Voice Encryption with Watermarking for Secure Speech Communication

[1]Pushpalatha.G.S, [2]Dr.Ramesh S, [3]A. Raganna
[1]Asst. Professor, [2]Professor, [3]Asst. professor
[1] Department of Electronics and Communication,
[1]Dr. Ambedkar Institute of Technology, Bangalore,India

Abstract: The Demand for protection of data is increasing with the digital evolution. The advent of internet technologies made the multimedia data like speech, image or video is vulnerable to attack from various corners. Providing security for such data is at most important. In order to provide end to end security for speech data, an integrated approach of compression, water-marking and encryption of a speech signal has been proposed in this paper. Watermark provides authentication and ownership verification. Advanced Encryption Standard (AES) encryption is performed to provide security for watermarked speech signal. In addition to security, Discrete Cosine Transform (DCT) compression is performed to reduce the data-size of a speech signal.

IndexTerms - DCT compression, speech watermarking, AES encryption, speech steganography, speech cryptography

1. INTRODUCTION

The increasing demand of multimedia applications in communication system has tiled the way for secure communication. This is necessary to overcome unauthorized modifications and unwanted disclosure while transmitting speech and other data, especially in wireless channels. In secure speech communication systems the redundancy of the language plays an important role. The more redundant the language, the easier it is for an intruder to decipher the information with ease and convenience. That is why many real-world cryptographic implementations use a compression program to reduce the size of the signal before encryption. Compression of signal to lower rates with good speech quality not only eliminates the redundancy issue but also provides a lower bandwidth signal, which solves multiple problems in communication and multimedia applications. The possible threats which could attack in passive or active way includes eavesdropping, modification, replay, masquerading, penetration and repudiation. Speech signal is an essential component of most multimedia applications. Thus speech encryption is becoming increasingly important in video conference, news broad casting, military etc. Today, competitors, hackers, or governmental institutions can intercept any GSM cell call with relatively little effort and hence speech signal requires substantial level of security. This paper deals with integrating of speech compression, watermarking and encryption for providing secure communication [1].

The compression, watermarking and encryption which are required for secure communication are briefly discussed in the following literature.

I. Compression

Voice compression is the technique of reducing the data bits, in order to reduce the storage requirements of digital audio files and also to improve the data rate transfer. Compression can be either lossy or lossless. Lossless compression reduces bits by identifying and eliminating statistical redundancy. No information is lost in lossless compression. Lossy compression reduces bits by identifying marginally important information and removing it. The process of reducing the size of a data file is popularly referred to as data compression, although it's formal name is source coding.

Audio data compression has the potential to reduce the transmission bandwidth and storage requirements of audio data. Audio compression algorithms are implemented in software as audio codecs. Lossy audio compression algorithms provide higher compression at the cost of fidelity and are used in numerous audio applications. These algorithms almost all rely on psychoacoustics to eliminate less audible or meaningful sounds, thereby reducing the space required to store or transmit them [5].

In both lossy and lossless compression, information redundancy is reduced, using methods such as coding, pattern recognition, and linear prediction to reduce the amount of information used to represent the uncompressed data.

The DCT, and in particular the DCT-II, is often used in signal and image processing, especially for lossy data compression, because it has a strong "energy compaction" property (most of the signal information tends to be concentrated in a few low-frequency components of the DCT) [5].

II. Watermarking

Digital watermarking is the technique of inserting specific information into signal, data, image or video. The specific information is known as watermark and usually used for ownership identification, authentication purpose and protection of integrity of data. Due to rapid progress in wireless communication systems, extreme prevalence mobile systems, and smart card technology, information is more vulnerable to abuse. For these reasons, it is important to make information systems secure to protect data and resources from malicious acts [2].

The watermark is extracted or detected at any point where identification or integrity is concerned. Traditional digital watermarking schemes are mainly based on spatial-domain or transform-domain, such as Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT).

