International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 9, Number 23 (2014) pp. 19469-19476

http://www.ripublication.com

Combined Speaker Recognition and Compression for Secured Speech Transmission

D. Ambika[1] and V. Radha[2]

Research Scholar[1], Department of Computer Science,
Avinashilingam Institute for Home Science and Higher Education for Women,
Coimbatore, India,
Professor[2], Department of Computer Science,
Avinashilingam Institute for Home Science and Higher Education for Women,
Coimbatore, India
ambikaphdscholar@gmail.com[1], radhasrimail@gmail.com[2]

Abstract:

The continued advancement and growth in the network processing ability have improved the competence and scope of speech communication in the network. Therefore the secure speech communication and security of communication information are required now a days. To provide more security for the speech signal, the Multiple Huffman Table (MHT) is enhanced in this paper, by incorporating an optimization procedure as a preprocessing step before compression. So before performing compression the MHT splits the speech signal into Most Significant Signal (MSS) and Least Significant Signal (LSS). MSS is considered to be the significant signal, so the MHT considers only the MSS for compression. From the compressed output, the coefficients are selected and protected using cryptographic cipher Advanced Encryption Standard (AES), because it will be difficult for attackers to recover any information from these coefficients. Finally the results were evaluated using Compression ratio (CR), Peak Signal-to-Noise Ratio (PSNR) and Normalized Root-Mean Square Error (NRMSE). From the results it can be seen that the Enhanced MHT with Encryption method provides better result.

Keywords: Secure Speech Communication, Multiple Huffman Coding, Advanced Encryption Standard

1 Introduction

Communication plays a dominant role and considered to be the significant feature in

the information society field. The rapid growth of information technlogy has created a demand for security and privacy in the network communication channel. More and more speech services through networks are realized now a days. Any type of multimedia data such as speech,text which is transmitted through the network needs to be protected fom any alteration,forgery and theft. Human beings are living in the information society and communication security and hence the confidentiality is required in public lifes such as network voice communication,mobile banking, online payment on the internet etc.The secure communication is used to protect state secrets,commercial secrets, individual privacy which is essentail for the nation and society. The speech communication becomes an important way to transfer information through voice and can be used for military,diplomatic,economic and as well as in research purposes. In digital communication systems,speech signals are encoded into binary sequences,which are transmitted through channels,and then decoded,recovered in the form of understandable speech. These conversions have unique advantages such as it avoids noise intereference during transmission and storage procedures,and it is easier to process,encrypt,regenerate and forward encode signals.

In communication, coding can be used as a signal processing which transforms a signal in an appropriate form that can be transmitted to channel. The coding can be divided into two types such as souce coding and channel coding.The source coding is used to improve the signal transmission and the storage efficiency.The channel coding is used to improve the transmission reliability and it can be called as reliability coding.The compression coding tries to identify and eliminate redundancy in order to reduce bit-rate and it can be lossy or lossyless.The redundancy reside in in the following aspects[1]:

- There is a strong correlation between signal samples, that is, the short-time spectrum is not flat.
- The dull resonance speech segment is quasi-periodic.
- The shape of the sound track and its change speed is limited.
- The probablity distribution of transmitted codes is uniform.

The traditional secure speech communication technology uses mainly the Analog Scrambling and Digitized Encryption. The Analog Scrambling segments the speech signal using frequency domain or time domain in order to change the signal into unintelligible form. Whereas in digitized encryption the speech is digitized and then it is encrypted. But both of the methods depend on the transformation of speech signal and cryptographic strength to protect the speech communication. The existing system [2] deals with the integrating of speech coding, with speaker authentication and encryption. Figure 1 show the existing method of speech communication and the steps involved are explained below:

