International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Speech Encryption using Advanced Encryption Standard for Secured Communication

Sujiya Sreedharan, Chandra Eswaran

Abstract: ***Securing speech communication is a great challenging task and interesting research area in recent years. Transmission of voice data over untrusted networks leads to attack the confidential information. Various speech encryption algorithms are in progress and even though speech data have been encrypted some additional features are still getting compromised to the eavesdroppers which are a get setback for security attacks. The state-of-art of Advanced Encryption Standard (AES) algorithm when compared to other encryption algorithm it has different types of cipher key length such as 128,192 and 256 bits respectively. Advanced Encryption Standard is analyzed statistically for revealing its superiority based on confusion and diffusion properties for evaluating the robustness against statistical attacks. A novel approach is developed for encrypting speech data using Advanced Encryption Standard (AES). Encrypted and the decrypted signal are evaluated based on some quality analysis for immunity checking against frequency-domain attack, brute force attack and statistical attack etc.***

Keywords: ***Speech Processing, Encryption, Advanced Encryption Standard (AES), Quality Analysis Metrics***

I. INTRODUCTION

The invention of encryption has been around for more than 2000 years were theory of computation plays a major role in the field of cryptography research where the entire protocol relies on theories of algebra. Cryptography is mainly invented for secured digital communication system to overcome the consequences of impersonization by eavesdroppers. Encryption is a coding of encrypted writing or secret writing of sensitive data so that only legitimate users can access and understand to initiate a task. Encryption is breached out in various research areas were the introduction of voice encryption today had evolved drastically in recent years with effective replacement of old analog method of encryption by complex algorithms. Digital speech encryption is one of the countermeasures implemented in the communication channel which has a advancement in upcoming technology where security is the major concern for protecting sensitive data from unauthorized access[1][2]. Each individual voice is unique and differ based on tone, articulation, pronunciation, frequency, rhythm and pitch information.

Manuscript published on 30 September 2019
* Correspondence Author
Sujiya Sreedharan*, Department of Computer Science, Bharathiar University, Coimbatore, India. Email: suji.sreedharan.
Dr.Chandra Eswaran, Department of Computer Science, Bharathiar University, Coimbatore, India. Email: crcspeech@gmail.com, chandra.e@buc.edu.in

Obliviously the males voice sound ranges at 65 to 260 Hertz, while females voice sounds at a range of 100 to 525 Hz where each individual voice is unique based on the different characteristic of accent. Speech signal is represented as analog and digital form. In analog representation the speech waveform represents the frequency and amplitude of the speech signal where digital represents the numeric form of analog signal (i.e) zero's and one's.

During digital transmission the information transmitted over communication channel are easily hackable [3]. So in order to over the vulnerability security countermeasures like voice encryption are proposed in recent study in the area of voice processing. Since speech data are redundant in nature compared to written text speech encryption is quite a difficult task to provide security against various statistical attacks [10]. In the security point of view any voice communication is vulnerable by two categories namely the one who listens the conversation and second the person who eavesdrops by communication transmission [4]. Due to advancement of digital world secure speech communications have become an essential issue to upgrade the performance of various speech encryption algorithms implemented on Digital Signal Processor (DSP) platform. Encryption algorithms are especially rely on performance based on parameters such as speech, security against statistical attacks, CPU cycle and energy consumption and encryption/decryption speed[5][6].

II. ENCRYPTION ALGORITHMS

Cryptography is a field of study of secret writing of data into codeword which are not understandable by the hacker. Various algorithms are written in the field of cryptography for secured communication. In recent years encryption mechanism based research are focused towards speaker recognition for secure speech communication in the digital world. Modern field of cryptography is categorized into symmetric and non-symmetric algorithm based on key mechanism. Symmetric key algorithmic flow perform same key mechanism for encryption and decryption process of speech signal [7]. In recent years, various speech data encryption techniques has been developed and are listed out in the following table 1 with their features[10].

