Engineering Science and Technology, an International Journal 20 (2017) 72–79

HOSTED BY

Contents lists available at ScienceDirect

Engineering Science and Technology, an International Journal

journal homepage: www.elsevier.com/locate/jestch

Full Length Article

A high capacity text steganography scheme based on LZW compression and color coding

Aruna Malik *, Geeta Sikka, Harsh K. Verma

Department of Computer Science & Engineering, National Institute of Technology, Jalandhar, India

ARTICLE INFO

Article history:
Received 15 February 2016
Revised 13 June 2016
Accepted 17 June 2016
Available online 9 August 2016

Keywords:
Color mapping
Text steganography
LZW compression
Capacity

ABSTRACT

In this paper, capacity and security issues of text steganography have been considered by employing LZW compression technique and color coding based approach. The proposed technique uses the forward mail platform to hide the secret data. This algorithm first compresses secret data and then hides the compressed secret data into the email addresses and also in the cover message of the email. The secret data bits are embedded in the message (or cover text) by making it colored using a color coding table. Experimental results show that the proposed method not only produces a high embedding capacity but also reduces computational complexity. Moreover, the security of the proposed method is significantly improved by employing stego keys. The superiority of the proposed method has been experimentally verified by comparing with recently developed existing techniques.

1. Introduction

With the expeditious progress of computer technology and the widespread usage of the Internet, it becomes more and more suitable for people to access and interchange all kinds of multimedia information like audio, video, and images. However, distribution of these types of information over public network (i.e., Internet) makes them vulnerable to attack. Thus, there is a need to have solution which can protect sensitive data. One such method is data hiding (information hiding) which plays an important role in information security over the internet. Data hiding is a process of embedding the secret data imperceptibly into the cover media by minimally modifying the elements of the cover media. Generally, data hiding involves both watermarking and steganography [1–3]. A watermarking scheme imperceptibly alters a cover object to embed a message about the cover object (e.g., owner's identifier). The main objective of the watermarking is to attain a high level of robustness i.e. it should be very difficult to remove a watermark without degrading the quality of the data object. Watermarking is mainly used for copyright protection, broadcast monitoring, document and image security, forensics, and piracy deterrence. On the other hand, steganography is the art and science of hiding secret communication [1]. A steganographic system thus embeds secret content in cover media (like text, image, audio, and video) so that its existence is not detected to an eavesdropper.

In text based steganographic methods, text is used as a cover media for hiding the secret data. Text steganography is one of the hardest areas of data hiding, since the human eye is very susceptible to any change between the original and the modified texts (stego-texts) and it can be easily detected [4]. There are mainly two parameters namely capacity (or bit rate), and security to analyze the performance of any text steganographic method [4]. Capacity refers to the amount of secret data that can be concealed in the carrier, and security relates to the ability of a masquerader to figure out the concealed information, In this paper, both capacity and security issues have been considered to analyze the performance of the proposed text based steganography method. The main objective of this work is to obtain significant increment in the amount of secret data which is hidden in the cover medium and also design and uses stego keys for security improvement. In order to achieve this objective, a forward mail platform or more specifically the email ids and the cover message both are used to hide the secret data. The message is chosen from a text base which contains wishes messages, poems, and jokes etc., which are mainly used at the happy or pious occasions or moments. The message is made colorful while hiding the secret data according to the secret data bit stream. Thus, there is no need to modify the semantic/syntax of the chosen message in terms of content. In the proposed work, for capacity increment LZW data compression technique is used for compressing the secret data, because LZW is frequently used in literature and it has good compression ratio. At the initial level,

* Corresponding author.
E-mail addresses: arunacsrke@gmail.com (A. Malik), sikkag@nitj.ac.in (G. Sikka), vermah@nitj.ac.in (H.K. Verma).

Peer review under responsibility of Karabuk University.

http://dx.doi.org/10.1016/j.jestch.2016.06.005
2215-0986/

Table 1
Color mapping.

S.No.	Color Name	Color	Bit representation
1	Red		0
2	Green		0
3	Blue		0
4	Aqua		0
5	Pink		0
6	Black		1
7	Dark Yellow		1
8	Indigo		1
9	Dark red		1
10	Lavender		1

LZW reads the data and attempts to match a sequence of data bytes as large as possible with an encoded string from the dictionary. The matched data sequence and its succeeding character are grouped together and then added to the dictionary for encoding later data sequences [5–7]. In proposed work, LZW is directly applied to the secret data, but in [8] it is applied after some mapping operations. Thus, the proposed method decreases the computational complexity and also increases the capacity of the secret data. The generated overhead is also diminished because of bypassing the mapping operation used in [8]. For the second purpose i.e., security improvement, stego-keys are employed. The stego keys are divided into two categories, namely constructed stego key which is used during embedding phase and previously constructed global stego key which is shared between sender and the receiver both beforehand. For hiding the secret data into the message of the email, the color coding table is used as shown in Table 1. For each bit, the table contains a set of colors. Thus, the table is partitioned into two parts, first part corresponds to the '0' bit and the other corresponds to '1' of the secret data bit stream. The element of the message is made colored according to the bit values of the bit stream using color coding table. Thus, the secret data is embedded in the message even without modifying the contents of the message. The rest of the paper is organized as follows. Section 2 will explain the related work. Section 3 discusses the proposed work and in Section 4, experimental results are discussed. Finally, in Section 5, the paper is concluded.