Spatial Domain Techniques: Spatial domain watermarking, in general, is easy to implement on computational point of view but too fragile to withstand large varieties of external attacks.
A robust, computationally efficient and blind digital image watermarking in spatial domain has been proposed by Santi Prasad Maity& Malay Kumar Kundu [7]. In their technique watermark insertion process exploits average brightness of the homogeneity regions of the cover image. Spatial mask of suitable size is used to hide data with less visual impairments and recovery process needs only one secret image.
Ibrahim Nasir, Ying Weng, Jianmin Jiang in their paper presents a new robust watermarking scheme for color image based on a block probability in spatial domain [8].
Spatial Watermarking Method, based on a Logarithmic transformation of an encrypted embedded mark is presented by HassenSeddik, MounirSayadi, FarhatFnaiech, and Mohamed Cheriet [9].
Transform Domain Techniques: Generally these techniques uses discrete cosine transform DCT or DWT, which provide higher imperceptibility and are much more robust to manipulations.
Saeed K. Amirgholipour, Ahmad R. Naghsh-Nilchi proposed an algorithm based on Joint DWT-DCT Transformation. A binary watermarked logo is scrambled by Arnold cat map and embedded in certain coefficient sets of a 3-level DWT transformed of a host image. Then, DCT transform of each selected DWT sub-band is computed and the PN-sequences of the watermark bits are embedded in the middle frequencies coefficients of the corresponding DCT block [10].
MahaSharkas, Dahlia ElShafie, and NadderHamdy proposed dual digital image watermarking technique which incorporates two watermarks in a host image for improved protection and robustness [11].
J.L. Liu et al. proposed watermarking by modifying the original image in transform domain and embedding a watermark in the difference values between the original image and its reference image [12].
John N. Ellinas proposed a robust watermarking algorithm using the wavelet transform and edge detection in 2007. The watermark embedding process is carried over the subband coefficients that lie on edges, where distortions are less noticeable, with a subband level dependent strength [13].
A paper by A. R. Stauffer, A. D. Lawson, entitled ”Speaker Recognition on Lossy Compressed Speech using the Speex Codec” presented in 2009 examines the impact of lossy speech coding with Speex on GMM-UBM speaker recognition (SR) [14].
Hussain Mohammed DipuKabir, Saeed Anwar, Abu Shahadat Md. Ibrahim, Md. Liakot Ali, Md. Abdul Matin proposed watermark with fast encryption for FPGA based secured speech communication in 2013[15].

III. Encryption

Encryption is the process of encoding messages or information in such a way that only authorized parties can read it [16]. Encryption doesn't prevent hacking but it reduces the likelihood that the hacker will be able to read the data that is encrypted.
Encryption can be broadly classified into two types namely, Symmetric Key Encryption and Public Key Encryption.
In Symmetric-key schemes, the encryption and decryption keys are the same. Thus communicating parties must agree on a secret key before they wish to communicate [17]. Ex: Advanced Encryption Standard (AES).
Public-key cryptography refers to a cryptographic algorithm which requires two separate keys, one of which is private and one of which is public. Although different, the two parts of this key pair are mathematically linked, but the parameters are chosen so that calculating the private key from the public key is either impossible or prohibitively expensive. The public key is widely distributed, while the private key is known only to its proprietor. Ex: RSA (Ron Rivest, Adi Shamir and Len Adleman) [18].
The generation of encryption secret keys with a high level of security is crucial to ensure secure data storage and is a challenging topic of investigation while technology has changed the world administration and education of under-developed countries. Hence secured technology is required for administrations. Mobile banking has also become common in developing and overpopulated countries. This increasing demand of online transactions face security challenges, thus trust and security issues of mobile banking find its importance for developing countries. There are many techniques like robust, fragile approaches in watermarking. The robust watermarking approach protects the copyright identifier of data in which watermarks are not easily removed by attacks. The fragile watermarking approach confirms content integrity. However a security problem arises due to many factors like copy-and-paste, counterfeiting and transplantation attacks, deletion and insertion attacks etc.
Encryption is needed to ensure protection of data in the medium. Traditional symmetric key encryption algorithms like Data Encryption Standard (DES) use small blocks size with complex permutations process to give secure output cipher text. Public key algorithms are not suitable for large amount of data due to its slow performance. Advanced Encryption Standard (AES) was announced by National Institute of Standards and Technology (NIST) on 2001. AES is one of the most secure algorithms used in symmetric key cryptography. It uses complicate repeated steps to prevent analytic attacks that can discover weakness in the algorithm and so attack any encrypted data. AES use high diffusion to eliminate any prediction of key. The only problem with AES is its sensitivity to noise due to its high diffusion [3]. The diffusion make elements within each block depend on each other (mix-column step). So if one element or more missed or corrupted by noise it will affect the surrounding elements and the error will propagates and increase in next round. Therefore the AES problem may appear in noisy channels only because of repeated rounds that cause propagation of error. Decreasing number of rounds or cancel mix column step will affect the security of algorithm so solution will be depend on other security aspects to protect algorithm when decreasing diffusion [3].