Features	Encryption Algorithms			
	DES	*Triple--DES*	*AES*	*Blowfish*
Block size	64	192	128,192,256	64
Key size	64	64	128	32-448
Rounds	16	16	10,12,14	16
Energy Consumption	Low	Highest	Medium	Lowest
Execution Speed	Slow	slowest	Medium	Fastest

Security	Crack ed	Not cracked but very slow	Not Cracked	Not Cracked
Encryption/ Decryption time	High	Highest	Moderate	Lowest
ThroughPut	Low	slowest	High	Highest

Figure 1: Advanced Encryption Standard algorithm representation

Joan Daemon and team first introduced AES algorithm where one key is used for both encryption and decryption mechanism. This algorithm is symmetric key cryptography where one key is used for both encryption and decryption. Among traditional encryption algorithm AES algorithm has 128,192 and 256 bit cipher key [8]. With speech signal S and with the key k as input, the encryption algorithm is as follows

$$CS=E_K(S) \quad (1)$$

The notation mentioned in equation (1) is the ciphered signal which is generated by using encryption algorithm E that applied on speech signal S with a specific function performed by key (K). The receiver invert the transformation by the possession of the key (K) and retrieves back the original signal for authentication mechanism

The decryption process is performed as follows

$$S=D_K(CS) \quad (2)$$

The popular Advanced Encryption Standard is an iterative process consists of different computational rounds for both encryption and decryption of speech signal. As AES encryption mechanistic as different key size the computational round counts vary based on the key size as mentioned in table 2.

III. ALGORITHM OF ADVANCED ENCRYPTION STANDARD:

AES algorithm performs four stage of process to complete one round process which is iterated 10 times for 128 bit, 12 rounds for 192 bit and 14 times for 256 bit length key mentioned in figure 2