2. Related work

In this section, we discuss some of the well-known text-based steganographic techniques for different languages. Wayner [9,10] proposes a mimic function based technique. In this technique, the inverse of Huffman code is used by inserting a data stream of randomly distributed bits. For improving the performance, it makes use of Van Wijngaarden Grammar and Context Free Grammar. Though, it provides resiliency against statistical attacks yet it suffers from invalid syntax problems. Maher [11] discuss a text based data hiding method which is known as TEXTO. This technique transformed PGP ASCII-armored ASCII data into English sentences. It converts the secret data into English words. Therefore, this method resembles with substitution cipher and reduces suspicion over the message. This work is extended into the articles like [12–14] which uses synonyms-based approach. In this technique, the cover text appears as original which has an appropriate accuracy of the chosen synonyms. Sun et al. [15] propose L–R scheme which uses the left and right components of Chinese characters. To conceal the secret data, the left and right components of characters are selected as candidates. If the secret data bit is "1", the scheme modifies the candidate by adjusting the space between the left and right components otherwise leaves unchanged. To improve the L–R scheme in terms of hiding capacity, Wang et al. [16] modify the scheme [15] by incorporating the up and down structure of Chinese characters as an extra candidate set. Apart from this, a reversible function is also added to obtain the original cover text after the initial hidden secret data has been extracted. Later, Wang et al. [17] used emotional icons (also called emoticons) in chat rooms over the Internet to hide the secret data. In this method, both the sender and the receiver design a table collaboratively that will be with them during their communication. The table consists of emoticons which are classified into several sets according to their meaning (like cry, smile, and laugh) and every emoticon belongs to one set. Each individual set is provided with a unique order number which is further used for hiding the secret data. Stutsman et al. [18] discuss an approach which uses noise or error to hide the secret data. The noise is naturally encountered in a machine translation (MT). The secret message is embedded by performing a substitution procedure on the translated text using translation variations of multiple MT systems. Samphaiboon [19] suggests a steganographic scheme for short text message. This method broadcasts the secret messages to multiple receivers in different locations at the same time via a stream of running short text messages displayed on a media output screen. It hides four secret data bits in the short text message. The receiver can extract the embedded bits, even without using OCR. Desoky [20] discusses a method called Listega which makes use of a textual list to hide the secret data by exploiting itemized data. In this method, firstly the message is encoded and then it is assigned to legitimate items in order to produce a stego text in the form of the litany. Thus, the stego-text in the form of the litany of items is linguistically and logically legitimate. To enhance the security, it used combinatorial based coding. The method maintains the originality of the cover text but suffers from the complexity point of view during extraction process. Another method known as UniSpach which uses spaces to hide the secret data is given by Por et al. [21]. It uses Unicode space characters to embed the secret data in Microsoft Word document. Additionally, to encode payload white spaces are taken as they appear throughout the document and the manipulation of white spaces have a trivial effect on the visual appearance of the document. Secret data is embedded using Unicode space characters [21] into spaces wherever spaces are found.

Rajeev et al. [22] discuss an efficient text steganographic scheme using Unicode characters which hides the secret data into the Microsoft word document. It hides the secret data into inter-word, inter-sentence, end of line, and inter-paragraph spacing. Firstly, the scheme selects Unicode space characters which do not give any awkward presence on visual attacks and then, it combines the selected characters with regular space characters to map the secret data bits to each combination. It normalizes the width of the Unicode character by reducing their font size so that their width equals the width of hair or Six-Per-Em Unicode characters. This in turn, increases the hiding capacity as the numbers of Unicode spaces which are embedded into the end of line and inter-paragraph spacing are increased. Later, Rajeev et al. [23] discuss an email based high capacity text steganography method using combinatorial compression. This method makes use of forward email platform to hide the secret data in email addresses. It uses the combination of Burrows Wheeler Transform (BWT) + Move to Forward (MTF) + LZW coding algorithm to increase the hiding capacity. To further increase the capacity, the numbers of characters of email id are also used to refer the secret data bits. Furthermore, the method adds some random characters just before the '@' symbol of email ids to increase the randomness. In 2016, Rajeev et al. [24] discuss a high capacity text based steganographic method using Huffman compression. The forward email platform is used to hide the secret data. It makes use of the number of characters used in email id to indicate the hidden secret data bits. So, to make optimal utilization of number of characters in email ids, the characters added to the email id, to indicate the secret data bits are taken from the processed secret data. Hence, the hiding capacity is

