ISSN (e): 2250 – 3005 || Volume, 08 || Issue, 6|| Jun – 2018 ||
International Journal of Computational Engineering Research (IJCER)

Improved Text Steganography Scheme Based On LZW Compression and Color Coding

Amanpreet Kaur[1], Sukhvir Kaur[2], Gunjan Sethi[3]
[1]Research Scholar, Department of Computer Science Engineering, CTIEMT, Jalandhar
[2]Assistant Professor, Department of Computer Science Engineering, CTIEMT, Jalandhar
[3]Assistant Professor, Department of Computer Science Engineering, CTIEMT, Jalandhar
Correspondence Author: Amanpreet Kaur

ABSTRACT
Steganography is a technology where new data compression, information theory and cryptography technologies are brought together to propitiate the need for privacy on the Internet. The art of information hiding has earned much attention in the recent few years as security of information has become a big issue in this internet world. As sharing of sensitive information via a common communication channel has become indispensable, Steganography is the art and science of hiding sensitive information into cover media. This paper presents a new mechanism of hiding the secret message using color mapping and symbol mapping tables using LZW compression. Multiple parameters like capacity and processing time have been evaluated. Various experiments have been conducted and the results are analyzed in this paper.
KEYWORD: Steganography, cryptography, plain text, encryption, decryption, cipher, LZW, compression, color coding.

Date of Submission: 21-05-2018 Date of acceptance: 05-06-2018

I INTRODUCTION

Steganography can be defined as a method of hiding secret data within a cover media so that other individuals fail to realize the existence of the secret data. In other words, steganography is the science of hiding information. It is often confused with cryptography as both are used to protect confidential information. The difference between the steganography and cryptography is in the appearance of the processed output; the output of steganography operation is not obviously visible but in cryptography the output is twisted so that it can draw attention. If a nefarious government or Internet service provider (ISP) is looking for encrypted messages, they can easily find them. Whereas the aim of cryptography is to make data unreadable by a third party, the aim of steganography is to hide the data from a third party. Steganalysis is the process to detect the presence of steganography. The key terminologies in the context of steganography are: plaintext is the original secret message that needs to be communicated; cover text is the larger and harmless looking data which is used as container for the plaintext; and the stego text is the data generated after embedding the plaintext into the cover text. The basic features expected from a steganography method are high embedding capacity, invisibility or perceptual transparency, undetectability, robustness (i.e. the ability of the algorithm to retain the data embedded in the cover), tamper resistance (the capability to prevent modification or deletion or embedding a different message), and the independence of the original cover . Off course, some of these requirements conflict and thus, any specific algorithm can satisfy only one or two of them. More specifically, embedding capacity, robustness, and undetectability are mutually conflicting and cannot all be achieved by one algorithm. Limitation of cryptography is that the third party is always aware of the communication because of the unintelligible nature of the text. Steganography overcomes this limitation by hiding message in an innocent looking object called cover object. Cryptography is the practice of protecting the contents of a message alone, steganography is concerned with concealing the fact that a secret message is being sent, as well as concealing the contents of the message. Steganography includes the hiding of information within computer files. In digital steganography, it is the science that involves communicating secret data in an appropriate multimedia carrier, e.g., image, audio, and video files. Media files are best for steganographic transmission because of their large size. The Steganography Method Used should have:

a) Imperceptibility: The video with data and original data source should be perceptually identical.

b) Robustness: The embedded data should pull through any processing operation the host signal goes through and preserve its safeness.
c) Capacity: Maximum data embedding rate.
d) Secrecy: Extraction of hidden information from the cover media must not happen without prior permission of intended user having password.
e) Accuracy: The extraction of the hidden data from the medium should be accurate and reliable.

Steganography is the art of hiding information in seemingly harmless carriers without drawing suspicion to the transmission of a hidden message. On the other hand, the art of discovering and rendering such harmless covert messages of hidden information is known as steganalysis. The primary goal of steganalysis is the identification of the existence of a hidden message, and then the identification of hot-spots to look for hidden information. Below are some of the common term which is necessary to understand any steganography system:

- Cover Media- It is the medium in which secret information is embedded in such a way that it is difficult to detect the presence of data.
- Stego- Media- It is medium obtained after embedding the secret information.
- Secret data- The data or information to be hidden in cover media.
- Steganalysis- Steganalysis is the process of detecting presence of secret information in cover media.