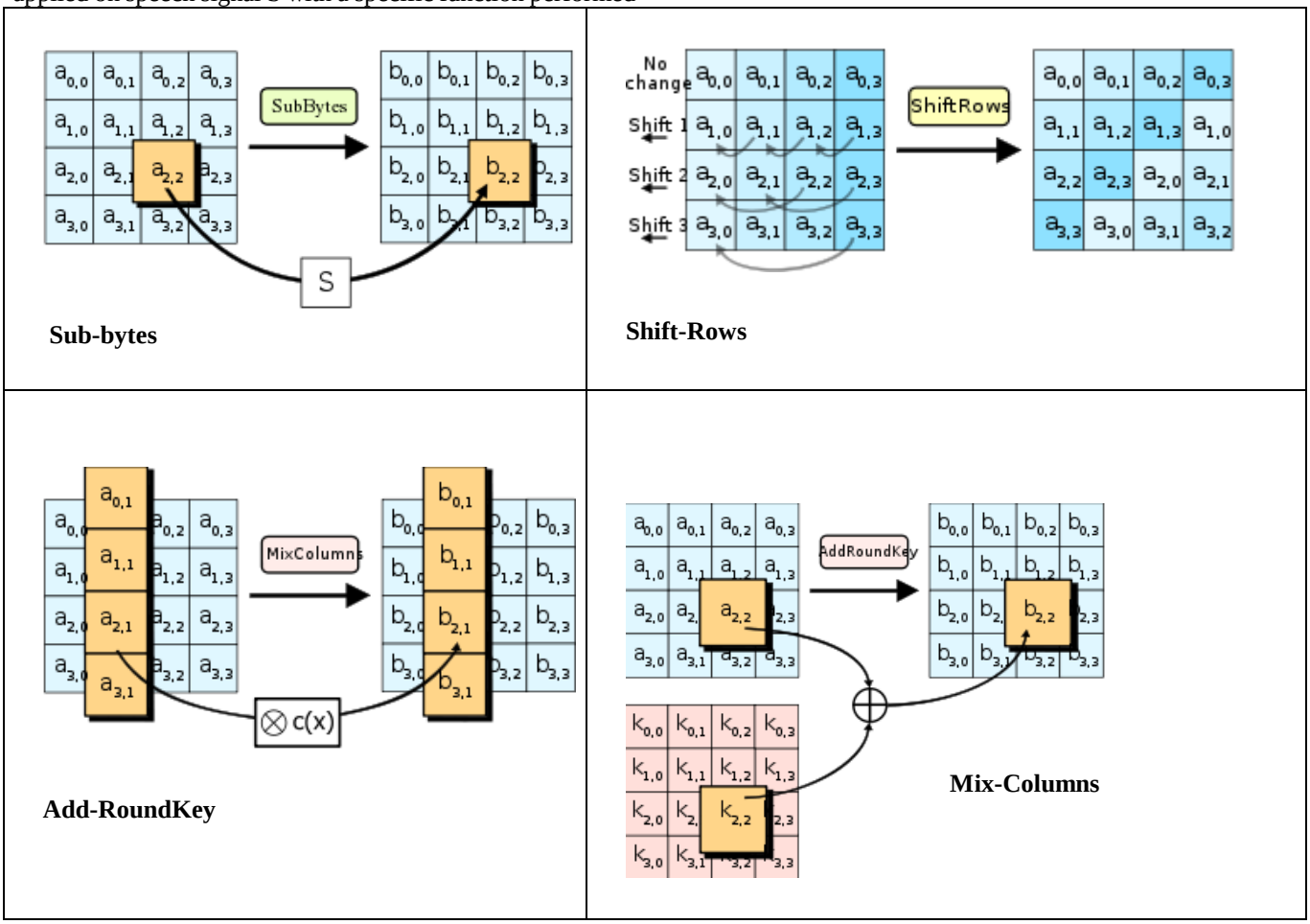
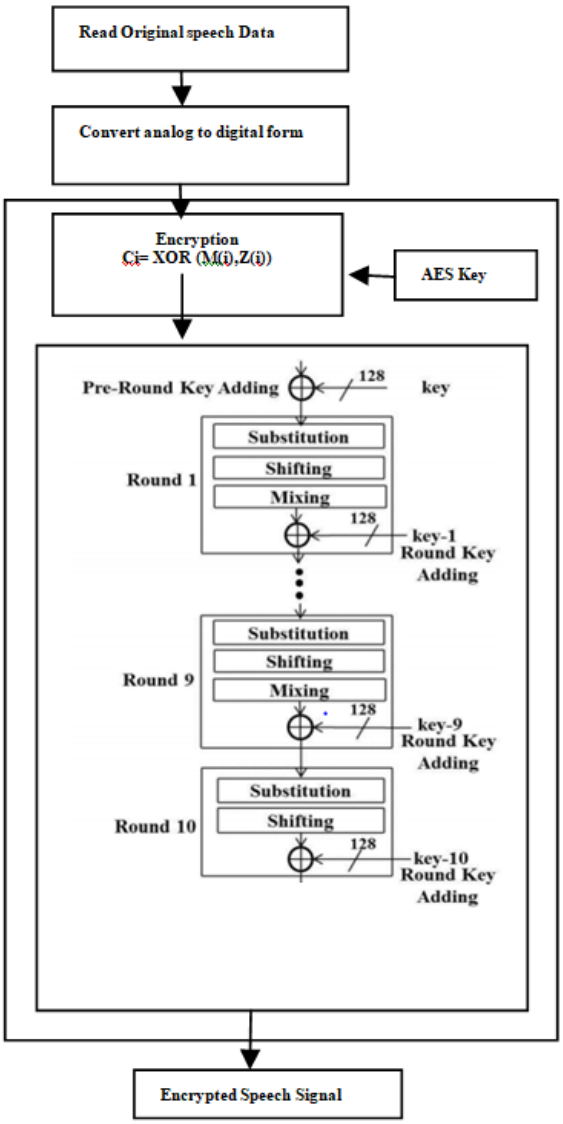


Figure 2: Algorithm flow of AES algorithm

1. **Sub-Bytes**: Transformation is a non-linear byte substitution for each byte of the block.
2. **Shift-Rows**: Transformation cyclically shifts (permutes) the bytes within the block
3. **Add-Round Key:** Transformation groups 4-bytes together forming 4-term polynomials and multiplies the polynomials with a fixed polynomial mod (x^4+1).
4. **Mix-Columns:** Transformation adds the round key with the block of data

IV. PROPOSED METHODOLOGY

A. Speech Encryption Algorithm

Pseudo code	Methodology flow of Speech Encryption
Input : Input speech signal **Output :Encrypt Speech signal** **Algorithm Flow:** • Read Original Speech Signal • Convert analog Speech to digital form • Digital form is divided into n no of blocks and stored in file M • T ← the length of M • Perform AES operation on the divided blocks • Generate Key Z_i by key generator algorithm • For i=1 to T make • Encrypt speech signal using key Z_i using the equation as follows $C_i = M_i \oplus Z_i$ and store in file C • End Process	