In this paper, an integrated approach of compression, watermarking and encryption algorithm over speech signal is presented.
Compression is needed to prevent the increase in data size of the data in transmission channel. Watermarking inserts specific information into signal which ensures ownership verification. Encryption provides security while the signal traveling via the transmission channel. AES encryption is performed for compressed and watermarked speech signal.

2. Proposed Methodology

The main objective of this work is to design and implement voice encryption with compression and watermarking for secure speech communication, thus to provide authentication, ownership identification and also to maintain confidentiality of the data, which provides security. It is represented in the block diagram as shown in fig. 1.

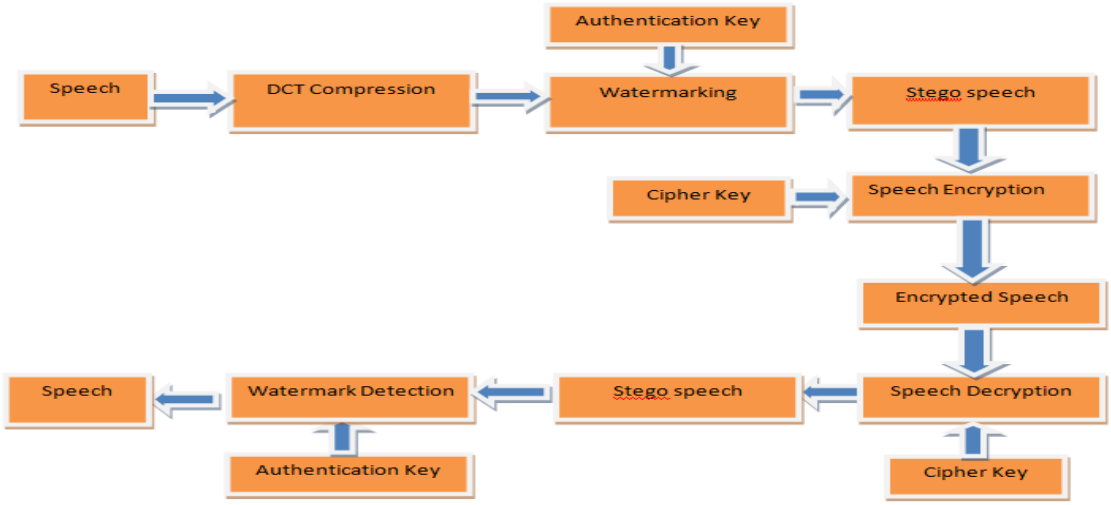

Fig.1. Block diagram of proposed secure communication system

The main components of the block diagram are as explained below.

2.1. Speech Compression

Voice compression is performed using DCT, which expresses a finite sequence of data points in terms of sum of cosine functions oscillating at different frequencies. In short, DCT is a technique for converting a signal into its elementary frequency components. It is widely used in image and audio compression.

Equation 1 represents DCT transform of the sequence and Equation 2 represents its corresponding Inverse DCT (IDCT) [19]. Equation 3 and 4 represents the value of α(k) for various values of k.

$$C(k) = \alpha(k)\sum_{n=0}^{N-1} x(n)\cos\left[\frac{\pi(2n+1)k}{2N}\right], 0 \le k \le (N-1) \quad (1)$$

$$x(n) = \sum_{k=0}^{N-1} \alpha(k)C(k)\cos\left[\frac{\pi(2n+1)k}{2N}\right], 0 \le n \le (N-1) \quad (2)$$

where

$$\alpha(k) = \sqrt{\frac{1}{N}} \text{ for } k = 0 \quad (3)$$

$$\alpha(k) = \sqrt{\frac{2}{N}} \text{ for } 1 \le k \le N-1 \quad (4)$$

In this paper the DCT compression technique is employed to compress the speech samples. Different compression factors like 2, 4, 8 etc., are used to compress the signal. For higher compression factors there is significant loss in speech quality.

2.2. Speech Watermarking

For the compressed speech samples, text data for example 'REVA' is embedded as watermark in spatial domain for authentication purpose to provide ownership verification. This paper uses spatial domain watermarking.

2.3. AES Encryption

The Advanced Encryption Standard (AES) specifies a Federal Information Processing Standard (FIPS) approved cryptographic algorithm that can be used to protect electronic data. AES algorithm is a symmetric block cipher that can encrypt (encipher) and decrypt (decipher) information.

