International Journal of Computer (IJC)

ISSN 2307-4523 (Print & Online)

© Global Society of Scientific Research and Researchers

http://ijcjournal.org/

A Study on Text Based Steganography Using Email Platform and Color Mapping

Amanpreet Kaur^{a*}, Sukhvir Kaur^b, Gunjan Sethi^c

^aResearch Scholar, Department of Computer Science Engineering, CT Group of institution, Shahpur, Jalandhar, Punjab, India 144020

^{b,c}Assistant Professor, Department of Computer Science Engineering, CT Group of institution, Shahpur,

Jalandhar, Punjab, India 144020

^aEmail: amanpreet.kamboj231@gmail.com

^bEmail: sukhsain.17@gmail.com
^cEmail: gunjan.ctit@gmail.com

Abstract

Steganography is the art of hiding of a message within another so that presence of hidden message is indistinguishable. The key concept behind steganography is that message to be transmitted is not detectable to the casual eye. This is also the advantage of steganography over cryptography. Modern digital steganography uses text, images, audio, video etc. as a cover medium. This paper presents a mechanism of storing the secret message using forward email platform and color mapping table. Various experiments have been conducted and the results are analyzed in this paper.

Keywords: Steganography; cryptography; plain text; encryption; decryption; cipher.

1. Introduction

Almost all computer users write and exchange documents written with applications using proprietary document formats such as Microsoft Office. Even though causes have not been addressed yet, it is well known among communities of expert computer users that unexpected information may hide into those documents and can be easily revealed. On the other hand, regular computer users are often unaware of the information leakage by their published documents. Classical steganography concerns itself with ways of embedding a secret message (which might be a copyright mark, or a covert communication, or a serial number) in a cover message (such as a video film, an audio recording, or computer code). The embedding is typically parameterized by a key; without knowledge of this key (or a related one) it is difficult for a third party to detect or remove the embedded material. Once the cover object has material embedded in it, it is called a stego object. Thus, for example, we might embed a mark in a cover text giving a stego text; or embed a text in a cover image giving a stego-image; and so on [14].

⁻⁻⁻⁻⁻

^{*} Corresponding author.

There has been a rapid growth of interest in this subject over the last two years, and for two main reasons. Firstly, the publishing and broadcasting industries have become interested in techniques for hiding encrypted copyright marks and serial numbers in digital films, audio recordings, books and multimedia products; an appreciation of new market opportunities created by digital distribution is coupled with a fear that digital works could be too easy to copy. Steganography can be defined as the art and science of hiding data so that no one knows about hidden data except the sender and receiver. It is used to hide data inside data. Steganography is not a new concept, it is in use since centuries and this technique has started as computer came into existence as different algorithms are used to hide data with different cover media [1][2]. The general specifications that must be kept in mind when implementing steganography are described below:

- The embedded data must not downgrade the quality of cover media. Quality of cover media should be such that it must look like original media. The media size should not increase tremendously as it can look suspicious to the casual viewer. It must have good Peak Signal to Noise Ratio.
- The data that is to be hidden should be embedded in the information part of cover media, not in the header. The data must be temper resistance to the attacks of the third party.
- Integrity of data should be maintained, it must be robust enough so that it must not be modified in between the way.
- A cover file must be of enough size so that it can hide a large amount of message.
- The message hidden must be invisible to the viewer.
- The hidden message must be undetectable during steganalysis process.

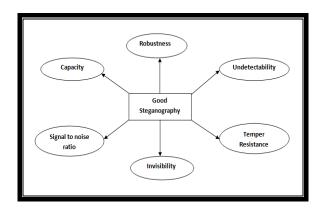


Figure 1: Features of a good Steganography [4]

1.1 Types of Steganography

With the emerging use of the Internet, it was important to secure the information using information technology. Cryptography was developed as a technique to hide the meaning of message. Various methods have been developed to encrypt and decrypt the message in order to hide its meaning. Modern steganography came into existence with the advent of personal computers in 1985 and classical steganographic problems began to solve electronically.

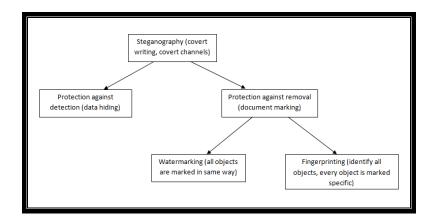


Figure 2: Types of Steganography [18].

