ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

A Secure Data Hiding and Extraction Technique from Digital Carrier Media

Aparna V. Navlakha[1], Dr. Avinash D.Gawande[2]

M.E. Student, Dept. of CSE, Sipna College of Engineering &Technology, Amravati, India[1]

Head, Dept. of CSE, Sipna College of Engineering &Technology, Amravati, India [2]

ABSTRACT: A new uncompressed Audio secure data hiding algorithm is proposed. This algorithm hides the secret messages (hidden text) within every day, seemingly innocuous objects (cover text) which is audio to produce a stego audio. The recipient of a stego audio can use his knowledge of the particular method of data hiding employed to recover the hidden secret message from the stego audio. The goal of information hiding is to permit parties to converse in such a way that an attacker cannot tell whether or not there's hidden meaning to their conversation. In the algorithm, embedding and detection operations are both executed entirely within the uncompressed domain, with no need for the decompression process. The new criteria using statistical invisibility of contiguous frames is used to adjust the embedding strategy and capability, which increases the security of proposed method. Therefore, the collusion resistant properties are obtained. Experimental results showed this method can be applied on decompressed Audio data hiding with high security properties.

KEYWORDS: Higher LSB, Data Hiding, Extraction, MSE, PSNR, WAV Audio.

I. INTRODUCTION

In present situation, Data security is the major part of computer world and it's a rapidly growing area in IT sector. Data hiding is the transmission of a secret message hidden within an ordinary carrier while not revealing its existence. The cover file (container) may be digital still image, audio file, or video. If we embed the secret message, then it can be transmitted across insecure lines or we can post in public places. For this reason, digital Audio is convenient choice for data hiding. Nowadays, in modern information systems such as multimedia sensor networks, covert communications becomes a greater threat to forensic analysis than ever. Thus it is necessary to find out methods to discover and discourage covert communications such as data hiding in multimedia networks that acquire highly correlated data.

This method will focus on the particular problem of the compressed Audio data hiding. General speaking, digital Audio appears in two main distinct encoding formats: uncompressed and compressed. The most famous compressed format by far is motion compensated compressed Audio, specifically the widely accepted standard MPEGx. It accomplishes compression through the elimination of spatial, temporal and statistical redundancies and with this compression operation. The Audio bit-stream consist of variable length codes (VLC) that represent various Audio segments. For Audio stream usually being offered in compressed form, data hiding algorithms that are not applicable in compressed bit-stream will require complete or at least partial decompression. But this is an unnecessary burden which can be avoided. If the requirement of strict compressed domain data hiding is to be met, the data hiding needs to be embedded in the compressed domain. Method begins with replacing the higher bit of audio wave file called as carrier file. Sample of audio wave file are taken & 5th layer bit is replaced with the message bit, but care is taken for not having too much quantization error.

Recently, there are large amount of Audio watermarking algorithms been proposed where some of them are applied for compressed Audio. To be useful, a data hiding technique should not be easily detectable. If we can detect the presence of secret message with higher probability than random guessing, the corresponding data hiding technique is considered to be invalid. Similar to cryptography, data hiding may suffer from the attack method (steganalysis). Much of the research work for the field of steganalysis has been carried out on images.. One approach is based solely on the blind steganalysis concept, which is produced by blind classifiers. The classifier should be trained to learn the differences between the features of cover and stego-image at first. Another approach is based on the first order statistics and is associated only with idempotent embedding.

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

There are also two Audio steganalysis methods using collusion principle. And because of Audio statistical invisibility properties, inspired us to design this data hiding. In this paper, we propose a secure uncompressed Audio data hiding architecture taking account of Audio statistical invisibility. Also the architecture is with a steganalysis module, operated in a closed-loop manner to improve anti- steganalysis capability of stego Audio with data embedded.

II. LITERATURE SURVEY AND RELATED WORK

Andreas Westfeld and Gritta Wolf [1], during this work authors have described a steganographic system which embeds secret messages into a Video stream. Normally the compression methods are used in Video for securing acceptable quality. But usually, compression techniques are lossy because reconstructed image may not be equal with the original. There are some drawback of compression and data embedding technique. Signal noise and irrelevance are common examples of data embedding. But compression methods try to remove noise of signal and irrelevance. If signal is compressed more, then there are fewer possibilities of data embedding. The authors have solved this problem, they have investigated a typical signal path for data embedding. In this method, security is established by indeterminism within the signal path.