Encryption converts data to an unintelligible form called cipher text; decrypting the cipher text converts the data back into its original form, called plaintext.

The input and output for the AES algorithm consist of sequences of 128 bits or 16 bytes (digits with values of 0 or 1). These sequences will sometimes be referred to as blocks and the number of bits they contain will be referred to as their length. The Cipher Key for the AES algorithm is a sequence of 128 bits.

3. Implementation Procedure

The proposed secure communication system has been implemented using MATLAB on a speech signal which has the properties mentioned in Table 1.

Table 1: Speech Signal Information obtained using MATLAB's "audioinfo" function

Parameter	Value
Filename	Speech.wav
CompressionMethod	'Uncompressed'
NumChannels	1
SampleRate	22050
TotalSamples	63039
Duration	2.8589
BitsPerSample	16

The compression, watermarking and encryption implementation steps are as mentioned below.

3.1. DCT Compression

- The speech signal is read using MATLAB and the speech samples are converted into 1D row vector
- A suitable compression factor is selected. Ex: 2
  [Note: For higher compression factors, there is a significant loss in speech fidelity]
- A block size of 8192 samples is selected for each run. DCT is applied to time domain speech samples in terms of block size of 8192. This results in frequency samples having block size of 8192, out of which lower 4096 frequency samples (blocksize/compressionfactor) are unaltered and upper 4096 frequency samples are replaced by zeros

Inverse DCT is applied to the resultant 8192 frequency samples to obtain 8192 time domainsamples. During this process, high frequency components are eliminated which results in speech compression.

3.2. Spatial Domain Watermarking

- User creates an authentication key and its equivalent ASCII values are to be determined. Ex: The authentication key 'REVA' is converted into its corresponding ASCII values 82, 69, 86, 65
- The ASCII equivalent values of the authentication key are then normalized in the range -1 to +1 to avoid clipping of the ASCII values. Ex: 82, 69, 86, 65 is divided by 1000 to get 0.0820, 0.0690, 0.0860, 0.0650 which is in the range -1 to +1
- The normalized ASCII values are then embedded into the compressed speech signal by replacing the speech samples at alternate positions, starting from the position which is equal to the length of the key. Ex: The normalized values of 'REVA' i.e., 0.0820, 0.0690, 0.0860, 0.0650 are embedded into speech signal by replacing the speech samples at positions 4, 6, 8 and 10. Note that the embedding is done with starting position as 4, which is equal to the length of the word 'REVA'
- The watermark will be detected at the destination

3.3. AES Encryption and Decryption

- The watermarked speech samples in the previous step are in the range -1 to +1 range. This range is converted into 0 to 255 range (00H to FFH hexadecimal) as required for AES algorithm for whole numbers in the range 0 to 255
- During this conversion, the resultant samples have both integer part and floating part
- So, the integer and fractional part are to be separated. Ex: In 125.7689, integer part 125 is retained as it is and 0.7689 is extracted for further manipulation
- 0.7689 is multiplied by 10000 to get 7689 and divided by 40 which results in 192.225. The fractional part is rounded to 192
- As a result a speech sample having value 125.7689 is converted into two samples 125 and 192. This procedure is repeated for all speech samples. The resultant speech signal will have integer part of all samples which is followed by fractional part of all samples. As a result, length of speech signal is doubled
- AES encryption is applied to this speech samples using 16 bytes crypto-key for every 16 bytes of speech samples. The resultant encrypted signal is again normalized in the range 0 to 1 and saved as "transmitter.wav" audio file with a sampling frequency of 22050 HZ

During decryption, the "transmitter.wav" which is in the range 0 to 1 is converted to 0 to 255 and AES decryption process is employed with the same 16-byte crypto-key used for encryption. In the decrypted samples, the modified fractional part Ex: 192 is divided by 10000 and multiplied by 40 to get .768

- It is added with integer part 125 to get 125.768. Similar procedure is repeated for all speech samples. The samples which were doubled earlier are now halved to get original speech samples
- These speech samples are in the range 0 to 255, are normalized in the range -1 to +1 and played using audio player to perceive the compressed, watermarked and decrypted speech samples

3.4. Watermark Authentication

- At destination, User is prompted to enter authentication key Ex: 'REVA' for ownership verification. Length of the key entered by the user is determined (Length of 'REVA' is 4) and ASCII equivalent of the key is calculated (82, 69, 86, 65)
- Using this length, the values in the decrypted speech samples at positions 4, 6, 8 and 10 are extracted. Ex: Extracted values will be 0.0820, 0.0690, 0.0860, and 0.0650. This is multiplied by 1000 to get 82, 69, 86 and 65. These values are compared with ASCII values of the key entered by the user. If the key matches, ownership of the speech signal is verified

4. Implementation Results

The implementation is done using MATLAB for different compression factors with different stego-text and crypto-key. The implementation results are as follows

4.1. Compression, watermarking, encryption, decryption and watermark identification with compression factor 2

Table 2 indicates the data used for Compression, watermarking, encryption, decryption and watermark identification with compression factor 2.