Input : Input Encrypted signal

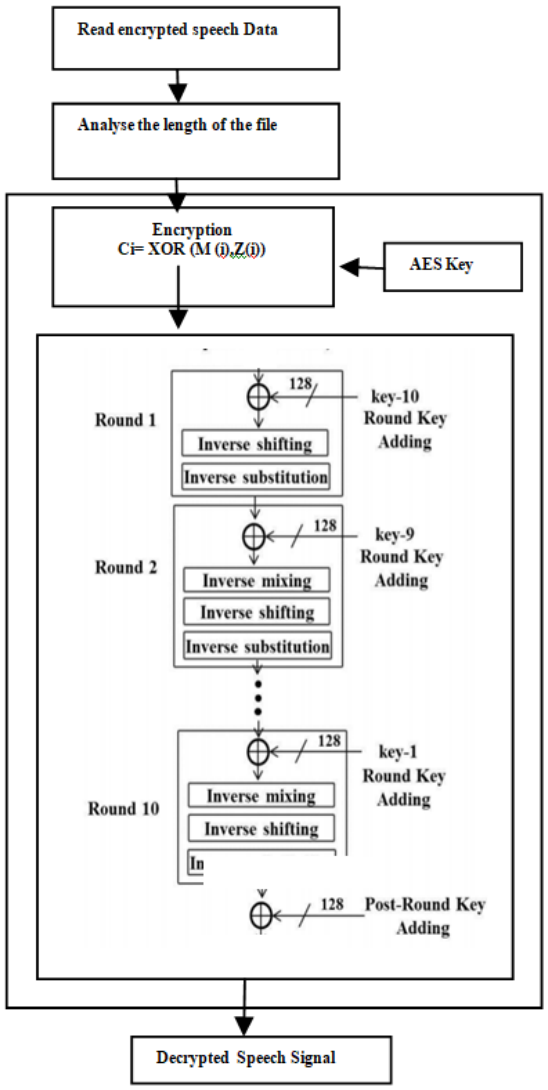
Output :Decrypted original Speech signal

Algorithm Flow:

- **Read the encrypted speech file C**
- **T ← the length of encrypted file**
- **Generate AES Key Z_i by key generator algorithm**
- **For i=1 to T make**
- **Decrypt the Speech signal using key Z_i using the equation as follows**

  $M_i = C_i \oplus Z_i$

- **Get the original speech data**
- **End Process**



B. Algorithm Implementation

The 128-bit speech signal is taken and the analog signal is converted to digital form and further processed with pre-round transformation. The basic pre-round operation is performed initially by XOR operation between the 128-bit input speech signal and 128-bit cipher key for encryption. The XOR-ed signal is passed on with further transformation rounds which has four distinct transformation namely (1) Substitution (2) Shifting (3) Mixing and (4) ad-round key. These transformation are performed on 128-bit speech data block which is clearly depicted in the above table 3

C. Substitution

Substitution operation is performed on the XOR-ed speech signal block independently on 128-bit speech block which has 16 separate byte-to-byte transformation. Substitution operation are performed on the pre-computed values saved in lookup table

D. Shifting

Shifting is followed by substitution transformation. Permutation function is processed to the bytes of the speech signal block in which the sequential order of the bytes are interchanged without the bits of the order are unchanged.

E. Mixing

Bitwise XOR operation is performed on the neighborhood bytes of the 128-bit speech data. The last round of the block is kept unchanged without performing mixing operation.

F. Add round key

The final round of AES is the important round in AES operation where the key expansion process is performed by Bit-wise XOR operation between the generated round key on the 128-bit output of the results obtained by the mixing operation performed previously.

G. Decryption process

Once the operation of input speech signal along with the cipher key is completed the encrypted speech signal is given as an input to the decryption module to invert it to original speech signal at the receiving end. Add round key transformation is the initial round of the decryption process. There is no inverse mixing operation in the first round of round transformation. Inverse mixing and inverse shifting operation are performed before inverse substitution Transformation. The lookup table values generated during encryption differs from the lookup table generated during decryption process. After completion of 10 rounds process of Decryption post-round transformation is performed as the final stage of operation. Finally bit-wise XOR operation is performed between cipher key and the tenth round processed output to achieve the original speech signal.

V. RESULTS AND SUMMARY

The proposed system is implemented using Matlab 2014a with four different speech samples taken randomly from TIMIT database recorded at 8 KHz which has taken 2 seconds of speech recording with 8000 samples per frame. The quality evaluation is tested with correlation test, Spectral Distortion, Signal to Noise Ratio (SNR), PSNR test, robustness test and randomness test to check the robustness of the system against various attacks.