Arup Kumar Bhaumik, Minkyu Choi, RosslinJ.Robles, and MaricelO.Balitanas[6], the main requirements of any data hiding system are security, capacity and robustness. It is very difficult to achieve all these factors together because these are inversely proportional to each other. Authors give focus on increasing security and capacity factor of data hiding. The data hiding method uses high resolution digital Audio as a cover signal. It provides the ability to hide a significant quality of information making it different from typical data hiding methods. They have used the large payloads like Audio in Audio and picture in Audio as a cover image.

Ahmed Ch. Shakir [7], the confidential communications over public networks can be done using digital media like text, images and audio on the internet. Simply hiding the contents of a message using cryptography was not adequate. Hiding of message provide an additional layer of security. To provide the more security the author suggested the new procedures in data hiding for hiding ciphered Information inside a digital colour bitmap image. He has used quadratic technique which depends on the locations concluded by the binary image, beside of public key cryptography. He had concluded that the conjunction between cryptography and data hiding produce immune information.

S.Suma Christal Mary [9], has proposed Real time Compressed Video Secure Steganography (CVSS) method using Video bit stream. In this, embedding and detection operations are both executed completely in the compressed. The proposed algorithm increases the safety because the statistical invisibility of contiguous frames helps to adjust the embedding strategy and capacity. At present we are hiding the data in audio format, so in the future implementation of uncompressed formats may possible as well. Multiple frames embedding are possible. Also we are embedding single frame at a time, but in future multiple frames embedding is also possible.

Saurabh Singh and GauravAgarwal [10], have presented a novel approach of hiding image in a Video. In this approach, one LSB of every pixel is replaced with one bit of secrete message. So it is very difficult to find that image is hidden in a Video of 30 frames per second. The analysis is very difficult because each row of image pixels are hidden in multiple frames of the Video. The intruder needs full video to unhide image. Authors have described the LSB algorithm in their paper. The proposed algorithm is very useful in sending sensitive information securely.

Sherly A. P. and Amritha P. P. [12], in their paper have proposed a new compressed video Steganographic scheme. In this scheme, data is hide in compressed domain. The novel embedding technique Triway Pixel Value Differencing (TPVD) is used to enhance the capacity of hidden secret information and for to providing an imperceptible stego-image for human vision. This method can be applied on compressed videos without degradation in visual quality.

Abdullah Bamatraf, Mohd. NajibMohd. Salleh and Rosziati Ibrahim [13] in their work describes A New Digital Watermarking method using various combination of Least Significant Bit (LSB) and Inverse Bit. Author proposed a new LSB based digital watermarking method with the combination of LSB and inverse bit. The experimental result shows that the proposed algorithm maintains the quality of the watermarked image. when combining different positions of LSB such as the second LSB, third LSB and fourth LSB and the various combination between them. The proposed algorithm is also tested using Peak signal-to noise ratio (PSNR).

Yusuf Perwej, Firoj Parwej, Asif Perwej[14] in their work describes An Adaptive Watermarking Technique for the copyright of digital images and it's protection. Authors proposing edge detection from Gabor Filter method, uses data hiding by the simple LSB substitution method. In the method a set of pixels that constitute a block jointly share the bits from the watermark .The values of mean square error (MSE) and peak signal to noise ratio (PSNR) are measured. The results shows that the method introduces low noise and hence ensures lesser visible distortions.

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

S.S.Verma, R.Gupta, G.Shrivastava[15] in their work presented a high capacity and high stego- signal quality audio steganography scheme based on the samples comparison in DWT domain where selected coefficient of a segment value is change by a threshold value depending on the embedding cipher text bit. The strength of their algorithm is depend on the segment size and their strength are enabled the algorithm to accomplish very high embedding capacity for different data type that can reach up to 25% from the input audio file.