The various techniques used in steganography are as follows:

1.1.1 Text Steganography

Text steganography [2,7] can be achieved by changing the text format, or by altering certain features of textual elements such as alphabets. The main goal is to develop such techniques that perform changes such that that original data must be reliably decodable yet largely invisible to the reader.

1.1.2 Image Steganography

To a computer, an image is a collection of numbers that constitute different light intensities in different areas of the image. The individual points are referred to as pixels.

1.1.3 Audio Steganography

Information hiding in audio is based on the interpretation of sound made by Human Auditory System.

1.1.4 Video Steganography

A video is a combination of audio and image. A continuous flow of image constitutes a video. Therefore, the techniques that can be applied on audio and image separately, they can be applied on video also.

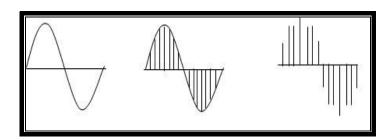


Figure 3: An analog signal converted into digital signal by sampling and quantization [19]

Some of the techniques used for audio steganography are- LSB encoding, Phase encoding, MP3 etc.

2. Related Work

Dickman [17] has worked upon the technique used to hide data known as steganography. The steganography technique uses another layer of security when used in combination with cryptography. The carriers that can be used for this purpose can be image, audio, and video. It also discuss some methods which are used to find whether steganography is implemented or not due to its misuse. In [18] paper the performance of some of the steganography tools is analyzed. Cryptography and Steganography techniques are compared. As both techniques are used to secure data, therefore these techniques can be used in combination. Steganography is a useful tool that allows covert transmission of information over the communications channel. Hidden image is generated by combining secret image with the carrier image. The hidden image is invisible to third party and difficult to detect without retrieval. Its uses, applications and its history are also given a brief description. In [19] paper author gives a brief description of steganography which is to hide existence of data. It discusses how TCP/IP header can be used in steganography. Steganalysis is a process to detect the presence of steganography. Different techniques with their limitations have been discussed .The vulnerability of different media toward different attacks have been considered. Every steganography technique is less or more vulnerable to attacks and therefore it lays emphasis on strong steganographic technique. Kaur et al. [20] have worked upon the performance evaluation of various techniques which are used in steganography. There are situations when different applications have different requirements of the steganography technique used. For example, applications may require absolute invisibility of the secret information and requires the message size must be short as compared to cover size, while other can require a larger secret message to be hidden. This paper gives an overview of image steganography, its uses and techniques. It also identifies the requirements of a good steganographic algorithm and briefly reflects on which steganographic techniques are more suitable for which applications. The techniques which are discussed are covered under image domain works on least significant bits. Thilagamani et al. [21] has conducted the survey on different clustering techniques to achieve image segmentation. Different clustering techniques are discussed which can be applied on images and databases. Clustering can be termed here as a grouping of similar images in the database. The process of clustering is done based on different attributes of an image such as size, color, texture etc. The purpose of clustering is to get meaningful result, effective use of storage and fast retrieval in various areas. Kavitha et al. [22] has published the information for the rapid development of data transfer through internet made it easier to send the data accurate and faster to the destination. One of the most important use of information technology and communication is to provide the security of the information. This paper gives a brief idea about the image steganography that makes use of Least Significant Bit to implement steganography which is used in combination with cryptography. Cryptography is applied before implementing steganography. Mritha et al. [23] has proposed the steganography approach for sequential data encoding and decoding in video images. This paper is aimed to transmit maximum hidden data without losing the video quality and size. In order to achieve this goal, in this system we are using the encryption key for sequential data encoding and decoding. The system is implemented in MATLAB environment. The performance of the steganography approach has been evaluated using video images in bit mapped (bmp) format in Red, Green, and Blue (RGB) components. The experimental results prove that the sequential encoding based steganography system is simple and produces imperceptible distortions in resulting bmp images.

3. Methodology

In this work, capacity and security issues have been taken into account. 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 method discussed here increases the hiding capacity and also reduces computational complexity.