Table 2: Data used for compression, watermarking, encryption, decryption and watermark identification of speech signal with compression factor 2

Factor	Value
Compression factor	2
Authentication key	'REVA'
Encryption Key	{'00' '01' '02' '03' '04' '05' '06' '07' '08' '09' '0a' '0b' '0c' '0d' '0e' '0f'}

The time domain and spectrogram plots of transmitted speech signal are as shown in Fig. 2 and Fig. 3 respectively.

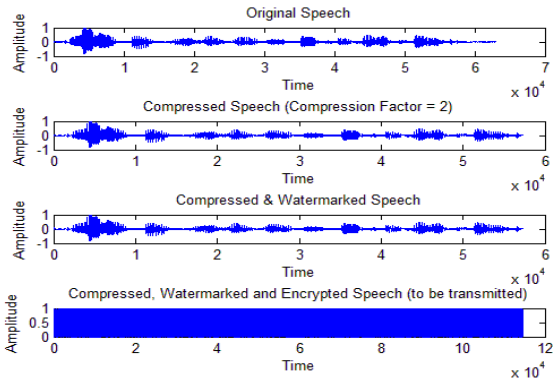

Fig 2. Time domain plot for compressed, watermarked and encrypted speech with compression factor 2

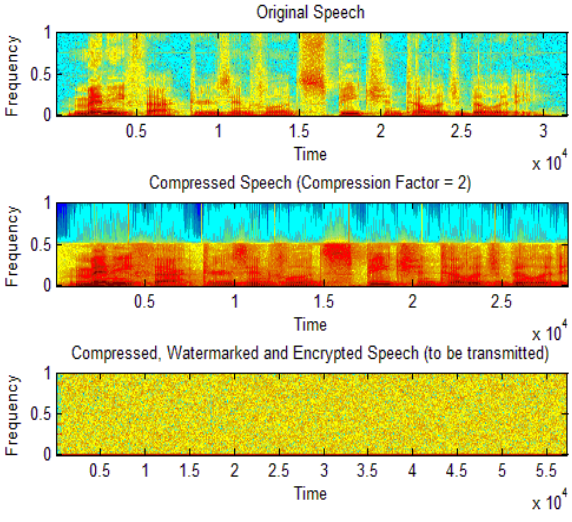

Fig 3. Spectrogram plot for compressed, watermarked and encrypted speech with compression factor 2

The time domain and spectrogram plot of received speech signal are shown in Fig. 4.

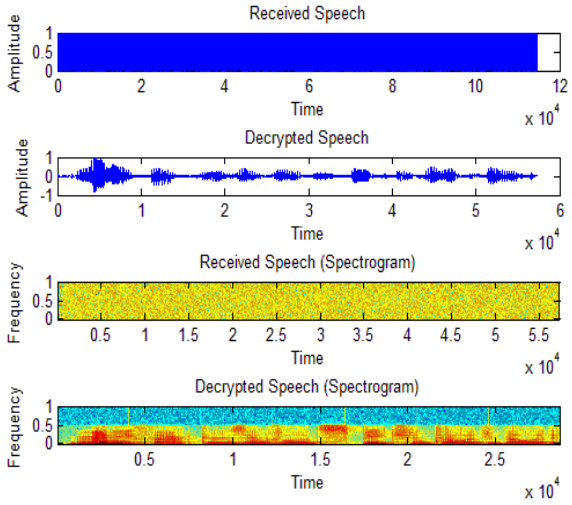

Fig 4. Time domain plot and Spectrogram plot for decrypted and authenticated speech with compression factor 2

The time domain and spectrogram plots of transmitted and received signals for compression factor of 2 shows that the original speech signal has been encrypted and decrypted successfully without a major loss in the original speech.

4.2. Compression, watermarking, encryption, decryption and watermark identification with compression factor 4

Table 3 indicates the data used for Compression, watermarking, encryption, decryption and watermark identification with compression factor 4.