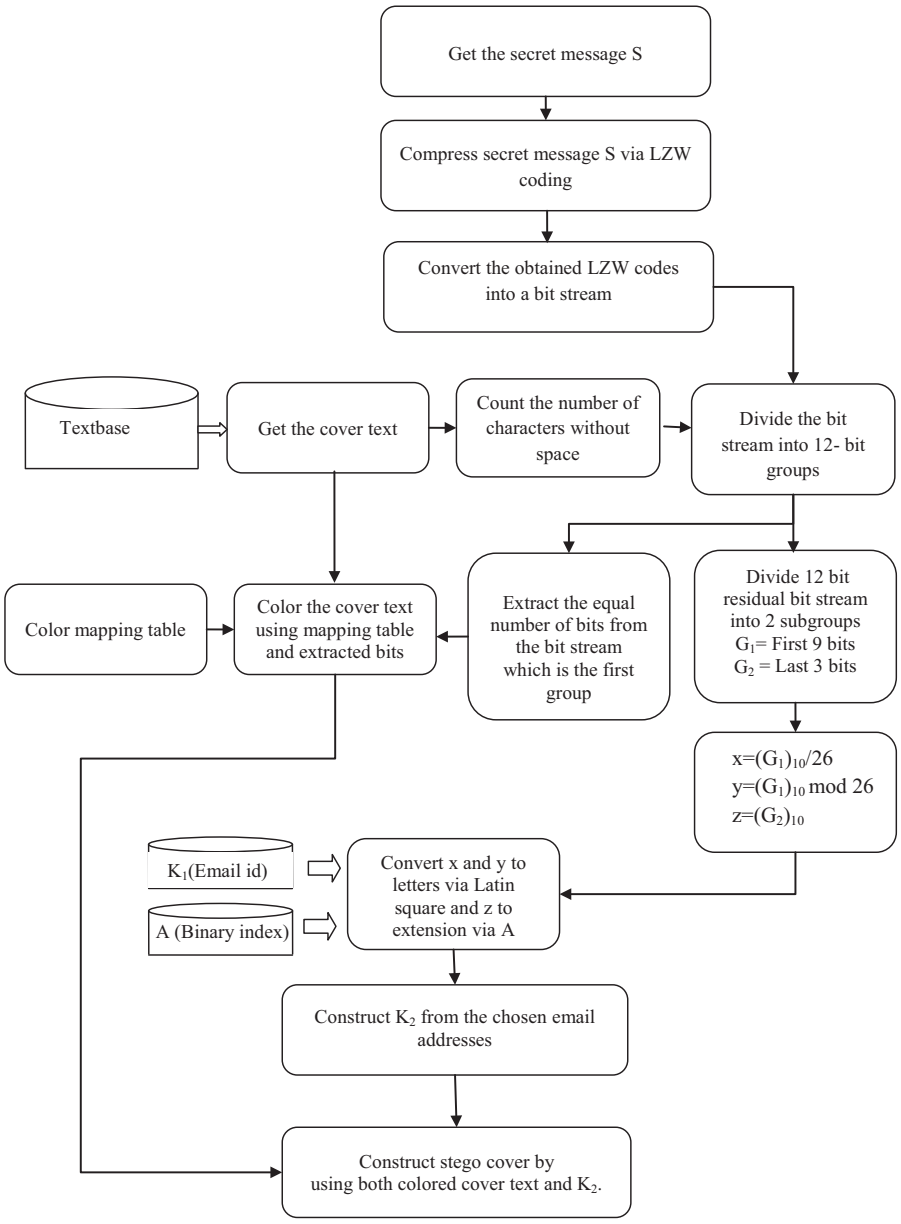

Fig. 1. Flowchart for embedding phase.

further increased. The new characters are appended just before the '@' symbol of email ids. Tutuncu et al. [25] discuss a text steganography method which is combination of lossless compression techniques and Vigenere cipher. It makes the use of email addresses to be the keys to reconstruct the secret message which has been embedded into the email text. After selecting the cover text that has highest repetition pattern regarding to the secret message the distance matrix is formed. The members of distance matrix are compressed by considering Run Length Encoding (RLE) + BWT + MTF + Run Length Encoding + Arithmetic Encoding lossless compression algorithms sequence. Later on, Latin Square is used to form stego key 1 and then Vigenere table is used to increase complexity of extracting stego key 1. Final step is to choose e-mail addresses by using stego key 1 and stego key 2 to embed secret message into forward e-mail platform. Mohamed [26] proposes a new steganographic algorithm for Arabic text based on features of Arabic text.

Satir and Isik [8] discuss a LZW compression based text steganography method to hide the secret data. This method hides the secret data into the email ids which are listed in Cc. During processing, the method incurs a lot of overhead which comes as a barrier in the path of achieving high hiding capacity. This method also suffered from complexity point of view. Another shortcoming of the Satir and Isik technique is that all the elements which are present in the secret data must also be available in the cover text otherwise the mapping operation will not get completed. Thus, before selecting the cover text, the sender has to look at this aspect of both the secret data and the cover text.

In our proposed work, capacity and security issues have been considered. The LZW algorithm is directly applied on the secret text and the obtained bit stream is hidden into email ids and in the message of the email. A color coding table is used to hide the secret data bits into the cover text of the email thus the notion of the content is not modified. The proposed method increases the hiding capacity and also reduces computational complexity. Moreover, security is also improved by employing stego keys. In the next section, the proposed method is discussed.