EXPERIMENTAL RESULTS

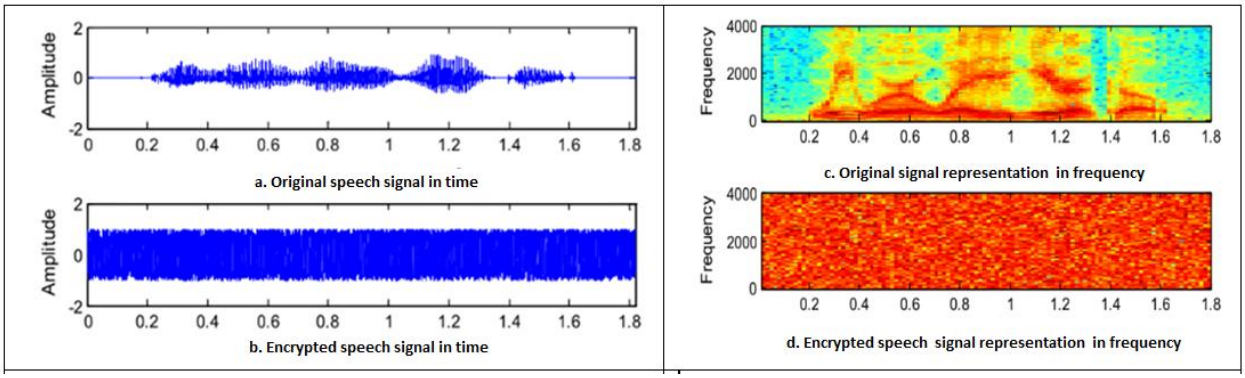


Table 2: Simulation results of original and encrypted speech signal in time and frequency

Figure a and c shows a clear speech signal without encryption based on time and frequency where figure b and d shows the encrypted speech signal based on time and frequency in which the speech signal presence are unable to predict where it seems like noise presence where no clues of residual intelligibility for the eavesdroppers to hack the actual conversation.

A. Quality measures of speech signal

Quality of encrypted speech signal is analyzed by large key space which should be infeasible to various attack infeasible. An attacker tries combination of keys and will be exhaustive in large number of combinations.

$$\frac{2^{128}}{365 \times 24 \times 60 \times 60 \times 1000 \times 10^6 \times 10^{21}\ \text{Years}} > 10.79 \quad (3)$$

B. Signal to Noise Ratio (SNR)

Signal to Noise ratio is calculated on encrypted speech signal for measuring the intelligibility of the speech signal.

$$\text{SNR} = 10\log_{10}\frac{\sum_{i=1}^{N_s} x^2(i)}{\sum_{i=1}^{N_s}(x(i)-y(i))^2} \quad (4)$$

If the value of the SNR values is near to zero the quality of decrypted signal is considered to be good where the original signal is be recovered overall from the decrypted process

C. Peak signal to Noise Ratio

Analysis between desired signals to background noise. Presence of noise persist more in an encrypted signal which has more negative PSNR value which indicates that the signal is more secured from eavesdropping attack.

$$\text{PSNR} = 10\log\frac{nx^2}{\|x-k\|^2} \quad (5)$$

x= maximum absolute square value of the original speech signal;
l= Length of the encrypted signal
x-k = energy level between encrypted signal and original speech signal

D. Correlation coefficient

Evaluation measure between the similarity score level between original and encrypted and measure of decrypted and original signal. It is calculated as follows

$$r_{xy} = \frac{c_v(x,y)}{\sqrt{D(x)}\sqrt{D(y)}} \quad (6)$$

Where $c_v(x,y)$ is the covariance between original signal denoted as x and encrypted signal denoted as y. D(x) and D(y) is the variance of the x and y. the numerical formulation is formulated as follows

$$E(x) = \frac{1}{N_s}\sum_{i=1}^{N_s} x(i)$$
$$D(x) = \frac{1}{N_s}\sum_{i=1}^{N_s} (x(i) - E(x))^2$$
$$c_v(x,y) = \frac{1}{N_s}\sum_{i=1}^{N_s} (x(i) - E(x))(y(i) - E(y)) \quad (7)$$

N_s = number of speech signal

From the analysis evaluation of correlation coefficient it is considered that the low value of r_{xy} denotes good encryption quality.