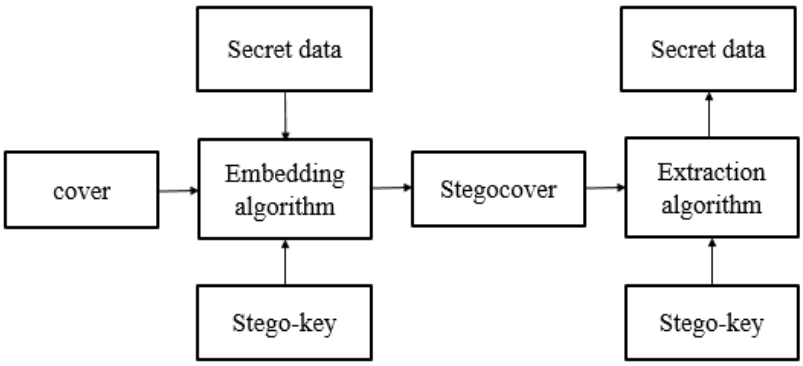

Figure1.General model of steganography [17]

II RELATED WORK

Moerland proposed a text steganography technique by using specific characters from the words. In this method, some specific characters from certain words are selected and are used to hide the secret information. For e.g. the first character of first word of each paragraph can be used to hide a secret message one character at a time such that by placing these characters side by side, we get the whole message [2]. Moerland also discussed about the text steganography approach by using punctuation marks. The idea behind this approach is to utilize the presence of punctuation marks like comma (,), semi colon (:), quotes („ “) etc. in the text for encoding a secret message. The use of punctuation marks is quite common in the normal English text and hence it becomes difficult for the intruder to recognize the presence of secret message in the text document. This accounts for the security of the technique [2]. Low, Maxemchuk, Brassil, Gorman [3] and Alattar [4] proposed a text steganography technique by using line shifting method. In this method, the lines of the text are shifted to some degrees say 1/300 inch up or down and then the information is hidden by creating a hidden unique shape of the text. Low, Maxemchuk, Brassil, Gorman [3] and Kim, Moon, Oh [5] proposed a text steganography technique using word shifting method. In this method, the information is hidden by shifting the words horizontally or by changing the distance between the words. Niimi, Minewaki, Noda, Kawaguchi proposed a technique that uses synonyms of certain words to hide the message in the English text. In this method, certain words from the text are selected, their synonyms are identified and then the words along with their synonyms are used to hide the secret message in the text [6]. Huang, Yan proposed a technique for hiding information by adding extra white-spaces in the text. These white spaces can be placed at the end of each line, at the end of each paragraph or between the words [7]. Shirali-Shahreza [8, 9] and Memon, Khowaja, Kazi [10] proposed a steganography method on Arabic, Persian and Urdu text. One of the characteristics of these languages is the abundance of points in its letters. One point letters can be used to hide the information by shifting the position of a point a little bit vertically high with respect to the standard point position in the text. Alla, Prasad proposed a Hindi text steganography technique. This technique is based on the fact that each language has its own characteristics. Every language is formed of combinations of one

or more vowels and consonants [11]. Shirali-Shahreza proposed a text steganography technique that hides secret message in the English text by using different spellings of the words. In English some words have different spelling in UK and US. For example "dialog" has different terms in UK (dialogue) and US (dialog) [12].

III METHODOLOGY

The proposed methodology uses the color mapping table and the symbol table for steganography. The color mapping table includes the boundary color and the filling color (shown as Table 1). The symbol table contains the various symbols that will be used for embedding (shown as Table 2). The LZW algorithm has been used for compression of the secret message. It will further increase the capacity of the proposed algorithm. The LZW algorithm is directly applied on the secret text and the obtained bit stream is encoded using color mapping table and the symbol table. A color coding table is used to hide the some portion of the secret data bits into the cover text, thus the notion of the content is not modified. Remaining bits of the secret message are stored using the symbol table. The method discussed here increases the hiding capacity and also reduces computational complexity. Figure 2 shows the proposed methodology used for embedding the secret message into the cover text.