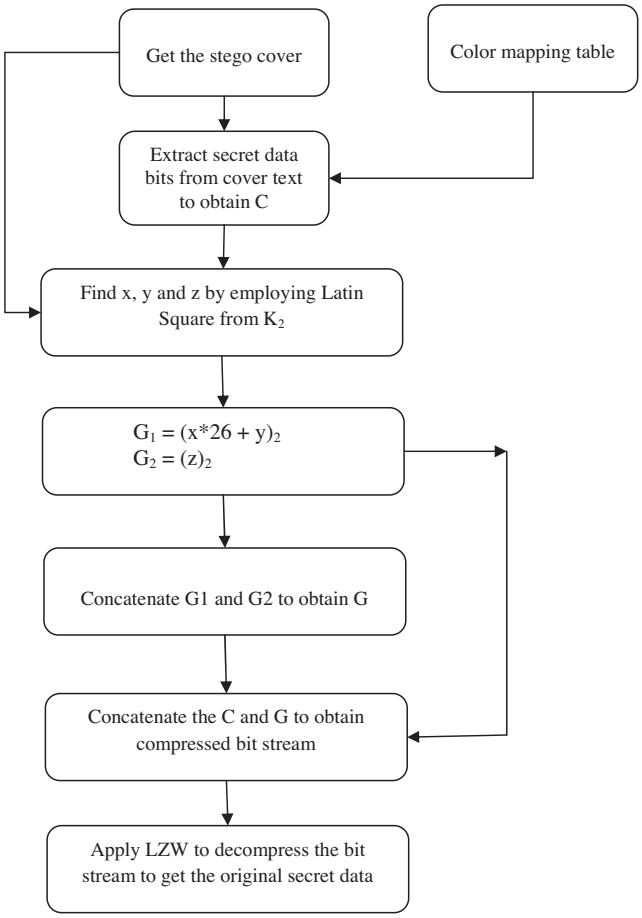

Fig. 2. Flowchart for extracting phase.

3. Proposed method

This section explains the proposed method which is divided into two phases embedding phase and extraction phase. Here, brief descriptions of the variables that are used in proposed technique are given below.

Let ***S***: secret message

T: Cover texts or message of email

K_1: A set of email addresses shared between the sender and the receiver which plays the role of public stego-key.

K_1 represent the set of email addresses which consists of the combinations of every pair of letters. Therefore, the maximum index can be $26 \times 26 = 676$. We can represent K_1 as follows:

K_1 = [aa...@gmail.com, ab...@gmail.com, ac...@gmail.com, ..., zv...@gmail.com, zy...@gmail.com, zz...@gmail.com]

A: Set of the second parts of email addresses (email extensions) such as hotmail.com, yahoo.com, etc.

A: [gmail.com (000), hotmail.com (001), yahoo.com (010), rediffmail.com (011), btinternet.com (100), aol.com (101), msn.com (110), verizon.net (111)]

3.1. Color coding table

This table is used to map secret data bits with the color and is shown as Table 1. It consists of four columns: A first column is a serial number which manages the rotation of the color during the mapping of the secret data bits onto the cover text in the form of color. The second column provides the color name and the third column represent the colors which are used in mapping. The fourth column gives the information about color to bit mapping.

3.2. Embedding phase

Step 1. Apply LZW algorithm on the secret message S.
Step 2. Convert the obtained LZW code into the binary form to generate a bit stream.
Step 3. Count the number of characters (without spaces) of cover text and extract the same number of bits from the bit stream.
Step 4. Now, construct a color coding table (as Table 1) unanimously with the receiver.
Step 5. Color each element of the cover text T as per the extracted bits of the bit stream using the designed table. The colors are used in such a way that every color of a set must be used before using any of them again.
Step 6. The residual bit stream is divided into groups of 12 bits and each group is further partitioned into 9 bits and 3 bits subgroups which are known as G_1 and G_2 respectively. If the bits in the bit stream are not in the multiple of 12 then append the required number of zeros to make it the nearest multiple of 12.
Step 7. Calculate x, y and z as follows:

$$x = (G_1)_{10}/26 \quad (1)$$

$$y = (G_1)_{10}/\text{mod } 26 \quad (2)$$

$$z = (G_2)_{10} \quad (3)$$

Step 8. The values of x and y are converted into letters by employing Latin Square (Fig. 4). Then two letters which are generated above are mapped to one email address by employing K_1 to obtain K_2. Using z email address extension is modified by employing A.
Step 9. Thus, using the T as a cover text and K_2 as the stego-key, the secret message is transmitted in the form of forward mail platform.