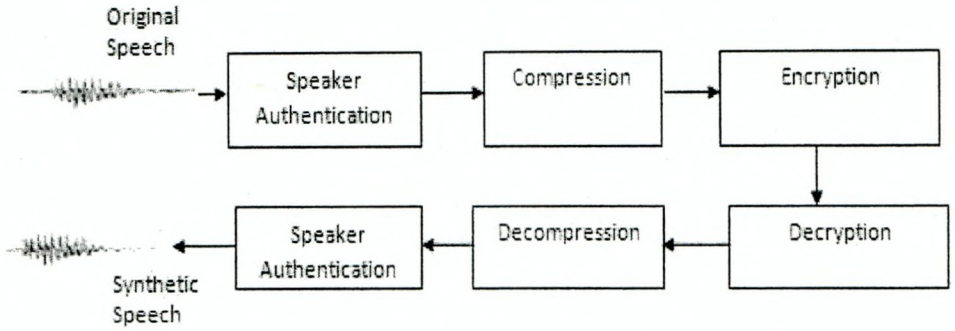

Figure1:Existing speech communication

- By using the speech recognition algorithms, the speaker is identified.
- Next the speech signal is compressed.
- After performing compression, the encryption is performed.
- Then the speech signal is transmitted from the sender to the receiver side.
- In the receiver side the signal is decrypted and decompressed.
- The speaker authentication is done to verify if there is any alterations done while transmission. If the alteration is found then the message is not considered by the receiver.

In order to improve and to provide more security this paper concentrates only on compression part which is enhanced using the Multiple Huffman Coding. The structure of the paper is organized as follows. In section 2 the Proposed Methodology is discussed. Section3 analyzes the Compression and Encryption. Performs evaluation is presented in Section 4. Finally, the conclusion is summarized in Section 5.

2 Proposed Methodology

- The first part in the proposed secure speech communication is to perform speaker recognition and the details regarding the recognition part is explained in [3].
- After performing speaker identification,if the sender is not an authenticated person there will be a security alert.
- In this paper the Huffman coder is enhanced by incorporating an optimization procedure as a preprocessing step before compression. So, the MHT splits the speech signal into Most Significant Signal (MSS) and Least Significant Signal (LSS). The MSS alone are considered for compression using Multiple Huffman Table (MHT) coding scheme.
- From the compressed output, the coefficients are selected and it is protected

using cryptographic cipher Advanced Encryption Standard (AES), so that it will be difficult for attackers to recover any information from these coefficients.

- After performing the compression, and encryption, the speech signal is transmitted from the sender to the receiver side.
- In the receiver side after performing the speaker authentication, the receiver decrypts and decompresses the data to get the secret speech signal. If the receiver is not an authenticated person there will be a security alert.
- To provide more seecurity , this paper concentrates only on the compression part which is highlighted in the Figure 2.