Table 3: Data used for compression, watermarking, encryption, decryption and watermark identification of speech signal with compression factor 4

Factor	**Value**
Compression factor	4
Authentication key	'Secret Key is 123!@#'
Encryption Key	{'2b' '7e' '15' '16' '28' 'ae' 'd2' 'a6' 'ab' 'f7' '15' '88' '09' 'cf' '4f' '3c'}

The time domain plots and spectrogram plots of to be transmitted speech signal are shown in Fig. 5 and Fig. 6 respectively.

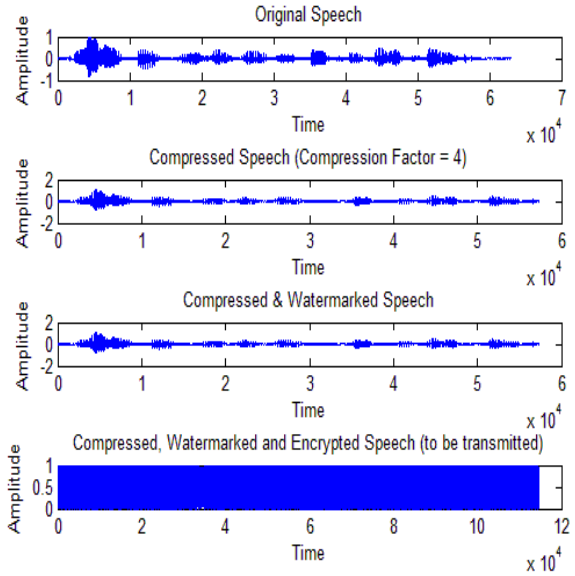

Fig 5. Time domain plot for compressed, watermarked and encrypted speech with compression factor 4

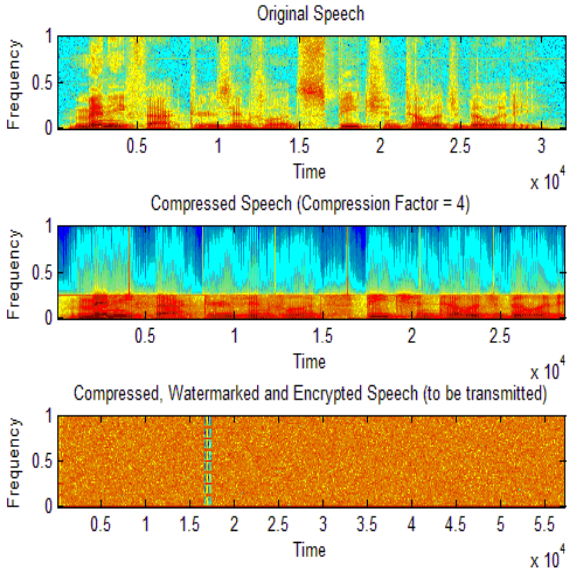

Fig 6. Spectrogram plot for compressed, watermarked and encrypted speech with compression factor 4

The time domain and spectrogram plots of received speech signal are shown in Fig. 7.

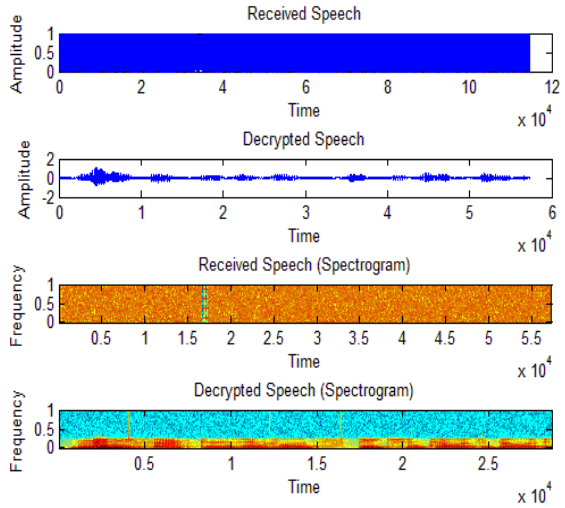

Fig 7. Time domain plot and Spectrogram plot for decrypted and watermark authenticated speech with compression factor 4

The time domain and spectrogram plots of transmitted and received signals for compression factor of 4 shows that the original speech signal has been encrypted and decrypted successfully without a major loss in the original speech.