The various steps used in embedding phase are shown in Fig. 1

3.3. Extraction phase

Step 1. Get the stego-cover and extract the secret data bit information into C from the colored cover text T using Table 1.
Step 2. Extract first two elements of K_2 and convert them to numbers by employing Latin Square and also extract email address extension to obtain z. Thus, G_1 and G_2 are obtained using following equations

$$G_1 = (x^*26 + y)_2 \quad (4)$$

$$G_2 = (z)_2 \quad (5)$$

Concatenate G_1 and G_2 to obtain G.
Step 3. Concatenate the C and G to obtain the compressed secret data bit stream.
Step 4. Apply LZW to decompress the bit stream to get the original secret data.

The various steps used in extraction phase are shown in Fig. 2

Thus, we get original secret data and the cover text T. Hence, this method is called a reversible text steganographic method.

To illustrate this method, an example is provided. We use a short length of a secret message in the example just to demystifying the method.

An illustrative Example 1

Consider a secret message S "**underlying physiological mechanisms**" and cover text "***Only boats catch connotes of the islands sober wines only ships wrap the slips on the cleats of twining***

Table 2
Values of x, y and z for illustrative example 1.

Number of mail ids required	x	Y	z
1	0	18	4
2	14	9	0
3	5	17	2
4	12	1	0
5	15	11	0
6	11	7	0
7	5	17	6
8	11	11	2
9	10	15	4
10	12	24	0

lines only flags flap in tags with color that assigns only passage on vessels"

Step 1. Apply the LZW compression on the secret message and convert the obtained LZW codes into binary form to get a bit stream. The obtained bit stream is given below:

"0101011001001111010001010100011001010011010011010
101101001001010010 1111010010000101000101001001010
110100101010001001010010100000100110101010000010010
0001001010010001000 1000010010011010100111001000110
0100010001001001010000100100111101001010010101000010
0111001010100"

Step 2. Count the number of characters of cover text and extract the same number of bits from the bit stream. Now, change the color of the cover text according to the extracted bits with the help of Table 1. Thus, we obtain the following text.

"Only boats catch connotes of the islands sober wines
Only ships wrap the slips on the cleats of twining lines
Only flags flap in tags with color that assigns
Only passage on vessels"

Step 3. Partition the residual bit stream into groups of 12 bits each as shown below.

0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, **1**, **0**, **0**, **0**, **1**, **0**, **0**, **0**, **1**, **0**, **0**, **0,** 0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, **1**, **0**, **0**, **1**, **1**, **1**, **0**, **0**, **1**, **0**, **0**, **0,** 1, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, **1**, **0**, **0**, **1**, **0**, **0**, **1**, **0**, **1**, **0**, **0**, **0,** 0, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, **1**, **0**, **0**, **1**, **0**, **1**, **0**, **0**, **1**, **0**, **1**, **0,** 1, 0, 0, 0, 1, 0, 0, 1, 1, 1, 0, 0, **1**, **0**, **1**, **0**, **1**, **0**, **0**, **0, 0, 0, 0, 0**

If the number of bits in the bit stream is not a multiple of 12, then required number of bits as '0' bit are added at the end of the bit stream. In this example, we have added five zeros at the end shown in red color.

Table 3
Comparison of embedding capacity of existing and proposed methods.

Method	Capacity	Explanations
Mimic function [9]	1.27	Evaluated by utilizing the subsequent sample secret message At http://www.spamimc.com
Sun et al. [15]	2.17	Evaluated in UNICODE format by considering sample in Wang et al. [17]
Wang et al. [16]	3.53	Evaluated in UNICODE format by considering sample in Wang et al. [17]
Listega [20]	3.87	Evaluated by utilizing samples in the article referred
Satir and Isik [8]	6.92	Evaluated by utilizing the cover text and sample secret message of Fig. 5 of this article
Rajeev et al. [23]	7.03	Evaluated by utilizing the cover text and sample secret message of Fig. 5 of this article
Rajeev et al. [24]	7.21	Evaluated by utilizing the cover text and sample secret message of Fig. 5 of this article
Proposed method	13.43	Evaluated by utilizing the cover text and sample secret message of Fig. 5 of this article

Step 4. Divide each group into two groups of 9 bits and 3 bits each known as G_1 and G_2 respectively, and compute the value of x, y, and z using Eqs. (1)–(3) respectively as shown in Table 2.

Step 5. Convert the values of x and y to the textual elements using Latin square and also convert the value of z to email extension by employing A. Thus, we get K_2 as follows.

K_2 = [assama@btinternet.com, pktee@gmail.com, htengg@yahoo.com, pebasc@gmail.com, tpteen@gmail.com, qmpreet@gmail.com, lxraj@msn.com, sssonu@yahoo.com, sxkaran@btinternet.com, vhsingh@gmail.com] as shown in Fig. 3.

The hiding capacity for illustrative example 1 is 6.03%.