III. PROPOSED METHODOLOGY

A. *Data Hiding in Audio Wave File*

One of the simplest algorithms with extremely high data rate of additional information is hiding the data inside the least significant bits (LSBs) of audio samples in the time domain. A given technique could shift the limit for transparent data hiding in audio in the fifth LSB layer, by using a two-step approach as shown in algorithm. Within the first step, a watermark bit is embedded by using a LSB coding method into the 5^{th} LSB layer of the host audio. In the next step, whatever the impulse noise generated by watermark embedding is shaped in order to alter its white noise properties. The original host audio bit in the 5^{th} layer replaces with the bit from the watermark bit stream by using standard LSB coding method. If the i^{th} LSB layer is used for embedding and the original and watermark bit are distinct then the error due to watermarking is 2^i-1 quantization steps (QS) (amplitude range is [-256 to 255]). If the original bit value is 0 and watermark bit value is 1 then the embedding error is positive and vice versa. The main idea of the proposed LSB algorithm that causes minimum embedding distortion of the host audio is watermark bit embedding. It is clear that, if only one of 8 bits in a sample is fixed and identical to the watermark bit, the other bits can be flipped in order to reduce the embedding error. For example, if the original sample value was $(00100000)_2=32_{10}$, and the watermark bit is one is to be embedded into 5th LSB layer, instead of value $(00110000)_2=48_{10}$, that would the standard algorithm produce, the proposed method produces sample that has value $(00011111)_2=31_{10}$, which is far more closer to the original sample. The extraction algorithm just reads the bit value from the predefined LSB layer of watermarked audio sample and retrieves the watermark bit . In the embedding algorithm, the 5th LSB layer is first modified by insertion of the present message bit and then the algorithm shown below is run. The case in which the bit a_i need not be modified at all, no action is taken with that signal sample. Underlined bits (a_i) stands for bits of watermarked audio.

B. *Procedure*

1. Select Input audio and Check for uncompressed format of an audio.
2. If input audio is uncompressed, extract samples from audio and convert it in binary form.
3. After converting in to the binary form that binary form are in consecutive way of 0 and 1.In sampling process this binary converted to 8 bits of group of samples**.**
4. To protect secret data need to generate key pattern and from that key pattern have to select number of keys for embedding purpose**.**
5. Hide secret data into 5th LSB bit position of selected energy samples. Energy samples are those samples whose decimal value is in between 16 to 234
6. Assembles audio samples so that result stego audio file will generate.
7. Save result stego audio file containing secrete data in an encrypted format.
8. Add stego audio into original audio & thus resultant audio with secrete information is generated. Then calculate MSE and PSNR.
9. At Receiver Side, Select Stego audio and convert it to binary form.
10. In resampling process this binary converted to 8 bits of group of samples**.**
11. Select proper key file for audio samples selection and specify the number of bits to be extracted
12. Extract secrete encrypted data from stego audio file. and decode it

The fig - 1 given below shows the way that how the concept works and different terms used for the process of data hiding are

- Carrier /Cover File :- A original message or a file in which hidden information will be stored in it
- Stego – Medium :- The medium in which information is hidden
- Embedded or Payload :- Information which is to be concealed
- Steganalysis :- The process of detecting hidden information inside a file.

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

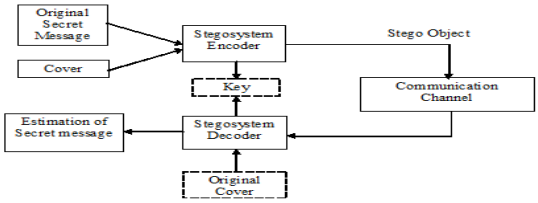


Fig-1 Fundamental procedure of data hiding

In the above diagram the secret message can be any text file or image or any audio wave file and then inputting the cover signal in which data is to be embedded. The cover signal must be sufficient large to cover the message. After selecting the input secret message and cover signal next, we find the length of the audio file as well as length of the text file. Before hiding the secret message into cover signal it must be converted into the other form so that it can't be interpretable by intruder. Before delivery of secret message to receiver , it must be converted back to its original form.

C. *5th Bit LSB Replacement Algorithm*

if 5th LSB!= Data Bit
if bit 0 is to be embedded
if ai-1 = 0 then ai-1, ai-2, ..., a0 = 11.....1
if ai-1 = 1 then ai-1, ai-2, ..., a0= 00.....0 and
If ai+1 = 0 then a i+1 = 1
else if a i+2 = 0 then a i+2= 1
.....
else if a7=0 then a7 = 1
else if bit 1 is to be embedded
if ai-1 = 1 then ai-1, ai-2, ..., a0 = 00.....0
if ai-1 = 0 then ai-1, ai-2, ..., a0= 11.....1 and
If ai+1 = 1 then a i+1 = 0
else if a i+2 = 1 then a i+2 = 0
.....
else if a7=1 then a7 = 0

.

This algorithm hides the secret messages, seemingly innocuous objects (cover text) which is audio to produce a stego audio. The block diagram of message hiding algorithm at sender is shown in fig-2.