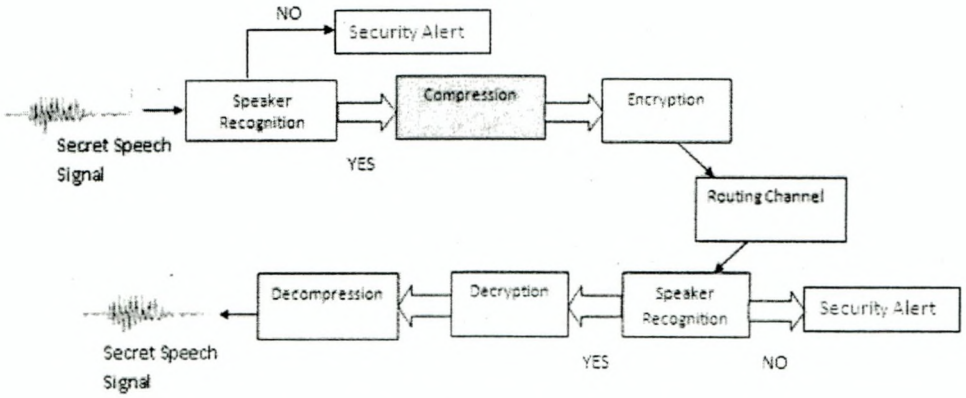

Figure 2 Proposed Secure Speech Communication

3 Compression and Encryption

Compression of signal to lower rates with good speech quality eliminates the redundancy issue and if there is more redundancy, then it is easier for an intruder to decipher the information. In order to avoid this many real-worlds cryptographic implementation uses a compression program to reduce the size of the signal before encryption [4]. Huffman is compatible with a great number of existing multimedia contents. It can be used for compressing data with variable - length codes. It can be implemented using a Look Up Table (LUT) or Multiple Look up Tables. The author Wu et al (2005) proposed a scheme which is based on encoding with Multiple Huffman Tables (MHT) can be used alternately in a secret order [7]. The encryption with reasonably high level of security and unaffected compression can be achieved simultaneously in MHT technique. One of the major advantages by using this kind of joint encryption-compression approach is that encryption and compression can be achieved in one single step, which simplifies the system design and makes it flexible for some advanced multimedia processing such as scalability and rate shaping.

In this paper the MHT is used to perform compression, because it is said to be the popular and widely used method which is used in multimedia standards such as JPEG, MPEG and MP3. The MHT is enhanced by incorporating an optimization procedure as a preprocessing step before compression. In the preprocessing step the

speech signal is divided into MSS and LSS. The MSS is said to be the most important signal and it is considered for compression using MHT. It also aims to increase the model space by maintaining the computational efficiency. While using this scheme, it makes use of standard coding tables, keep the structure of the Huffman tree but enlarge the model space through tree mutation [8]. After performing the compression, the compressed speech signal is divided into selected and other coefficients. Only the selected coefficients are encrypted using the AES algorithm as encryption is a very common technique for promoting the security and it is a much stronger method of protecting speech communication than any form of scrambling and it can be Symmetric or Asymmetric Encryption.

The AES is used to achieve confidentiality. It is an encryption standard, which uses a symmetric key cipher where both sender and receiver use a single key for encryption and decryption [5]. It operates on 128-bit blocks of data while the key length can be 128,192 or 256 with total rounds of 10, 12 or 14 respectively. This algorithm in each round scrambles the bytes of the state either row wise or column wise except the final round by applying four transformations such as: the Sub Bytes, the Shift Rows, the Mix Columns, and the Add Round Key, while the final round does not have the Mix Columns transformation. The Sub Bytes is a nonlinear transformation, which computes the multiplicative inverse of each byte followed by an affine transformation. Shift Rows is a permutation function in the Cipher round. In Shift Rows, bytes in each row of the state are shifted cyclically to the left. The number of places each byte is shifted differs for each row.

4. Performance Evaluation

All the experiments were performed using MATLAB 7.1 on different speech signals which are collected from MIT corpus. In each session, 54 speech samples were collected per user. Totally 5,184 examples from enrolled users (2,592 per session) and 2,700 imposter examples from users not in the enrollment set. Within the enrolled set of 48 speakers, 22 were female while 26 were male where each session yielded 54 speech samples per user. For the imposter set of 40 speakers, 17 were female while 23 were male. The utterances were recorded in three acoustic environments: office, lobby, and street intersection via two types of microphones: external earpiece headset and built-in mobile device. In this paper the Objective analysis is done for randomly ten speakers and they are evaluated using Compression Ratio (CR), Peak Signal to Noise ratio (PSNR), and Normalized Root Mean Square Error Rate (NRMSE) and the formulas are given in equation 1,2 and 3 [6]. The figure 3, 4 and 5 represents the performance of the techniques in a graphical form.

A. Compression Ratio

$$CR = \frac{Length(x(n))}{Length(r(n))} \qquad (1)$$

where x(n) is the original signal and r(n) is the reconstructed signal [6]

B. Peak Signal to Noise Ratio(PSNR)

$$PSNR = 10log_{10}\frac{(NX^2)}{\|X-r\|^2} \qquad (2)$$

where N is length of the reconstructed signal , X is the maximum absolute square value of the signal x and || x-r | |^2 is the energy of the difference between the original and reconstructed signals [6].