4. Experimental results

In text steganography, the hiding capacity is a major decisive parameter for performance analysis of the algorithm. In this section, the hiding capacity of the proposed scheme is calculated and compared with recently developed data hiding schemes [8, 9, 15, 16 and 20, 23, 24]. The hiding capacity is calculated by dividing the number of bits of the secret message with the total number of bits used to construct the entire stego-cover. The hiding capacity for illustrative example 1 is 6.03%, which is calculated by using Eq. (6).

$$\text{Capacity} = \frac{\text{bits of secret message}}{\text{bits of stego cover}} \tag{6}$$

The proposed scheme bypasses the mapping operation of [8] which used to incur some overhead in terms of extra information. Thus, the proposed algorithm is improved in terms of computational

---------- Forwarded message ----------
From: **Aruna Malik** <arunacsrke@gmail.com>
Date: Tue, Aug 5, 2014 at 4:09 PM
Subject: Sweet Poem
To: <cpsinghrke1954@gmail.com>
Cc: assama@btinternet.com, pktee@gmail.com, htengg@yahoo.com, pebasc@gmail.com, tpteen@gmail.com, qmpreet@gmail.com, lxraj@msn.com, sssonu@yahoo.com, sxkaran@btinternet.com, vhsingh@gmail.com

"Only boats catch connotes of the islands sober wines
Only ships wrap the slips on the cleats of twining lines
Only flags flap in tags with color that assigns
Only passage on vessels"

With Regards
Aruna Malik, Research scholar, NIT Jalandhar

Fig. 3. The constructed stego cover for the illustrative example 1.

Rows	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
2	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
3	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
4	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
5	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
6	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
7	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
8	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
9	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
10	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
11	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
12	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
13	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
14	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
15	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
16	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
17	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
18	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
19	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
20	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
21	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
22	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
23	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
24	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
25	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
26	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y

Fig. 4. Latin Square.

complexity. Moreover, it is clearly evident from the experimental results that the proposed scheme achieves better hiding capacity. The scheme uses the color coding to embed the secret data additionally in the body of the email message. The proposed scheme is implemented in MATLAB running on the Intel® Core 2 Duo 2.20 GHz CPU, and 3GB RAM hardware platform. The novelty of this work lies into the hiding of secret message on the basis of colors of cover text and also bypassing the mapping phase. A sample secret message (S) of 198 characters and a cover text (T) having 847 characters without quotation marks and with spaces used in our experiments are given below.

The Table 3 shows the comparison of the existing methods with proposed technique in terms of capacity. From the Table 3, it is clearly evident that proposed scheme achieves 13.43% capacity in comparison to [8] which achieves only 6.92% for the same secret message and cover text given in Fig. 5. The proposed scheme is also less computational complex because the mapping operation phase of [8] which has a lot of computation is bypassed and we directly apply the LZW compression algorithm to secret message. Thus, the overhead data resulted in the mapping phase of [8] are also getting diminished in proposed algorithm. Moreover, the proposed method also performs better than that of the Rajeev et al. schemes [23,24] by considering the same cover text and secret data as given in Fig. 5.

For the illustrative purpose, we have taken 10 colors in the color coding table as shown in Table 1. The colors can be increased or decreased as per the choice of the sender and receiver both. In this work, each element of the cover text is colored by a single color. However, we can use multiple colors for coloring a single element. For example, if we want to use multiple colors for a single element, the one way to do this is to use two different colors, one for the border and another for the inner part of the element as per the requirement of the bit stream. Thus, hiding capacity is increased because one bit is hidden in a single element of the cover text. Fig. 6 shows the capacity (in percent) comparison of different methods. Our proposed method performs much better as compared to other existing methods.

4.1. An extension of the proposed method

In this subsection, we discuss an extended method of the proposed method by considering the two different colors for a single character through an illustrative example 2.

Illustrative Example 2: In this example, we will hide secret data into the email ids and cover text of forward mail platform. In email ids, the secret data is embedded in the email extensions and first two characters. In case of cover text, we embed the secret data by coloring them using some specific colors as defined in Table 4 which shows seven colors with their names. Here, we use two different colors for each character of cover text i.e., fill color, boundary

"behind using a cover text is to hide the presence of secret messages the presence of embedded messages in the resulting stego text cannot be easily discovered by anyone except the intended recipient"

(a) Secret message

"in the research area of text steganography, algorithms based on font format have advantages of great capacity, good imperceptibility and wide application range. However, little work on steganalysis for such algorithms has been reported in the literature. based on the fact that the statistic features of font format will be changed after using font-format-based steganographic algorithms, we present a novel support vector machine-based steganalysis algorithm to detect whether hidden information exists or not. this algorithm can not only effectively detect the existence of hidden information, but also estimate the hidden information length according to variations of font attribute value. as shown by experimental results, the detection accuracy of our algorithm reaches as high as 99.3% when the hidden information length is at least 16 bits."

(b) Cover text

Fig. 5. Secret message and cover text.

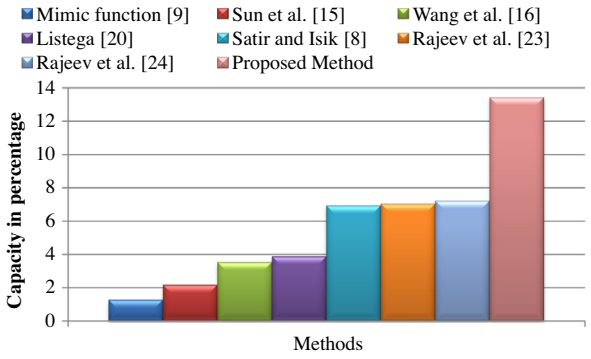

Fig. 6. Capacity (in percent) comparison of different methods.

color to embed secret data. Consider the combination of the seven colors excluding the same color combination and build a color mapping table. We consider the 32 different color combinations for hiding the secret data and represent these 32 color combinations by five bit binary representation as shown in Table 5.