3.1 Embedding phase

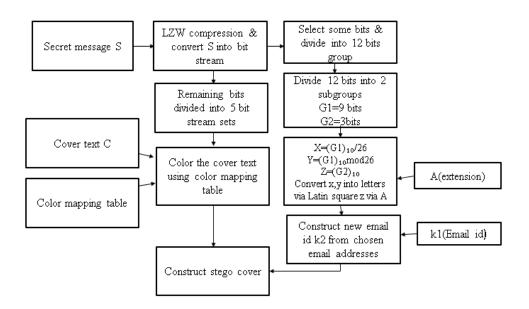


Figure 4: Methodology for embedding phase [16]

Figure 4 shows the methodology used for embedding the secret message into the cover text.

The embedding phase of proposed method is divided into two steps:

- Embed some bits of secret text into email ids.
- Embed remaining bits using color coding table.

Let S: secret message

T: Cover texts or message of email

K1: A set of email addresses shared between the sender and the receiver which plays the role of public stego-key.

Embedding phase of secret text using email ids:

Step 1. Apply LZW algorithm on the secret message S.

Step 2. Convert the obtained LZW code into binary format as shown in figure 5.

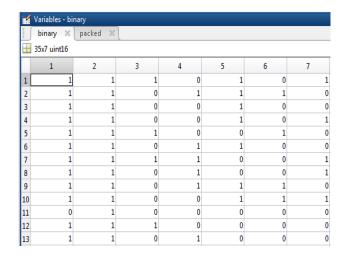


Figure 5: Binary format of compressed secret message

Step 3. Fetch some bits from the bit stream of secret message and use it for embedding using email ids. Remaining bits of secret message will be used for color coding phase. In existing methodology, secret data is hidden into two parts:

- Using email ids
- Using colored cover message.

Step 4. Binary data of compressed secret message is divided into 2 parts. First Half bits of binary data will be used by Latin square table for encryption of email-id. Second half bits will be used by the color cover message

Step 5. Select few bits of secret data from the bit stream and partition these bits stream into groups of 12 bits each. Divide each group into two groups of 9 bits and 3 bits each known as G1 and G2 respectively, and compute the value of x, y, and z

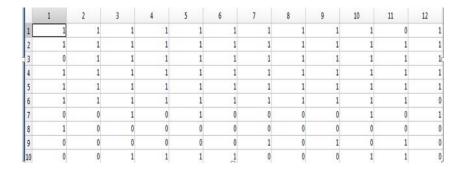


Figure 6: 12 bit Groups

Step 6. Compute x, y and z by using Latin square table. 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.