C. Normalized Root Mean Square Error (NRMSE)

$$NRMSE = \sqrt{\frac{(x(n)-r(n))^2}{(x(n)-\mu x(n))^2}} \qquad (3)$$

where x (n) is the speech signal, r(n) is the reconstructed signal and x(n) is the mean of the speech signal[6]. The results were analyzed for various methods where H is the Huffman coding. It is the existing Huffman code where it performs only the compression. HMT- is the enhanced Multiple Huffman Table Coding, where it performs only the compression using the multiple Huffman table. HMT –E is the multiple Huffman table with Encryption. It performs the compression with encryption using the AES encryption algorithm. From the results it can be seen that the enhanced HMT when combined with encryption provides better result for all the speakers.

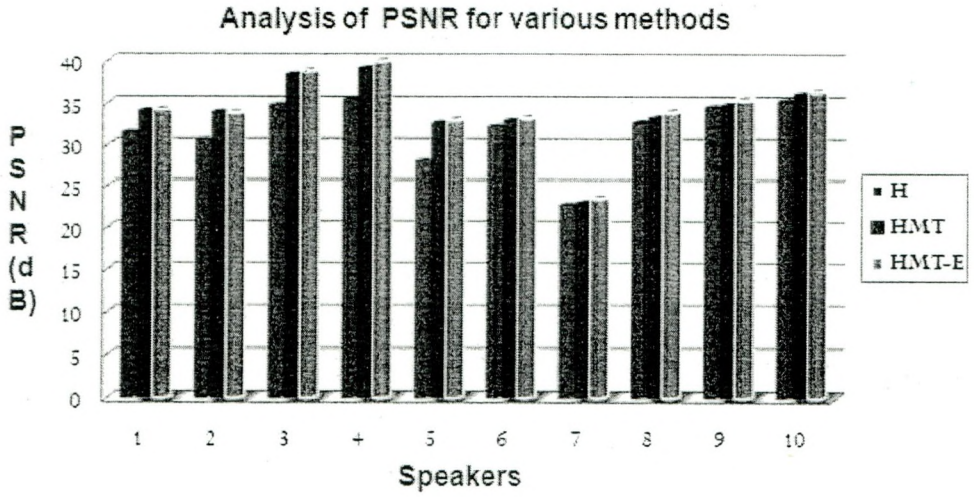

Figure 3 Performance Evaluation Based on PSNR

Huffman Coding - H
Multiple Huffman Table Coding - HMT
HMT with Encryption - HMT-E

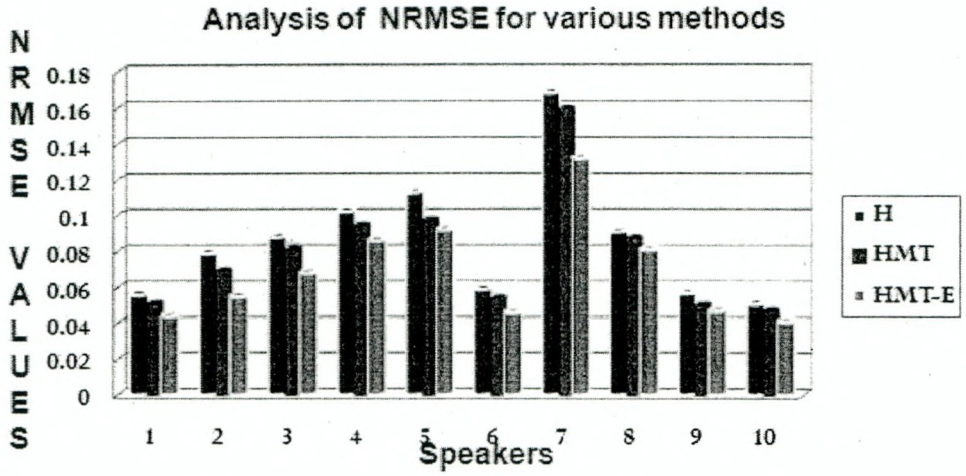

Figure 4 Performance Evaluation Based on NRMSE

Huffman Coding - H
Multiple Huffman Table Coding - HMT
HMT with Encryption - HMT-E