Consider a secret message S "**underlying physiological mechanisms**" and cover text "**you cant believe in god until U believe in yourself**".

Step 1. Apply the LZW compression on the secret message and convert the obtained LZW codes into binary form to get a bit stream. The obtained bit stream is given below:

"010101100100111101000101010001100101001 1 010011010 101101001001010010011110100100001010001010010010101 110100101010001001010010 1 0000010011010101000001001 000010010100100010001000010010011010100111001000110 010001000100100101000010010011111010010100101010001 00111001010100"

Table 4
Seven colors with their names.

S. No.	Color Name	Color
1.	Rackley	
2.	Violet	
3.	Rose	
4.	Green	
5.	Black	
6.	Blue	
7.	Scarlet	

Table 6
Values of x, y and z for illustrative example 2.

Number of mail ids required	x	y	z
1	16	07	04
2	20	17	05
3	10	05	05
4	01	04	05
5	24	07	04

In this work, we will hide secret data into the email ids and cover text of forward mail platform. Thus, we will hide few bits in email ids and others in cover text message.

Step 2. Select few bits of secret data from the bit stream and partition these bits stream into groups of 12 bits each as shown below.

"010101100100, 111101000101, 010001100101, 001101001101, 010110100100"

Step 3. Divide each group into two groups of 9 bits and 3 bits each known as G_1 and G_2 respectively, and compute the value of x, y, and z using Eqs. (1)–(3) respectively as shown in Table 6.

Step 4. Convert the values of x and y to the textual elements using Latin square and also convert the value of z to email extension by employing A. Thus, we get K_2 as follows.
K_2 = [qheet@btinternet.com, vsert@aol.com, mhder@aol.com, ehgdw@aol.com, clawe@btinternet.com]

Step 5. Divide the remaining bit stream into five bit stream sets:

"10100, 10011, 11010, 01000, 01010, 00101, 00100, 10101, 10100, 10101, 00010, 01010, 01010, 00001, 00110, 10101, 00000, 10010, 00010, 01010, 01000, 10001, 00001, 00100, 11010, 10011, 10010, 00110, 01000, 10001, 00100, 10100, 00100, 10011, 11010, 01010, 01010, 10001, 00111, 00101, 01000"

If the number of bits in the bit stream is not a multiple of 5, then required number of bits as '0' bit are added at the end of the bit stream. In this example, we have added one zero at the end shown in red color.

Step 6. Now, change the boundary color for first five bit stream set and fill color for next five bit stream set to embed secret data into the cover text with the help of Table 4. Thus, we obtain the following text.

Table 5
Color mapping for five bit code representation.

S. No.	Boundary line colors		Fill area Color		Binary code	S. No.	Boundary line Color		Fill area Color		Binary code
	Name	Color	Name	Color			Name	Color	Name	Color	
1	Rackley		Violet		00000	17	Green		Blue		10000
2	Rackley		Blue		00001	18	Green		Rackley		10001
3	Rackley		Rose		00010	19	Green		Rose		10010
4	Rackley		Black		00011	20	Green		Black		10011
5	Rackley		Scarlet		00100	21	Black		Violet		10100
6	Violet		Blue		00101	22	Black		Blue		10101
7	Violet		Rackley		00110	23	Black		Rackley		10110
8	Violet		Rose		00111	24	Black		Rose		10111
9	Violet		Black		01000	25	Black		Scarlet		11000
10	Violet		Scarlet		01001	26	Blue		Violet		11001
11	Rose		Violet		01010	27	Blue		Rackley		11010
12	Rose		Blue		01011	28	Blue		Rose		11011
13	Rose		Rackley		01100	29	Blue		Black		11100
14	Rose		Black		01101	30	Blue		Scarlet		11101
15	Rose		Scarlet		01110	31	Scarlet		Violet		11110
16	Green		Violet		01111	32	Scarlet		Blue		11111

Fig. 7. The constructed stego cover for the illustrative example 2.

Step 7. The final forward mail platform for transmitting the secret data is as follows (see Fig. 7).

Thus, the embedding capacity of the proposed method in illustrative example 2 based on the email ids, different colors for each character and the forward mail platform is =((33 * 8)/(303 * 8)) * 100 = 10.89%.