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

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

 $z=(G_2)_{10}$

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	В	C	D	E	F	G	н	I	J	K	L	M	N	0	P	Q	R	S	Т	U	v	w	x	Y	Z
2	В	C	D	E	F	G	H	I	J	K	L	M	N	0	P	Q	R	S	T	U	v	w	X	Y	Z	A
3	C	D	E	F	G	H	1	J	K	L	M	N	0	P	Q	R	S	T	U	\mathbf{v}	w	X	Y	Z	A	В
4	D	E	F	G	H	I	J	K	L	M	N	0	P	Q	R	S	T	U	v	w	X	Y	Z	A	В	C
5	E	F	G	H	I	J	K	L	M	N	0	P	Q	R	S	T	U	v	w	X	Y	Z	A	В	C	D
6	F	G	H	I	J	K	L	\mathbf{M}	N	O	P	Q	R	S	T	U	v	w	\mathbf{x}	Y	Z	A	В	C	D	E
7	G	H	I	J	K	L	M	N	0	P	Q	R	S	T	U	v	w	X	Y	Z	A	В	C	D	E	F
8	H	1	J	K	L	M	N	0	P	Q	R	S	T	U	v	w	\mathbf{x}	Y	Z	A	В	C	D	E	F	G
9	I	J	K	L	M	N	0	P	Q	R	S	T	U	v	w	X	Y	Z	A	В	C	D	E	F	G	H
10	J	K	L	M	N	0	P	Q	R	S	T	U	v	w	X	Y	Z	A	В	C	D	E	F	G	H	I
11	K	L	M	N	0	P	Q	R	S	T	U	v	w	X	Y	Z	A	В	C	D	E	F	G	H	I	J
12	L	M	N	0	P	Q	R	S	T	U	v	w	X	Y	Z	A	В	C	D	E	F	G	H	I	J	K
13	M	N	0	P	Q	R	S	T	U	\mathbf{v}	w	X	Y	Z	A	В	C	D	E	F	G	H	1	J	K	L
14	N	0	P	Q	R	S	T	U	v	w	X	Y	\mathbf{z}	A	В	C	D	E	F	G	H	1	J	K	L	M
15	O	P	Q	R	S	T	U	v	w	X	Y	Z	A	В	C	D	E	F	G	H	I	J	K	L	M	N
16	P	Q	R	S	T	U	v	w	\mathbf{x}	Y	Z	A	В	C	D	E	F	G	H	I	J	K	L	M	N	0
17	Q	R	S	T	U	v	w	X	Y	Z	A	В	C	D	E	F	G	H	I	J	K	L	M	N	0	P
18	R	S	T	U	v	w	X	Y	Z	A	В	C	D	E	F	G	H	I	J	K	L	M	N	0	P	Q
19	S	T	U	\mathbf{v}	w	X	Y	Z	A	В	C	D	E	F	G	H	I	J	K	L	M	N	0	P	Q	R
20	T	U	v	w	X	Y	Z	A	В	C	D	E	F	G	H	1	J	K	L	M	N	0	P	Q	R	S
21	U	\mathbf{v}	w	\mathbf{x}	Y	Z	A	В	C	D	E	F	G	H	1	J	K	L	M	N	0	P	Q	R	S	T
22	v	w	X	Y	Z	A	В	C	D	E	F	G	H	1	J	K	L	M	N	0	P	Q	R	S	T	U
23	w	X	Y	Z	A	В	C	D	E	F	G	H	I	J	K	L	M	N	0	P	Q	R	S	T	U	v
24	X	Y	Z	A	В	C	D	E	F	G	H	I	J	K	L	M	N	0	P	Q	R	S	T	U	v	w
25	Y	Z	A	В	C	D	E	F	G	H	I	J	K	L	M	N	0	P	Q	R	S	T	U	\mathbf{v}	w	x
26	Z	A	В	C	D	E	F	G	Н	I	J	K	L	M	N	O	P	Q	R	S	T	U	v	w	X	Y

Figure 7: Latin Square table

Sample of K2 is specified as follows:

```
'ursama@aol.com'
'vstee@verizon.net'
'mxengg@verizon.net'
'xubasc@verizon.net'
'yvteen@verizon.net'
'zwpreet@msn.com'
'jiraj@aol.com'
'rdsonu@gmail.com'
'inkaran@yahoo.com'
```

Figure 8: Encrypted email ids

Step 7. Divide the remaining bit stream into five bit stream sets. 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.

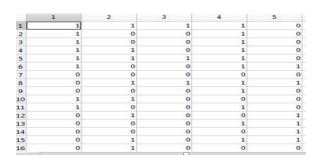


Figure 9: Representation of remaining 5 bits.

Step 8. 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 color mapping table.

 Table 1: Color mapping table

S. NO	BOUN DARY COLO R NAME	CO LO R	FILL AREA COLO R NAME	COL OR	BIN AR Y CO DE	S. NO.	BOUNDAR Y COLOR NAME	FILL AREA COLOR NAME	COLO R	BINA RY CODE
1	RED		GREEN		0000	17	CYAN	GREEN		10000
2	RED		BLUE		0000	18	CYAN	BLUE		10001
3	RED		CYAN		0001	19	CYAN	YELLOW		10010
4	RED		YELLO W		0001	20	CYAN	MAGENT A		10011
5	RED		MAGE NTA		0010 0	21	YELLOW	RED		10100
6	GREEN		RED		0010 1	22	YELLOW	GREEN		10101
7	GREEN		BLUE		0011 0	23	YELLOW	BLUE		10110
8	GREEN		CYAN		0011 1	24	YELLOW	CYAN		10111
9	GREEN		YELLO W		0100 0	25	YELLOW	MAGENT A		11000
10	GREEN		MAGE NTA		0100 1	26	MAGENTA	RED		11001
11	BLUE		RED		0101 0	27	MAGENTA	GREEN		11010
12	BLUE		GREEN		0101 1	28	MAGENTA	BLUE		11011
13	BLUE		CYAN		0110 0	29	MAGENTA	CYAN		11100
14	BLUE		YELLO W		0110 1	30	MAGENTA	YELLOW		11101
15	BLUE		MAGE NTA		0111 0	31	BLACK	RED		11110
16	CYAN		RED		0111 1	32	BLACK	GREEN		11111

3.2 Extraction Phase

Step 1. Get the stego-cover and extract the secret data bit information from the color mapping table. Fetch the

bits using the boundary color and filling color of the alphabet and stored inside A.