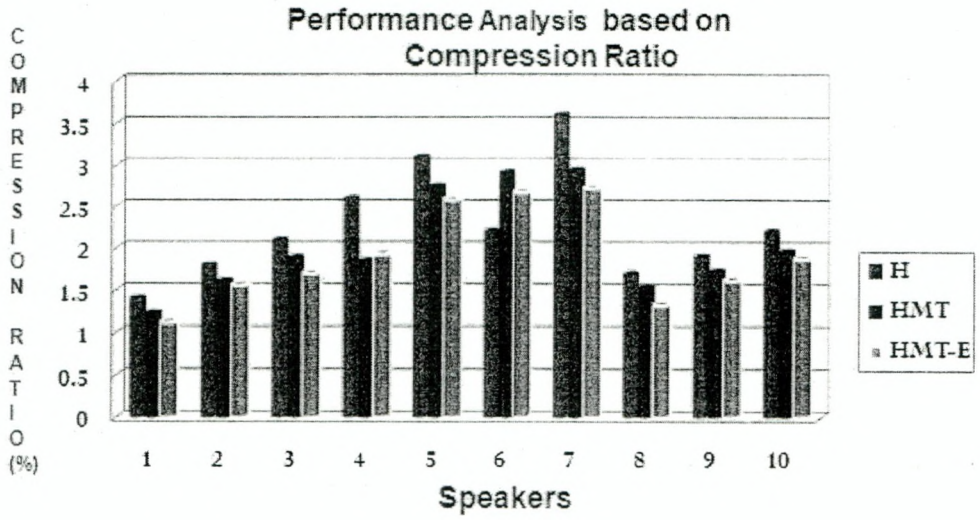

Figure 5 Performance Evaluation Based on Compression Ratio

Huffman Coding - H
Multiple Huffman Table Coding - HMT
HMT with Encryption - HMT-E

5 Conclusion

In order to provide more security for the speech communication, this paper focuses on enhancing Multiple Huffman Coding Table by incorporating an optimization procedure as a preprocessing step before compression. The compression part helps to reduce the size of the speech signal, in order to reduce the bandwidth requirement during transmission through mobile network. To provide protection, the signal is encrypted using AES which is practically impossible for attackers to recover any information. In general a good reconstructed signal should produce high PSNR and low NRMSE which means the signal have low error and high reliability. The results were analysed for various speakers with various methods. From the results it can be seen that the proposed method provides better result for all the speakers.

References:

1. Zhijun Wu," Information Hiding in Speech Signals for Secure Communication", Copyright@ 2015, Science Press Published by Elsevier Inc.
2. Akella Amarendra Babu et al ," Robust Speech Processing in EW Environment", International Journal of Computer Applications (0975 – 8887), Volume 38– No.11, January 2012.
3. D.Ambika and V.Radha,", "Vector Quantization in Language Independent Speaker Identification Using Mel-Frequency Cepstrum Co-Efficient", Lecture Notes in Electrical Engineering Volume 284, 2014, Pp 171-182.
4. D.Ambika, V.Radha," Secure Speech Communication – A Review", International Journal of Engineering Research and Applications, Vol. 2, Issue 5, September- October 2012, pp.1044-1049.
5. Hyubgun Lee, Kyounghwa Lee, and Yongtae Shin," AES Implementation and Performance Evaluation on 8-bit Microcontrollers", (IJCSIS) International Journal of Computer Science and Information Security, Vol. 6 No. 1, 2009.
6. D.Ambika and V.Radha," A Comparative Study between Discrete Wavelet Transform and Linear Predictive Coding", World Congress on Information and Communication Technologies,pp 966-970,2012.
7. C.P. Wu and C.C. J. K. Kuo, "Design of integrated multimedia compression and encryption systems", IEEE Transactions in Multimedia, vol. 7, no. 5, pp.28–839, 2005.
8. Shaimaa A. El-said et al," Securing Multimedia Transmission Using Optimized Multiple Huffman Tables Technique", International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 4, No. 1, March 2011.