1\}$

m : integer constant

The value of random number, S is denoted as

$S = \{ Si \mid 1 \leq i \leq m \}$

The last type is Pseudorandom Number Function (PRNF). This is used to produce a pseudorandom string of bits of some fixed length such as fixed length keys.

Our proposed scheme converts an ASCII code of a hidden message into (x,y,z) representation using a Quotient Reminder (QR) theorem. The ASCII code of hidden message is represented by a random value to avoid a repeating representation of similar characters of a hidden message. According to [20], it would be a better idea to place the characters at desired random positions rather than placing the alphabet in particular positions, making it difficult for the attacker to break the hidden message. Hence, the message is more secured. This various representation of a repetition character may increase the range of x,y,z representation and also the embedding capacity. The proposed technique is suitable for any languages that uses A to Z characters like English, Bahasa Melayu, Bahasa Indonesia or other languages.

The Quotient Remainder theorem says: Given any integer V, and a positive integer B, a unique integers Q and R might exist such as

$$V = B * Q + R \text{ where } 0 \leq R < B \quad (1)$$

For instance, if V = 317 and B = 15

21_ → Q
15) 317 → V
315
2 → R

If the quotient , Q is divided for a second time, we will get

1_ → Q'
15) 21 → Q
15
6 → R'

$$\text{hence, } Q = BQ' + R' \quad (2)$$

if we replace (2) in (1), hence, V = BQ + R can be written as

$$V = B(BQ' + R') + R \quad (3)$$

If $V = v$, $B = b$, $Q' = q'$, $R' = r'$ and $R = r$, hence equation (3) can be written as follow:

$$v = b(bq' + r') + r \quad (4)$$

where

b : constant value
q' : the quotient of second division
r : the remainder of first division
r' : the remainder of second division

if $q' = x$, $r' = y$ and $r = z$, hence equation (4) can be written as follow

$$v = b(bx + y) + z \quad (5)$$

Equation (5) is used to convert any value to (x,y,z) form and vice versa. This equation is used in the proposed technique to convert a random value into (x,y,z) form to avoid a redundant representation of a repeating hidden message characters.

5. Analysis and Result

The location value of each character in cover text will represent a homographic value of each character. All characters are counted as a location value except a white space.

For instance, Fig. 5 and Fig. 6 show a cover text sample and a part of homographic values of the cover text sample.

Production at the huge nickel deposit at Voisey's Bay in remote Labrador is still years away, but already it risks falling behind schedule because of environmental concerns and pressure from aboriginal groups.

Inco Ltd, the Toronto-based nickel giant that won control over the spectacular nickel, copper and cobalt property after a bidding war last spring, planned to start open pit production by 1998 and full-scale underground mining by 2000.

Fig. 5: Sample of cover text

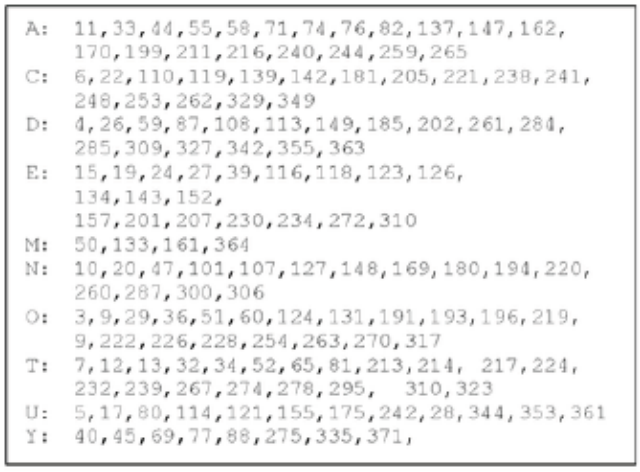


Fig. 6: Part of homographic value of a cover text Sample

Fig. 6 shows each character in cover text is assigned with a various values of the location of a cover text. This value will represent a random number of a character of a hidden message. A particular random number, v using PRNG is used during assigning the value. This random value, v is converted into x,y,z form using equation 5.

Table 6 shows the sample of a hidden message "MEETYOUATTEN" represented in form of (*x*,*y*,*z*) using this scheme by assuming *b* = 15.