3.1 Embedding Phase

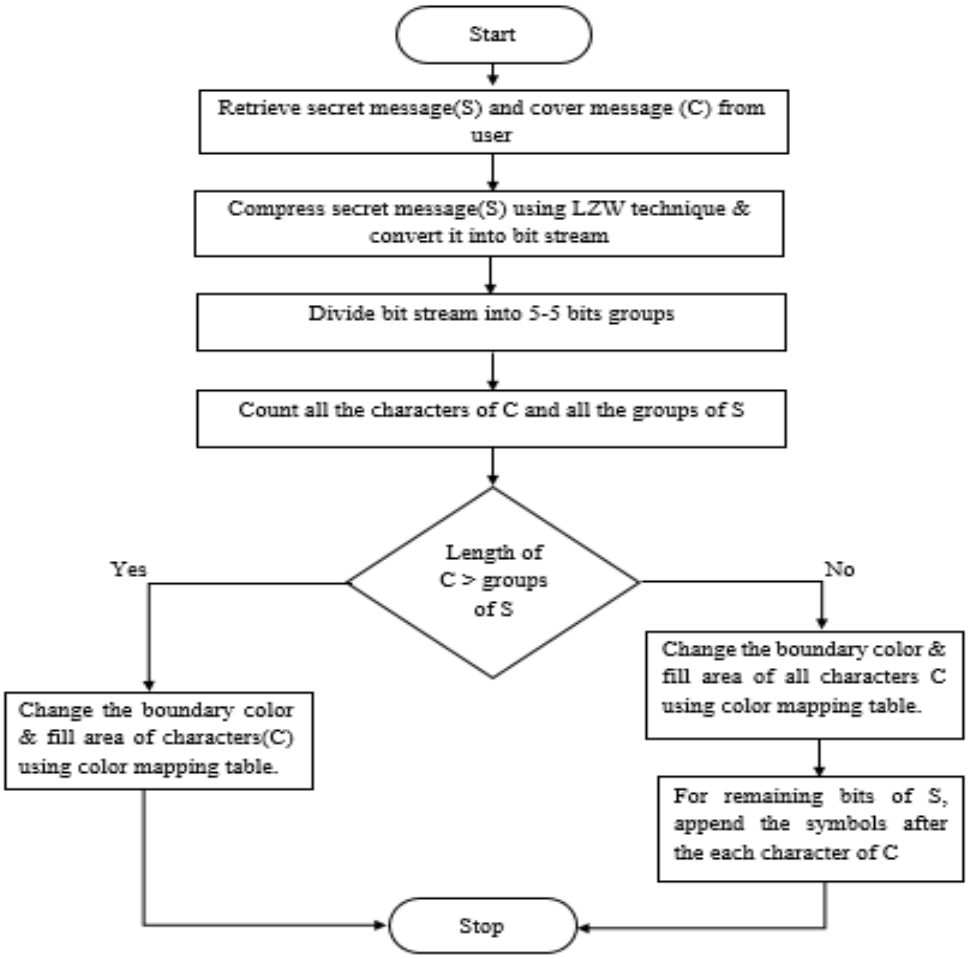

Figure 2. Proposed methodology for embedding phase

The embedding phase of proposed method is mentioned below:
Step 1. Retrieve the secret message (S) from the user.
Step 2. Retrieve the cover text message (C) from the user.
Step 3. Compress secret message(S) using LZW technique & convert it into bit stream
Step 4. Divide bit stream into 5 bits groups. If the bits in the bit stream are not in the multiple of 5 then append the required number of zeros to make it the nearest multiple of 5.
Step 5. Construct the color mapping table (as table 1) using different combinations of boundary color and filling color.
Step 6. Construct the symbol table (as table 2) using different symbols available on the keyboard.

Step 7. Count the number of characters in the cover message and number of groups of the compressed secret message.
Step 8. If size of cover text is larger than the size of secret message, then change the boundary color and filling color of the cover text characters using color mapping table.
Step 9. But if size of cover text is smaller than the size of secret message, then first change the boundary color and filling color of the cover text characters using color mapping table till the length of cover text and for the remaining bits of secret message, use the symbol table and append the symbol between two characters of the cover text.