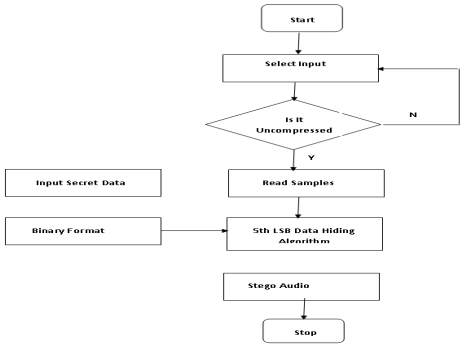


Fig-2 Flow chart of message hiding at sender side

Then recipient of a stego audio can use his knowledge of particular method of data hiding employed to recover original secret message from stego audio. The block diagram of message recovery algorithm at receiver side is shown in fig- 3.

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

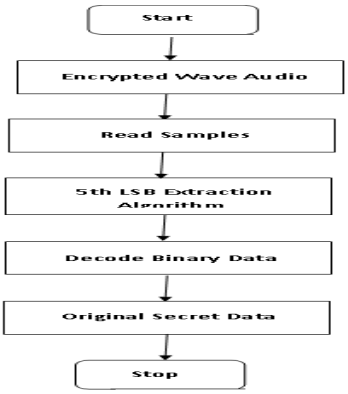


Fig-3 Flow chart of message recovery at receiver side

IV. EXPERIMENTAL RESULTS

The proposed data hiding algorithm using audio is implemented with MATLAB. We select five different audio files of different sizes and hide secret data in that files and measure various parameters like maximum data hiding capacity of each audio file which is calculated using total energy samples, time required for audio to binary conversion, sampling, key generation, embedding and extracting secret data, etc. We also measure mean square error (MSE) and peak signal to noise ratio (PSNR) of each audio file. Our results also show that the size of audio file and stego file which contains secret data are identical that means there is no change in file size before and after embedding secret data are same.

The table - 1 represents the various attributes of audio files like name, size, it's sample rate, total number of samples and duration in seconds.

Sr. No.	i/p Audio name (.wav)	Size of i/p audio (kb)	Sample Rate	Total Samples	Duration in sec
1	applause	18.1	8000	18480	2.31
2	babycry	18.1	11025	18476	1.67
3	goodbye	27.8	11025	28504	0.65
4	hello	15.6	11025	15976	0.36
5	new1094	175	22050	179660	2.037

Table - 1: Audio File Information

The table - 2 represents the various attributes of samples in selected audio files like total number of samples which are selected for embedding purpose in selected audio, total energy samples in selected samples , data hiding capacity in selected samples, etc . It also shows the total number of characters we can hide in that particular selected samples and the time required for audio to binary conversion and sampling process

Sr. No.	i/p Audio file (.wav)	Samples selected to-from	Total samples selected	Total energy samples	Energy samples in selected samples	Data hiding capacity (%)	Total char hide	Time for audio to binary (sec)	Time for sampling (sec)
1	applause	4-3000	2996	18480	2996	100	374	12.88	3.29
2	babycry	2-1000	998	18470	995	99	124	5.38	2.32
3	goodbye	6-3000	2994	2201	2201	73	275	13.09	3.82
4	hello	0-2000	2000	9380	956	47	119	7.52	2.62
5	new1094	8-500	492	68244	206	41	25	3.28	1.78

Table - 2:Selected Samples Information

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

The table - 3 gives the data hiding information like total number of secret character wish to hide in selected file, length of key file, time for embedding and merging and comparisons between samples before and after merging

Sr. No.	i/p Audio file (.wav)	Total Char	Length of key file (bits)	Time needed to generate keys (sec)	Time to Embed (sec)	Time needed for merging (sec)	Total samples before merging	Total samples after merging
1	applause	29	232	0.15	2.05	0.63	18480	18480
2	babycry	12	96	0.09	1.05	0.54	18476	18476
3	goodbye	33	264	0.16	1.92	0.67	28504	28504
4	hello	27	216	0.12	1.89	0.58	15976	15976
5	new1094	5	40	0.06	1.76	3.34	179660	179660

Table - 3: Data Hiding Information

The table - 4 calculates mean square error and peak signal noise ratio and time needed to calculate MSE and PSNR. MSE is a measure used to quantify the difference between the initial and the distorted or noisy audio. PSNR is the ratio between the maximum power of a signal and the power of corrupting noise that affects the fidelity of its representation

Sr. No.	i/p Audio file(.wav)	MSE	PSNR	Time needed for Calculation (sec)
1	applause	0.0113	43.5	0.03
2	babycry	0.0077	45.17	0.03
3	goodbye	0.0165	41.86	0.04
4	hello	0.0137	42.66	0.03
5	new1094	0.0012	53.24	0.26