4.3. Compression, watermarking, encryption, decryption and watermark identification with compression factor 8

Table 4 indicates the data used for Compression, watermarking, encryption, decryption and watermark identification with compression factor 8.

Table 4: Data used for compression, watermarking, encryption, decryption and watermark identification of speech signal with compression factor 8

Factor	**Value**
Compression factor	8
Authentication key	'MATLAB SOFTWARE'
Encryption Key	{'00' '01' '02' '03' '04' '05' '06' '07' '08' '09' '0a' '0b' '0c' '0d' '0e' '0f'}

The time domain plots and spectrogram plots of to be transmitted speech signal are shown in Fig. 8 and Fig. 9 respectively.

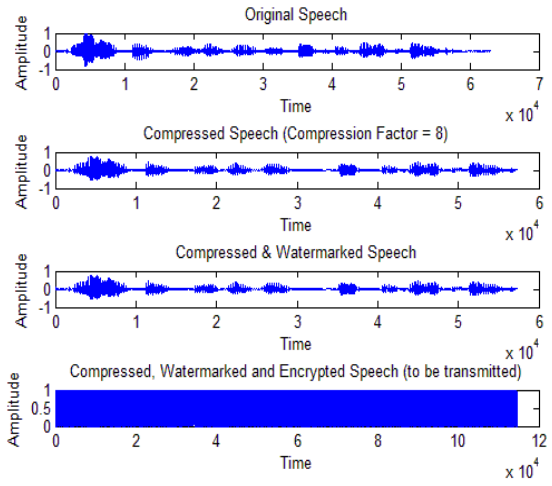

Fig 8. Time domain plot for compressed, watermarked and encrypted speech with compression factor 8

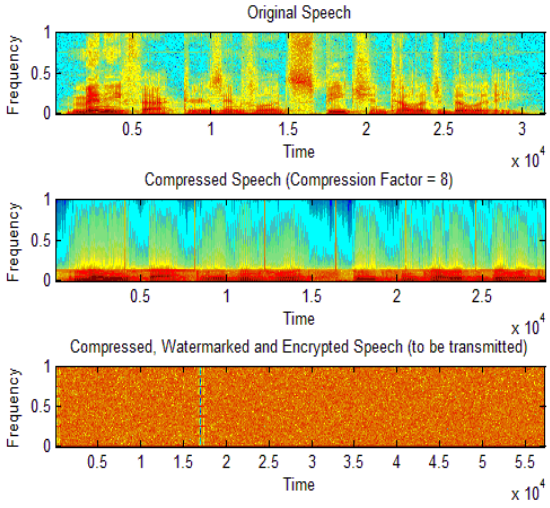

Fig 9. Spectrogram plot for compressed, watermarked and encrypted speech with compression factor 8

The time domain and spectrogram plots of received speech signal are shown in Fig. 10.

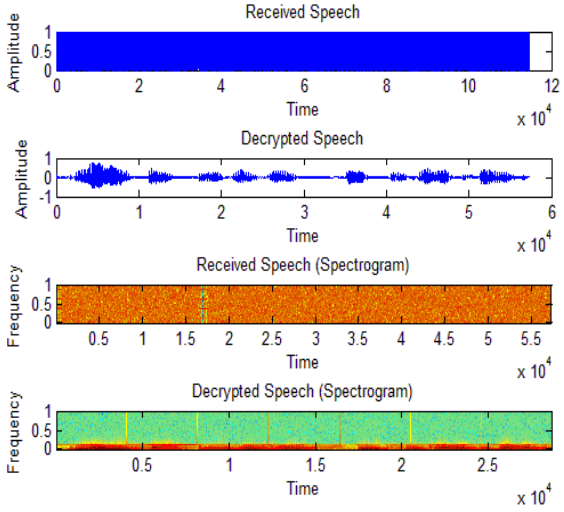

Fig 10. Time domain plot and Spectrogram plot for decrypted and watermark authenticated speech with compression factor 8

The time domain and spectrogram plots of transmitted and received signals for compression factors of 2, 4 and 8 are represented in the figures from 1 to 10. It has been observed from the plots that with the increasing compression factor the original signal has experienced a losssy compression, although which maintains the same level of encryption and original data can be decrypted without much loss, thus providing the security for the original speech data.