3.2 Extraction Phase
Step 1. Get the stego-cover and extract the secret data bit information into S1 from the colored cover message by using Table 1
Step 2. Extract all symbols from stego-cover and extract the secret data bits into S2 from symbols by using Table 2.
Step 2. Combine both S1 and S2 to obtain the compressed secret data bit stream.
Step 3. Apply the LZW to decompress the secret message to obtain the secret message in original format.

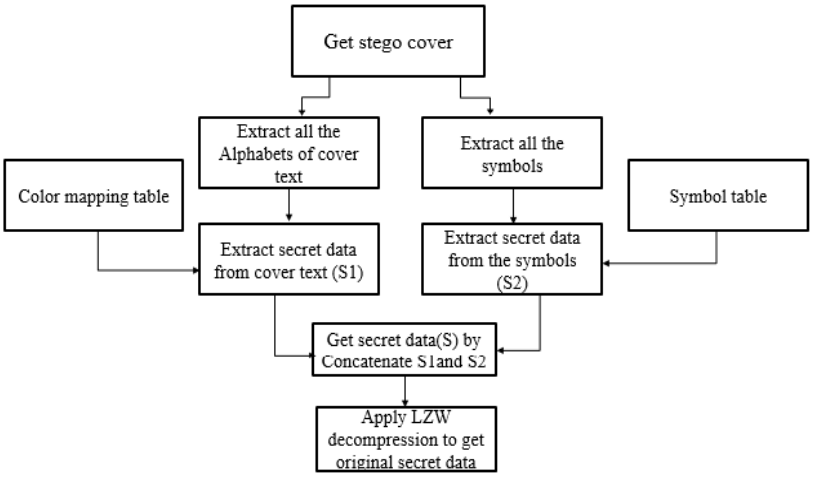

Figure 3. Methodology for extraction phase.

Example 1.
Let secret message S is "**underlying physiological mechanism**" and cover message C is "**you cant believe in god until u believe in yourself**"
Step 1. Apply LZW algorithm on the secret message S.
Step 2. Convert the obtained LZW code into binary format as shown below:
11111111110111111111111101111111111111111111111111111111111111111111111000101000010110000000000000000010101000111000110111101001010010110101111010011000000111100010110001001001001000110000101011010000010010101100100110100011110111110011101011111

Step 3. Partition of the compressed secret message into groups of 5 bits as shown below:
11111,11111,01111,11111,11110,11111,11111,11111,11111,11111,11111,11111,11111,11111,10001,01000,01011,00000,00000,00000,00101,01000,11110,00110,11110,10010,10010,11010,11110,10011,00000,01111,00010,11000,10010,01001,00011,00001,01011,01000,00100,10101,10010,01101,00011,11011,11100,11101,01111.

If the bits in the bit stream are not in the multiple of 5 then append the required number of zeros to make it the nearest multiple of 5.
Step 4. Count the number of bits in the cover message and the compressed secret message and compare cover message with secret message.
Step 5. Size of cover text is smaller than the size of secret message, then first change the boundary color and filling area color of the cover message characters using color mapping table(table 1) till the length of cover message .