5. Conclusion

In this paper, we have proposed a new technique for text steganography that uses LZW compression and color coding approach for hiding the secret data in the forward mail platform. There are several advantages of the proposed method. Firstly, it performs better in term of computational complexity as there is no need to use mapping operation, this saves the time required to process mapping operation. Secondly, the employment of LZW directly on the secret data diminished the chances of overhead generation thereby increasing the embedding capacity. To further increase the capacity, colors are used in the cover text or message of the email to hide some part of the secret data bit stream. Security of the proposed method has been increased by employing stego keys. The proposed method can be applied to any language by reproducing the text directory and modify the Latin Square to the respective language. Hence, it is not language specific. The proposed method has further advantages that it preserves the novelty of the cover media while transmitting and therefore it changes neither definition nor appearance of the cover text so the text is relevant, linguistic and grammatically accurate and authorized. Compared to other techniques the experimental results of proposed data hiding scheme present the embedding capacity increased up to 13.43%. Therefore, this substantial performance improvement demonstrates the effectiveness of the proposed algorithm.

References

[1] N. Johnson, Z. Duric, S. Jajodia, Information hiding, and watermarking- attacks countermeasures, in: Advances in Information Security, 2001.
[2] F.K. Mohamed, A parallel block-based encryption schema for digital images using reversible cellular automata, Int. J. Eng. Sci. Technol. 17 (2014) 85–94.
[3] C. Karri, U. Jena, Fast vector quantization using a Bat algorithm for image compression, Int. J. Eng. Sci. Technol. (2015).
[4] W. Bender, D. Gruhl, N. Morimoto, A. Lu, Techniques for data hiding, IBM Syst. J. 35 (1984) 313–336.
[5] Z.H. Wang, H.R. Yang, T.F. Cheng, C.C. Chang, A high-performance reversible data-hiding scheme for LZW codes, J. Syst. Softw. 86 (2013) 2771–2778.
[6] T.A. Welch, A technique for high-performance data compression, IEEE Comput. 17 (1984) 6–17.
[7] D. Salomon, Data Compression, Springer-Verlag, 2002.
[8] E. Satir, H. Isik, A compression-based text steganography method, J. Syst. Softw. 85 (2012) 2385–2394.
[9] P. Wayner, Mimic functions, Cryptologia 16 (1992) 193–214.
[10] P. Wayner, Disappearing Cryptography, 2nd ed., Morgan Kaufmann, Menlo Park, 2002, pp. 81–128.
[11] K. Maher, Texto, 1995.
[12] K. Winstein, Lexical steganography through adaptive modulation of the word choice hash, secondary education at the Illinois Mathematics and Science Academy http://alumni.imsa.edu/keithw/tlex/lsteg.ps1999.
[13] B. Murphy, C. Vogel, The syntax of concealment reliable methods for plain text information hiding, in: Proceedings of the SPIE International Conference on Security, Steganography and Watermarking of Multimedia Contents, vol. 6505, 2007.
[14] H. Nakagawa, K. Sampei, T. Matsumoto, S. Kawaguchi, K. Makino, I. Murase, Text information hiding with preserved meaning—a case for Japanese documents, IPSJ Trans. 42 (2001) 2339–2350.
[15] X.M. Sun, G. Luo, G. Huang, Component-based digital watermarking of Chinese texts, in: Proceedings of the 3rd International Conference on Information Security, 2004, pp. 76–81.
[16] Z. Wang, C. Chang, C. Lin, M. Li, A reversible information hiding scheme using left-right and up-down Chinese character representation, J. Syst. Softw. 82 (2009) 1362–1369.
[17] Z.H. Wang, T.D. Kieu, C.C. Chang, M.C. Li, Emoticon-based text steganography in chat, in: Proceedings of Asia-Pacific Conference on Computational Intelligence and Industrial Applications, vol. 2, 2009, pp. 457–460.
[18] R. Stutsman, M. Atallah, C. Grothoff, K. Grothoff, Lost in just the translation, in: Proceedings of the ACM Symposium on Applied Computing, 2006, pp. 23–27.
[19] N. Samphaiboon, Steganography via running short text messages, Multimed Tool Appl. 52 (2011) 569–596.
[20] A. Desoky, Listega: list-based steganography methodology, Int. J. Inf. Secur. 8 (2009) 247–261.
[21] L.Y. Por, K. Wong, K.O. Chee, UniSpaCh a text-based data hiding method using Unicode space characters, J. Syst. Softw. 85 (2012) 1075–1082.
[22] R. Kumar, S. Chand, S. Singh, An efficient text steganography scheme using Unicode Space Characters, Int. J. Forensic Comput. Sci. 10 (2015) 8–14.
[23] R. Kumar, S. Chand, S. Singh, An Email based high capacity text steganography scheme using combinatorial compression, in: 5th IEEE International Conference CONFLUENCE 2014: The Next Generation Information Technology Summit, 25th-26th Sept., 2014, pp. 336–339.
[24] R. Kumar, S. Chand, S. Singh, A high capacity Email based text steganography scheme using Huffman compression, in: International Conference on Signal Processing & Integrated Networks, 2016.
[25] K. Tutuncu, A.A. Hassan, New approach in E-mail based text steganography, Int. J. Intell. Syst. Appl. Eng. 3 (2015) 54–56.
[26] A.A. Mohamed, An improved algorithm for information hiding based on features of Arabic text: a Unicode approach, Egypt. Inf. J. 15 (2014) 79–87.