Step 2. Extract first two elements of K2 and convert them to numbers by employing Latin Square and also extract email address extension to obtain z. Thus, G1 and G2 are obtained

$$G1 = (x * 26 + y)_2$$

$$G2 = (Z)_2$$

Step 3. Concatenate G1 and G2 to obtain G.

Step 4. Append A and G together to obtain the secret message in compressed format.

Step 5. Apply the LZW to decompress the secret message to obtain the secret message in original format.

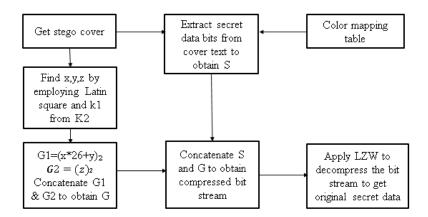


Figure 10: Methodology for extraction phase [16].

4. Experimental Analysis

Multiple number of experiments have been conducted using different lengths of cover text and secret message. As we have studied that some bits of the message are stored inside email ids and remaining bits are stored using color mapping table. So we have used 2 approaches over here. For the experiments mentioned in table 2, the secret message is exactly divided into 2 parts which means 50% message is stored using email ids and remaining 50% is stored using color mapping table.

The experiments are carried out in Matlab environment and the results of the experiments are mentioned below in table 2 and table3. The secret message is compressed using LZW compression to increase the overall capacity. Capacity [16] is calculated using the given formula.

Capacity = bits of secret message/bits of cover message (1)

Table 2: Calculation of capacity (1:2)

S.NO.	Length Of Secret Message In Characters	Length of Compressed Secret Message in characters	Length Of Cover Text In Characters	Length Of Email Id	Total Length(Cover Text + Email)	Message Encoded (Yes/No)	Message Decoded (Yes/No)	Capacity
1	35	35	181	161	342	YES	YES	10.23
2	35	35	109	161	270	YES	YES	12.96
3	35	35	57	161	218	YES	YES	16.05
4	35	35	25	161	186	NO	N.A	N.A
5	198	135	227	FAILED	N.A	NA	N.A	N.A
6	208	143	227	FAILED	N.A	NA	N.A	N.A

In table 2, we have divided the secret message into exactly 2 parts and results are mentioned. Out of 6 experiments, only 3 experiments have been conducted successfully. But for the experiments in table 3, 25% of the secret message is stored using email ids and remaining 75% is stored using color mapping table. Out of 6 experiments, 1 experiments failed to store the message whereas 5 experiments have completed successfully.

Table 3: Calculation of capacity (1: 4)

s.no.	Length Of Secret Message In Characters	Length of Compressed Secret Message in characters	Length Of Cover Text In Characters	Length Of Email Id	Total Length(Cover Text + Email)	Message Encoded (Yes/No)	Message Decoded (Yes/No)	Capacity
1	35	35	181	85	266	YES	YES	13.15
2	35	35	109	85	194	YES	YES	18.04
3	35	35	57	85	142	YES	YES	24.64
4	35	35	25	85	110	NO	N.A	N.A
5	198	135	227	427	654	YES	YES	20.64
6	208	143	227	443	670	YES	YES	21.34

Table 2 and table 3 illustrates the results of the conducted experiments. Table 3 is providing better results than

table 2. It means when message is divided by 4 and is split into 25:75 ratio, it produces better results. Out of the total 6 experiments, only 5 experiments are conducted successfully and are able to store the secret message inside the cover text. Some experiments failed while encoding the email ids and other failed while color coding mapping.

5. Conclusion

In this research paper we have stored the secret message into the cove text using a hybrid version of email ids and color mapping technique. There are too many flaws in this technique. As it is evident from the results that this mechanism failed frequently and we need a cover text of appropriate size to store the secret message. We need some robust mechanism to store the secret message that should not fail too frequently. In the next work our main aim is to increase the capacity of this mechanism and decrease the failure ratio which will improve the overall performance and reliability of the mechanism.