5. Conclusion

An integrated approach of compression, watermarking and encryption of a speech signal provides security for a speech to a great extent. Compression ensures less buffer and memory space usage and thus high data transmission rate. The watermark is not easily detectable and cannot be suspected as it is merged with the original speech without increasing the actual length of the speech. The encryption is performed with highly efficient AES algorithm so, if any attempt to decrypt with wrong key results in noise itself. Thus it prevents network threats like eavesdropping, Man-in-the-Middle attack, etc. It is possible to extract original speech signal only when all the information like proposed algorithm, watermark detection key and key for decryption is known. Thus it provides highly secured and efficient algorithm for speech transmission.

Future Scope

Existing algorithms provides security and authentication for a speech independently. Naturally the integrated approach of speech encryption and watermarking further improves the overall security.

Acknowledgement

The authors would like to thank the organization for providing the support. The author would also thank Mr. Mohan Kumar N and Manivannan K for their contributions.

References

[1] Roger Sutton, "Secure Communication: Applications and Management", John Wiley and Sons Ltd, 2002, 4-8
[2] Ingemar J. Cox, Matthew L. Miller, Jeffrey A. Bloom, Jessica Fridrich, Ton Kalker,"Digital Watermarking and Steganography", 2nd Edition, Elsevier Inc, 2008, 35-38
[3] William Stallings, "Cryptography and Network Security", 4th Edition, Pearson Education India, 2011, 134-194
[4] Christopher H. Sterling,"Military Communications: From Ancient Times to the 21st Century", ABC-CLIO Inc, 2008, 409-412
[5] Omar Adil Mahdi, Mazin Abed Mohammed, Ahmed Jasim Mohamed, "Implementing a Novel Approach an Convert Audio Compression to Text Coding via Hybrid Technique", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 6, No 3, November 2012, ISSN (Online): 1694-0814
[6] Sasmita Mishra, AmitavMahapatra, Pranati Mishra,"A Survey on Digital Watermarking Techniques", International Journal of Computer Science and Information Technologies, Vol. 4 (3), 2013, 451-456
[7] Santi Prasad Maity, Malay Kumar Kundu, "Robust and Blind Spatial Watermarking in Digital Image"
[8] Ibrahim Nasir, Ying Weng, Jianmin Jiang,"A New Robust Watermarking Scheme for Color Image in Spatial Domain", SITIS '07, 3rd International IEEE Conference on 16-18 Dec. 2007
[9] HassenSeddik, MounirSayadi, FarhatFnaiech, Mohamed Cheriet,"A New Spatial Watermarking Method, based on a Logarithmic transformation of An Encrypted embeded Mark", Seventh IMACS Seminar on Monte Carlo Methods (MCM2009)
[10] Saeed K. Amirgholipour, Ahmad R. Naghsh-Nilchi,"Robust Digital Image Watermarking Based on Joint DWT DCT", Convergence and Hybrid Information Technology, ICCIT – 2008
[11] MahaSharkas, Dahlia El Shafie, NadderHamdy,"A Dual Digital-Image Watermarking Technique", World Academy of Science, Engineering and Technology, 2005
[12] Jiang-Lung Liu, Der-Chyuan Lou, Ming-Chang Chang, Hao-Kuan Tso, "A robust watermarking scheme using selfreference image", Computer Standards & Interfaces (2006), 356–367
[13] John N. Ellinas, "A Robust Wavelet-Based Watermarking Algorithm Using Edge Detection", World Academy of Science, Engineering and Technology, 2007
[14] Lawson, A. D, A. R. Stauffer,"Speaker Recognition on Lossy Compressed Speech Using the Speex Codec", ISCA 6-10 September, Brighton UK, 2009
[15] Hussain Mohammed DipuKabir, Saeed Anwar, Abu Shahadat Md. Ibrahim, Md. Liakot Ali, Md. Abdul Matin,"Watermark with Fast Encryption for FPGA Based Secured Realtime Speech Communication", Consumer Electronics Times Jan. 2013, Vol. 2 Iss. 1, 75-84
[16] OdedGoldreich, "Foundations of Cryptography: Volume 2, Basic Applications", Cambridge university press, 2004.
[17] Joan Daemen, Vincent Rijmen,"Specification for the Advanced Encryption Standard (AES)", Federal Information Processing Standards Publication 197, 2001
[18] Ron Rivest, Adi Shamir, Leonard Adleman, "RSAES-OAEP Encryption Scheme", RSA Laboratories, RSA Security Inc, 2000
[19] John G. Proakis, Dimitris G. Manolakis, "Digital Signal Processing, Principles, Algorithms and Applications", 4th Edition, Prentice Hall of India, 2007, 495-501