Table 6: The representation of (*x*,*y*,*z*) using the proposed scheme

Character	Random Value	Value of *(x,y,z)*
M	364	(1,9,4)
E	157	(0,10,7)
E	272	(1,3,2)
T	310	(1,5,10)
Y	77	(0,5,2)
O	317	(1,6,2)
U	175	(0,11,10)
A	240	(1,1,0)
T	323	(1,6,8)
T	81	(0,5,6)
E	152	(0,10,2
N	203	(1,13,8)

Table 6 shows the value of *x*,*y*,*z* as represented by each character of the hidden message. It seems that repetitive characters "E" and "T" are represented by a different value (0,10,7), (1,3,2), (0,10,2) and (1,5,10), (1,6,8), (0,5,6) respectively.

By using this scheme, the possibility of representing the same value for a repetitive characters can be avoided. Table 7 shows various representations of a hidden message "MEETYOUATTEN" using the above cover text.

Table 7: Various representation value of hidden message using the proposed scheme

	M	E	E	T	Y	O	U	A	T	T	E	N
C1	161	27	39	7	88	326	361	44	224	316	257	132
C2	364	104	134	267	69	140	344	240	12	184	112	362
C3	133	104	70	13	40	360	17	82	214	213	104	107
C4	364	352	85	184	275	317	114	11	190	274	279	20
C5	364	230	308	52	69	164	328	58	34	81	157	362
C6	133	116	126	12	88	228	175	44	12	214	272	127

Table 7 shows various representations of a hidden message characters. For every C_n, a repetitive character is represented by a different value. For instance, for C_5 representation, characters E and T are represented by various values which are 230, 308, 157 and 52, 34, 81 respectively. These values will be converted to (*x*,*y*,*z*) form using equation (5). The most important idea in this scheme is the representation of different values for a repeating character in the hidden message.

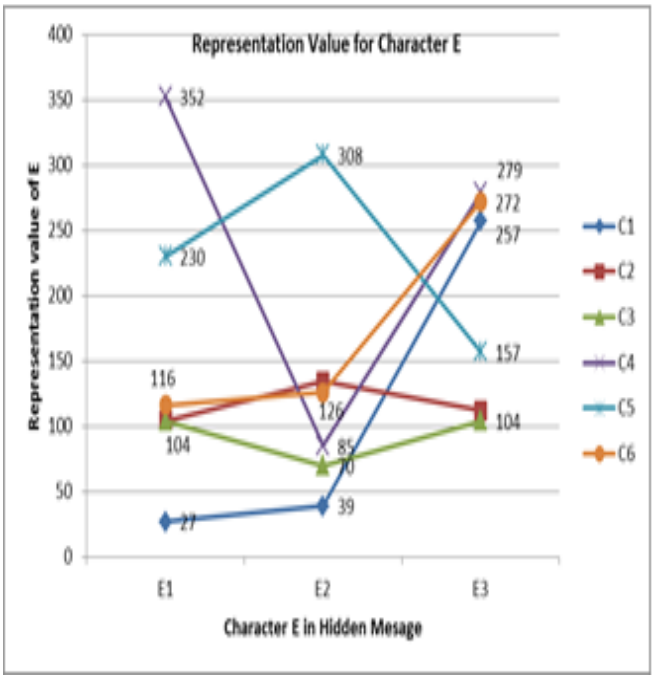


Fig. 7: Sample Representation Value for Character 'E'

Figure 7 shows 6 different representations of characters "E" in the hidden message. All of the E characters are represented by different values for each Cn. For each Cn there is no horizontal straight line which represented a repetitive value. For instance, for C1, the value of E1, E2 and E3 is 27,39 and 257 respectively.

6. Conclusion

In this paper, we present a new scheme to represent a hidden message character in form of x,y,z. This scheme has an advantage of representing various different values for a similar characters of a hidden message. This proposed scheme tends to boost up the embedding capacity by using this multiple representation of the repetitive characters of the hidden message. In future research, this study will expand the scheme by mapping the hidden characters with the value of x,y and z and hide it in the cover text to generate a stego text.

References

[1] S. Mahato, D. A. Khan, and D. K. Yadav, "A Modified Approach to Data Hiding in Microsoft Word Documents by Change-Tracking Technique," Journal of King Saud University Computing and Information Science, 2017.

[2] R. Gupta, S. Gupta, and A. Singhal, "Importance and Techniques of Information Hiding: A Review", International Journal Computer Trends Technology, vol. 9, no. 5, pp. 260–265, 2014.