Table - 4:Calculation of MSE and PSNR

The table - 5 represents samples selection in stego audio files at receiver side. It also time required for audio to binary conversion and resampling process

Sr. No.	Stego Audio file (.wav)	Samples selected to-from	Total samples selected	Time for audio to binary(sec)	Time for resampling (sec)
1	applause1	4-3000	2996	12.88	3.29
2	babycry1	2-1000	998	5.38	2.32
3	goodbye1	6-3000	2994	13.09	3.82
4	hello1	0-2000	2000	7.52	2.62
5	new10941	8-500	492	3.28	1.78

Table - 5 : Samples Selected at Receiver Side

The table - 6 gives the data retrieving information like total number of secret characters to extract in selected stego file, length of key file, time required for extracting and decoding.

Sr. No	Stego Audio file (.wav)	Total Char	Length of key file (bits)	Time for Extraction (sec)	Time needed to decode (sec)
1	applause1	29	232	0.034	0.021
2	babycry1	12	96	0.026	0.019
3	goodbye1	33	264	0.031	0.019
4	hello1	27	216	0.029	0.018
5	new10941	5	40	0.027	0.017

Table - 6 : Secret Data Extraction Information at Receiver Side

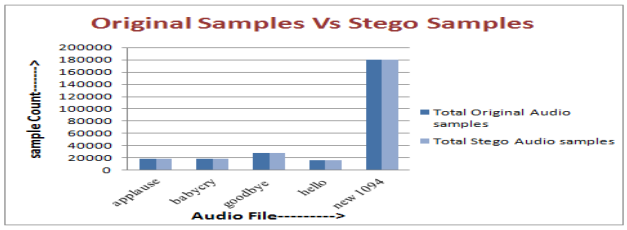
The Graph - 1 represents the comparison between original samples and stego samples which contains secret data. It is found that there is no change in length before and after embedding secret data.

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

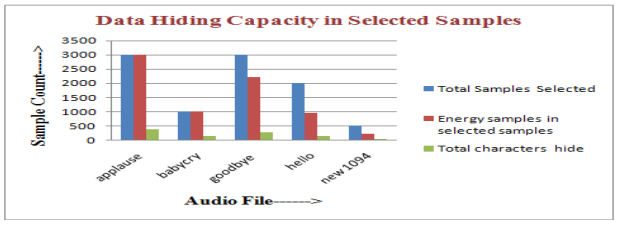
(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016



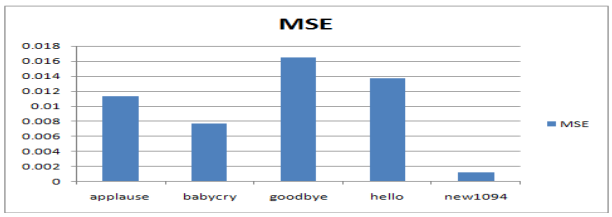
Graph - 1: Comparison between original Samples to Stego Samples

The Graph - 2 represents total number of samples which are selected for embedding purpose in selected audio, total energy samples in selected samples and number of characters of secret data sent to receiver side.



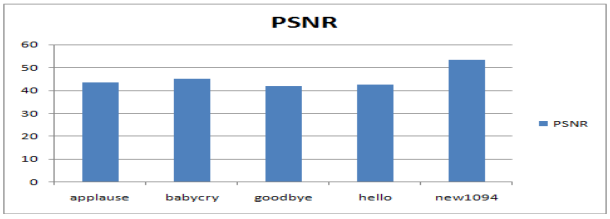
Graph - 2: Maximum Data Hiding Capacity in Audio

The Graph - 3 calculates MSE of different audio file selected for embedding purpose. A lower MSE value indicates that the reconstruction is of higher quality. If MSE is zero, then the original and stego files are same.



Graph - 3: MSE Calculation of different audio file

The Graph - 4 calculates PSNR of different audio file selected for embedding purpose. A higher PSNR value indicates that the reconstruction is of higher quality.



Graph - 4: PSNR Calculation of different audio file

V. CONCLUSION

An application that require high-volume embedding with robustness against certain statistical attacks, the proposed method has been used. We proposed a data hiding technique in an uncompressed audio. The main focus of it is to develop a system with extra security features where secret message can be hidden by data hiding technique. It means

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

this system produce a decent, economical technique for concealing the information during transmission. It is a blind scheme and its affect on audio quality or coding efficiency is almost negligible. It is highly configurable, thus it satisfies the necessities such as capacity, security and robustness which are intended for hiding high data capacities. So we can transmit resulting stego-audio without revealing that secret data is being exchanged. The proposed algorithm is analysed using statistical framework to show its level of security and also to prove its efficiency. It presents a theme that may transmit giant quantities of secret data and supply secure communication between two communication parties. It can be easily extended, resulting in better robustness, better data security and higher embedding capacity.