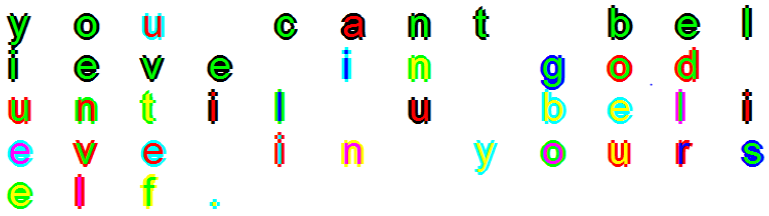

Figure 4.Stego cover with color mapping

Table 1. Color mapping table

S. NO.	BOUND - ARY COLOR NAME	COLOR	FILL AREA COLOR NAME	COLOR	BINARY CODE	S. NO.	BOUNDAR Y COLOR NAME	COLOR	FILL AREA COLOR NAME	COLOR	BINARY CODE
1	RED		GREEN		00000	17	CYAN		GREEN		10000
2	RED		BLUE		00001	18	CYAN		BLUE		10001
3	RED		CYAN		00010	19	CYAN		YELLOW		10010
4	RED		YELLOW		00011	20	CYAN		MAGENTA		10011
5	RED		MAGENTA		00100	21	YELLOW		RED		10100
6	GREEN		RED		00101	22	YELLOW		GREEN		10101
7	GREEN		BLUE		00110	23	YELLOW		BLUE		10110
8	GREEN		CYAN		00111	24	YELLOW		CYAN		10111
9	GREEN		YELLOW		01000	25	YELLOW		MAGENTA		11000
10	GREEN		MAGENTA		01001	26	MAGENTA		RED		11001
11	BLUE		RED		01010	27	MAGENTA		GREEN		11010
12	BLUE		GREEN		01011	28	MAGENTA		BLUE		11011
13	BLUE		CYAN		01100	29	MAGENTA		CYAN		11100
14	BLUE		YELLOW		01101	30	MAGENTA		YELLOW		11101
15	BLUE		MAGENTA		01110	31	BLACK		RED		11110
16	CYAN		RED		01111	32	BLACK		GREEN		11111

Step 6. For the remaining bits of secret message, use the symbol table (Table 2) and append the symbol between two characters of the cover text as shown in figure 5.

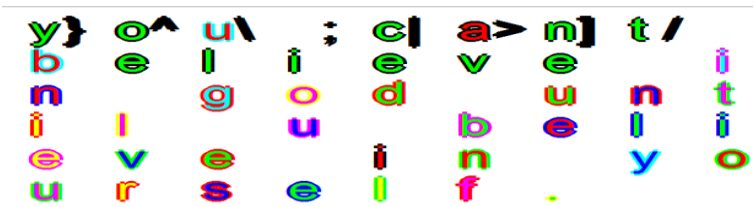

Figure 5. Stego cover with Symbol

For the security purposes and for the convenience of the user, the size of the symbols have been reduced so that they are not visible to the human eye as shown in figure 6.

Figure 6. Stego cover with small size of symbols

Table 2. Symbol table

S. No	SYMBOL	BIT REPRESENTATION	S. No	SYMBOL	BIT REPRESENTATION
1	~ (Tilde)	00000	17	/	10000
2	!	00001	18	\	10001
3	@	00010	19	+	10010
4	#	00011	20	- (hyphen)	10011
5	$	00100	21	_ underscore	10100
6	%	00101	22	=	10101
7	^	00110	23	[	10110
8	&	00111	24	]	10111
9	*	01000	25	{	11000
10	(	01001	26	}	11001
11	)	01010	27	‘ single quote	11010
12	, comma	01011	28	“ double quote	11011
13	. Dot	01100	29	: colon	11100
14	<	01101	30	; semicolon	11101
15	>	01110	31	| piping	11110
16	?	01111	32	÷	11111

IV EXPERIMENTAL ANALYSIS

In text steganography, capacity is the important parameter for performance analysis of the proposed method and existing method. Multiple number of experiments have been conducted using different lengths of cover text and secret message. The experiments have been conducted on system having Intel i7 processor with 16 GB of RAM. As we have studied that some bits of the secret message are encoded using color mapping Table we use two different color for single character to embed the five bits of secret data and remaining bits of secret data are stored using symbol table we use one symbol to embed five bits of secret data. The experiments are carried out in Matlab environment and the results of the experiments for existing method [18] are mentioned below in table 3 and results of experiments for proposed method are mentioned in table 4. Capacity is calculated by dividing number of bits of secret message with number of bits of cover message [18]. Capacity is calculated for different length of cover message and secret message by using following formula.

Capacity = bits of secret message/bits of cover message

Table3. Results of existing methodology

S. No.	Length Of Secret Message In Characters	Length of Compressed Secret Message in characters	Length Of Cover Text In Characters	Length Of Email Ids In Characters	Total Length(Cover Text + Email Ids)	Message Encoded (Yes/No)	Message Decoded (Yes/No)	Capacity (%)	Processing Time (in seconds)
1	35	35	181	85	266	YES	YES	13.15	0.799
2	35	35	109	85	194	YES	YES	18.04	0.814
3	35	35	57	85	142	YES	YES	24.64	0.778
4	51	20	57	52	109	YES	YES	18.34	0.745
5	51	20	35	52	87	YES	YES	22.98	0.756
6	51	20	25	52	77	NO	N.A	N.A	N.A
7	51	20	28	52	80	NO	N.A	N.A	N.A
8	185	137	181	433	614	NO	N.A	N.A	N.A
9	198	135	847	427	1274	YES	YES	10.59	0.738
10	198	135	227	427	654	YES	YES	20.64	0.782
11	198	135	212	427	639	NO	N.A	N.A	N.A
12	208	143	227	443	670	YES	YES	21.34	0.792
13	209	144	227	fail	N.A	N.A	N.A	N.A	N.A
14	85	71	181	215	396	YES	YES	17.92	0.805
15	86	72	181	215	396	YES	YES	18.18	0.796