[3] K. I. Rahmani, A. Kumar, and G. Manisha, "Study of Cryptography and Steganography System", International Journal Engineering Computing Science, vol. 4, no. 8, pp. 13685–13687, 2015.

[4] M. Htet and S. W. Phyo, "A Novel Text Steganographic Technique Using Specific Alphabets", Journal of Computer Science, vol. 2, no. 1, pp. 1–11, 2016.

[5] R. B. Krishnan, P. K. Thandra, and M. S. Baba, "An Overview of Text Steganography", in 4th International Conference on Signal Processing, Communications and Networking (ICSCN), 2017, pp. 1–5.

[6] W. Bhaya, A. M. Rahma, and D. Al-nasrawi, "Text Steganography Based on Font Type in MS-Word Documents", Journal of Computer Science, vol. 9, no. 7, pp. 898–904, 2013.

[7] Shivani, V. K. Yadav, and S. Batham, "A Novel Approach of Bulk Data Hiding using Text Steganography", Procedia Computer Science, vol. 57, pp. 1401–1410, 2015.

[8] M. Khairullah, "A Novel Text Steganography System in Financial Statements", International Journal of Database Theory Applications, vol. 7, no. 5, pp. 123–132, 2014.

[9] I. Stojanov, A. Mileva, and I. Stojanovi, "A New Property Coding in Text Steganography of Microsoft Word Documents A New Property Coding in Text Steganography of Microsoft Word Documents", in

SECURWARE 2014 : The Eighth International Conference on Emerging Security Information, Systems and Technologies, 2014, no. November.

[10] S. Chaudhary, M. Dave, and A. Sanghi, "Review of Linguistic Text Steganographic Methods", International Journal of Recent Innovation Trends Computer Communication., vol. 4, no. 7, pp. 377–381, 2016.

[11] B. K. Ramakrishnan, P. K. Thandra, and A. V. S. M. Srinivasula, "Text steganography : A Novel Character-Level Embedding Algorithm Using Font Attribute," in Security And Communcation Networks, 2017.

[12] S. Bhattacharyya, P. Indu, S. Dutta, A. Biswas, and G. Sanyal, "Text Steganography using CALP with High Embedding Capacity," J. Glob. Res. Comput. Sci., vol. 2, no. 4, pp. 29–36, 2011.

[13] K. K. Mandal, A. Jana, and V. Agarwal, "A New Approach of Text Steganography Based on Mathematical Model of Number System," in International Conference on Circuit, Power and Computing Technologies, 2014, pp. 1737–1741.

[14] K. K. Mandal, S. Koley, and S. Dhar, "A Mathematical Model for Secret Message Passing Using Steganography," in 2016 IEEE International Conference on Computational Intelligence and Computing Research, 2016, vol., no., pp. 1–6.

[15] S. Koley and K. K. Mandal, "A Novel Approach of Secret Message Passing Through Text Steganography," in International Conference on "Signal Processing, Communication, Power and Embedded System (SCOPES-2016), 2016.

[16] S. Koley and K. K. Mandal, "Number System Oriented Text Steganography in Various Language for Short Messages," in Computational Intelligence, Communications, and Business Analytics: International Conference, 2017, 1st ed., vol. 1, pp. 552–566.

[17] M. A. Tariq, A. T. Abbasi, A. Khan, and B. Ahmad, "Boosting the Capacity of Web based Steganography by Utilizing Html Space Codes : A blind Steganography Approach Boosting the Capacity of Web based Steganography by Utilizing Html Space Codes : A blind Steganography Approach," IT Ind., vol. 5, no. December, pp. 29–36, 2017.

[18] R. Ray, J. Sanyal, D. Das, and A. Nath, "A New Challenge of Hiding any Encrypted Secret Message Inside any Text / ASCII File or in MS Word File : RJDA Algorithm," in International Conference on Communication System and Network Technologies, 2012, pp. 889–893.

[19] M. Y. Elmahi, T. M. Wahbi, and M. H. Sayed, "Text Steganography Using Compression and Random Number Generators," Int. J. Comput. Appl. Technol. Res., vol. 6, no. 6, pp. 259–263, 2017.

[20] M. J. Stephen, P. Reddy, D. Naidu, S. Sonali, and Heymaraju, "More Secured Text Transmission with Dual Phase Message Morphing Algorithm," in Proceedings of the International Conference on Information Systems Design and Intelligent Applications, 2012, pp. 845–852.