References

- [1] H.Kabetta, B.Y. Dwiandiyanta, Suyoto, "Information hiding in CSS: A secure scheme text-steganography using public key Cryptosystem", IJCIS, Vol.1, No.1, pp. 13-22, December 2011.
- [2] M Garg," A Novel Text Steganography Technique Based on Html Documents", International Journal of Advanced Science and Technology, Vol. 35, pp.129-138, October, 2011.
- [3] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding", IBM Systems Journal, Vol. 35, No. 3&4, pp. 313-336, 1996.
- [4] P.Singh, R.chaudhary and A.Agarwal," A Novel Approach of Text Steganography based on null spaces", IOSR Journal of Computer Engineering, Vol. 3, No. 4, pp. 11-17, July-Aug. 2012.
- [5] M.S. Shahreza," A New Method for Steganography in HTML Files", Advances in Computer, Information, and Systems Sciences, and Engineering, pp. 247-251, 2007, Springer.
- [6] R. Kumar, A. Malik, "A Space based reversible high capacity text steganography scheme using Font type and style", International Conference on Computing, Communication and Automation, IEEE, 2016, pp.1090-1094.
- [7] R. Saniei, K. Faez, "The Capacity of Arithmetic Compression Based Text Steganography Method", Iranian Conference on Machine Vision and Image Processing, pp.38-42, 2013.
- [8] H. Singh, P.K. Singh, K. Saroha, "A Survey on Text Based Steganography", Proceedings of the 3rd National Conference; INDIACom-2009, 2009, pp. 26-27.
- [9] J.A. Memon, K. Khowaja, and H. Kazi, "Evaluation of steganography for Urdu /Arabic text", Journal of Theoretical and Applied Information Technology, Vol. 4,No. 3, pp. 232-237, 2008.
- [10] K. Bennett, "Linguistic Steganography: Survey, Analysis, and Robustness Concerns for Hiding Information in Text", Purdue University, CERIASTech Report 2004-13.

- [11] 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, April 1995, pp. 853 860.
- [12] 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, pp. 685-695, June 2004.
- [13] 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, pp. 1237-1245, December 2001.
- [14] Y.H. Yu, C.C. Chang, I.C. Lin," A new steganographic method for color and grayscale image hiding", Computer Vision and Image Understanding, Vol. 107, No. 3, pp. 183-194, 2007.
- [15] M. H. Shirali-Shahreza, and S. Shirali-Shahreza, "A Robust Page Segmentation Method for Persian/Arabic Document", WSEAS Transactions on Computers, Vol. 4, No. 11, pp. 1692-1698, Nov. 2005.
- [16] A. Malik, G. Sikka and H. K. Verma, "A high capacity text steganography scheme based on LZW compression and color coding," Engineering Science and Technology, An International Journal, Vol.20, No. 1, pp. 72-79, 2017.
- [17] S.D. Dickman," An overview of steganography", Department of Computer Science, James Madison University Infosec Techreport, 2007.
- [18] A. Kumar, K. Pooja," Steganography- A Data Hiding Technique", International Journal of Computer Applications (0975 8887), Vol. 9, No.7, pp. 19-23, 2010.
- [19] S Das, S Das, B Bandyopadhyay, S Sanyal," Steganography and Steganalysis: different approaches", arXiv preprint arXiv:1111.3758, 2011.
- [20] J. Kaur and S. Kumar," Study and Analysis of Various Image Steganography Techniques", International Journal of Computer Science and Technology (IJCST), Vol.2, No.3, pp.535-539, 2011.
- [21] S. Thilagamani, N. Shanthi," A Survey on Image Segmentation Through Clustering", International Journal of Research and Reviews in Information Sciences, Vol. 1, No. 1, pp.14-17, March 2011.
- [22] K.K Kavitha, A Koshti, P Dunghav," Steganography Using Least Signicant Bit Algorithm", International Journal of Engineering Research and Applications (IJERA), Vol. 2, Issue 3, pp. 338-341, May-Jun 2012.
- [23] M Ramalingam, N.A.M Isa," A steganography approach for sequential data encoding and decoding in video images", International Conference on Computer, Control, Informatics and Its Applications, pp. 120-125, 2014.