Table 4. Results of proposed methodology

S. No.	Length Of Secret Message In Characters	Length Of Compressed Secret Message In Characters	Length Of Cover Text In Characters	Message Encoded (Yes/No)	Message Decoded (Yes/No)	Symbols Used	Capacity (%)	Processing Time (in seconds)
1	35	35	181	YES	YES	NO	19.33702	0.259
2	35	35	109	YES	YES	NO	32.11009	0.274
3	35	35	57	YES	YES	YES	61.40351	0.245
4	51	20	57	YES	YES	YES	35.08772	0.263
5	51	20	35	YES	YES	YES	57.14286	0.271
6	51	20	25	YES	YES	YES	80	0.266
7	51	20	28	YES	YES	YES	71.42857	0.268
8	185	137	181	YES	YES	YES	75.69061	0.278
9	198	135	847	YES	YES	NO	15.93861	0.274
10	198	135	227	YES	YES	YES	59.47137	0.269
11	198	135	212	YES	YES	YES	63.67925	0.270
12	208	143	227	YES	YES	YES	62.99559	0.259
13	209	144	227	YES	YES	YES	63.43612	0.265
14	85	71	181	YES	YES	NO	39.22652	0.281
15	86	72	181	YES	YES	NO	39.77901	0.279

We have carried out 15 experiments and results obtained are mentioned in table 3 and table 4. In table 3, only 11 experiments have been conducted successfully. In table 4, all the experiments have been conducted successfully and no error has been reported. All the experiments were encoded and decoded successfully. The capacity of all the experiments have been mentioned over here. The capacity ranges from 15% - 80%. Capacity of proposed method has been increased. Figure 7 depicts the comparison of capacity of the existing work and the proposed work.

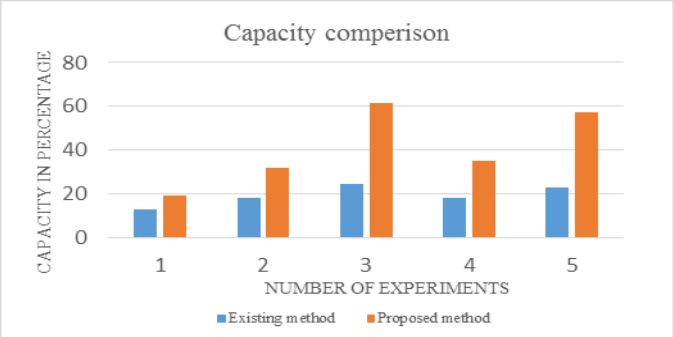

Figure 7. Capacity comparison

Processing time has been evaluated for the existing work and the proposed work. The processing time is evaluated in seconds. The processing time is evaluated for all the experiments. The results are mentioned in the table 3 and table 4.

From the table 3 and table 4, it is evident that the proposed work is taking much lesser processing time in comparison with the existing work. Figure 8 depicts the comparison of processing time of the existing work and the proposed work.

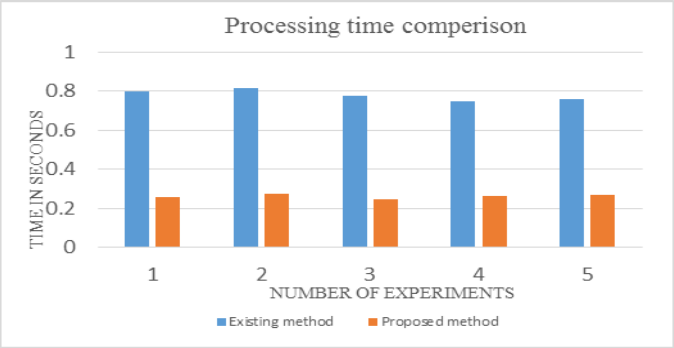

Figure 8. Processing time comparison

V CONCLUSION

A new improved text Steganography scheme based on LZW compression and color coding is proposed in this research work. Color mapping table and symbol tables have been used in this research work. Multiple parameters like processing time, capacity have been computed. The proposed method provides the better results than the existing schemes in terms of capacity and embedding performance. It has been checked that in our proposed techniques the capacity has been increased and processing time has been substantially reduced. In the future work, we can combine this research work with the rapidly growing cloud computing framework where user data needs a high end security.