REFERENCES

1. Andreas Westfeld and Gritta Wolf," Steganography in a Video Conferencing System", Information Hiding 1998, LNCS 1525, pp. 32-47, 1998. Springer-Verlag Berlin Heidelberg 1998.
2. Neil F. Johnson and SushilJajodia, "Exploring Data Steganography: Seeing the Unseen" published in Journal Computer, Volume 31 Issue 2, February 1998
3. Cheok Yan Cheng, "Introduction On Text Compression Using Lempel, Ziv, Welch (LZW) method, updated 2001-5-17
4. Y. J. Dai., L. H. Zhang and Y. X. Yang.: A New Method of MPEG Video Watermarking Technology. International Conference on Communication Technology Proceedings (ICCT), 2003.
5. D. C. Wu and W. H. Tsai: A steganographic method for images by pixel-value differencing, Pattern Recognition Letters, Vol. 24, pp. 1613–1626, 2003
6. Arup Kumar Bhaumik, Minkyu Choi, RosslinJ.Robles, and MaricelO.Balitanas, " Data Hiding in Video", International Journal of Database Theory and Application Vol. 2, No. 2, June 2009
7. Ahmed Ch. Shakir," Stego Encrypted Message in Any Language for Network Communication Using Quadratic Method", Journal of Computer Science 6 (3): 320-322, 2010 ISSN 1549-3636 © 2010 Science Publications
8. D. P. Gaikwad and Dr. S.J. Wagh, "Color Image Restoration for Effective Steganography", i-manager's Journal on Software Engineering, Vol. 4l , pp.65-71, 3l January - March 2010.
9. S. Suma Christal Mary, "Improved Protection In Video Steganopgraphy Used Compressed Video Bitstream ," International Journal on Computer Science and Engineering Vol. 02, No. 03, pp. 764-766, ISSN: 0975-3397, 2010
10. Saurabh Singh and GauravAgarwal,"Hiding image to Video," A new approach of LSB replacement", International Journal of Engineering Science and Technology Vol. 2(12), 6999-7003, 2010.
11. Steganography on new generation of mobile phones with image and video processing abilities, as appeared Computational Cybernetics and Technical Informatics , 2010 International Joint Conference in Timisoara, Romania ISBN: 978-1-4244- 7432-5, 27-29 May 2010.
12. Sherly A P and Amritha P P, "A Compressed Video Data hiding using TPVD", International Journal of Database Management Systems(IJDMS) Vol.2, No.3, August 2010
13. Abdullah Bamatraf, Rosziati Ibrahim and Mohd. NajibMohd. Salleh, " A new Digital watermarking algorithm using combination of LSB and Inverse bit", Journal of Computing, volume 3, issue 4, april 2011.
14. Yusuf Perwej, FirojParwej, AsifPerwej, " An Adaptive Watermarking Technique for the copyright of digital images and Digital Image Protection,The International Journal of Multimedia & Its Applications (IJMA) Vol.4, No.2, April 2012
15. S.S.Verma, R.Gupta, G.Shrivastava, "A Novel Technique for Data Hiding in Audio Carrier by Using Sample Comparison in DWT Domain", 978-1-4799-3070-8/14 4th International Conference on Communication Systems and Network Technologies, 2014.
16. Nedeljko Cvejic, Tapio Seppanen, "Reduced distortion bit-modification for LSB audio steganography" Signal Processing, 2004. Proceedings. ICSP '04, 7th International Conference on (Volume:3), 2004.
17. Ajay B. Gadichal, "Audio Wave Steganography", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, Issue-5, November 2011.
18. S. A. Laskar and K. Hemachandran, "High Capacity data hiding using LSB Steganography and Encryption", International Journal of Database Management Systems (IJDMS) Vol.4, No.6, December 2012.

BIOGRAPHY

Ms Aparna V. Navlakha is a student pursuing M.E. from Computer Science and Engineering Department ,SIPNA College of Engineering, Amravati, Maharashtra, India

Dr Avinash D. Gawande is Head of Computer Science and Engineering Department, SIPNA College of Engineering, Amravati, Maharashtra, India