REFERENCES

[1]. 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.

[2]. T. Moerland, "Steganography and Steganalysis", www.liacs.nl/home/tmoerlan/privtech.pdf, May 15, 2003.

[3]. S.H. Low, N.F. Maxemchuk, J.T. Brassil, and L. O'Gorman, "Document marking and identification using both line and word shifting", Proceedings of the Fourteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '95), vol.2, 2-6 April 1995, pp. 853 - 860.

[4]. 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.

[5]. 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.

[6]. 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.

[7]. D. Huang, and H. Yan, "Inter word Distance Changes Represented by Sine Waves for Watermarking Text Images", IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, no. 12, December 2001, pp. 1237-1245.

[8]. M. H. Shirali-Shahreza, and S. Shirali-Shahreza, "A New Approach to Persian/Arabic Text Steganography", Proceedings of 5th IEEE/ACIS international Conference on Computer and Information Science and 1st IEEE/ACIS, June 2006.

[9]. M. H. Shirali-Shahreza, and S. Shirali-Shahreza, "A Robust Page Segmentation Method for Persian/Arabic Document", WSEAS Transactions on Computers, vol. 4, Issue 11, Nov. 2005, pp. 1692-1698.

[10]. J.A. Memon, K. Khowaja, and H. Kazi, "Evaluation of steganography for Urdu /Arabic text", Journal of Theoretical and Applied Information Technology, pp 232-237.
[11]. K. Alla, and R.S.R Prasad, "An Evolution of Hindi Text Steganography", Sixth International Conference on Information Technology New Generations, 2009 (ITNG'09), Digital Object Identifier: 10.1109/ITNG.2009.41, 2009, Page(s): 1577 - 1578.
[12]. M. Shirali-Shahreza, "Text Steganography by Changing Words Spelling", International Journal of Advanced Communication Technology, 2008 (ICACT 08), Volume: 3, Digital Object Identifier: 10.1109/ICACT. 2008.4494159, 2008, Page(s): 1912 - 1913.
[13]. Z.H. Wang, C.C. Chang, D. Kieu, and M.C. Li, "Emoticon-based Text Steganography in Chat", Second Asia Pacific Conference on Computational Intelligence and Industrial applications, 2009.
[14]. C.-H. Yang and M.-H. Tsai, (2010) "Improving Histogram-based Reversible Data Hiding by Interleaving Predictions", IET Image Processing, Vol.4. Iss. 4 pp. 223-234.
[15]. Venkata Abhiram.M, Sasidhar Imadabathuni, U.Padmalochini, Maheedhar Imadabathuni and Ramya Ramnath (2009), "Pixel Intensity Based Steganography with Improved Randomness", International Journal of Computer Science and Information Technology, Vol 2, No 2, pp.169-173.
[16]. G.Sahoo & Rajesh Kumar Tiwari (2009) "Hiding Secret Information in Movie Clip: A Steganographic Approach", International Journal of Computing and Applications, Vol. 4, No.1, pp 103-110. [17] David Salomon, "Data Hiding in Text", Data Privacy and Security. (Springer), pp. 245-267, 2003.
[17]. David Salomon," Data Hiding in Text", Data Privacy and Security. (Springer), pp. 245-267, 2003.
[18]. A. Malik, G. Sikka and H. K. Verma, "A high capacity text steganography scheme based on LZW compression," Engineering Science and Technology, An International Journal, pp. 72-79, 2017.

Amanpreet Kaur." Improved Text Steganography Scheme Based On LZW Compression and Color Coding." International Journal of Computational Engineering Research (IJCER), vol. 08, no. 06, 2018, pp. 26-34.