E. Spectral Distortion

Frequency domain evaluation between the implementation between the original and processed signal is spectral distortion. It is basically measured in dB between the distance between the processed and original signal. The spectral distortion is calculated as follows:

$$SD = \frac{1}{M}\sum_{m=0}^{M-1}\sum_{i=Nm}^{Nm+N-1} |V_x(i) - V_y(i)| \quad (8)$$

$V_x(i)$- spectrum of the original speech signal ;$V_y(i)$ – Spectrum of the distorted signal; N – segment length of M In speech signal

F. Experimental results and discussion

File name	SNR	PSNR
Speaker1.wav	-23.89dB	-24.296
Speaker2. wav	-21.79dB	-28.128
Speaker3. wav	-24.01dB	-28.235
Speaker4. wav	-22.28dB	-28.264

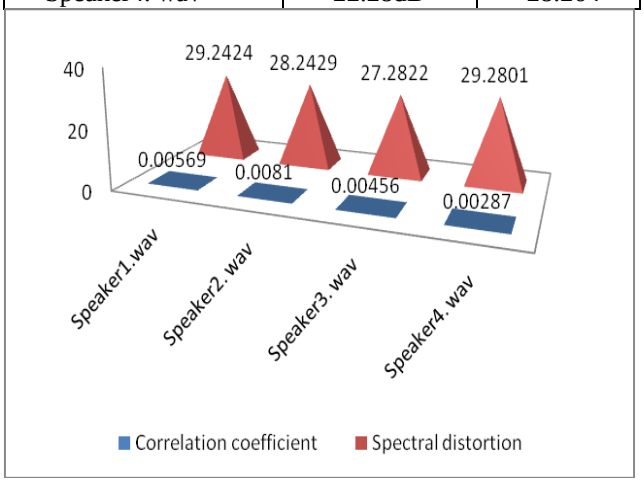


Table 3: SNR and PSNR values of encrypted signal

The measure of Residual Intelligibility between the encrypted speech signal and the decrypted speech signal is the measure of Signal to Noise Ratio (SNR). Generally low valued SNR value indicates high level of Noise presence whereas high SNR values indicate good quality of encryption and decryption quality with less occurrence of noise presence. The SNR [1, 2] is calculated using PSNR compares the level of desired signal to the level of background noise. Encrypted signal's sound is very noisy, having a negative PSNR. [9] So the encrypted signal is more secure then original signal.

File name	Correlation coefficient	Spectral distortion
Speaker1.wav	0.00569	29.2424
Speaker2. wav	0.00810	28.2429
Speaker3. wav	0.00456	27.2822
Speaker4. wav	0.00287	29.2801

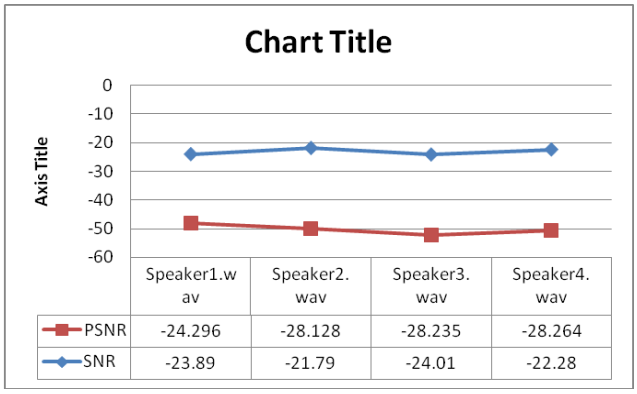


Table 4 : Correlation coefficient and Spectral distortion values of encrypted signal

G. Correlation Coefficient Analysis.

Correlation coefficient (CC) is one of the statistical measures which determine the encryption quality of the cryptosystem. This analysis measures the correlation between the two speech samples whose value lies between −1 and +1. The correlation coefficient being near zero indicates the weakest relationship between the two samples and it is not possible to predict the secret key by the attackers [1]. The correlation coefficient values between original and encrypted speech signals and their comparison with other algorithms are tabulated in Table 4. Figure 6 shows the correlation coefficient distribution of original and encrypted speech signal. It has been observed that the correlation values are closer to zero which indicates the good encryption quality.

VI. CONCLUSION

The proposed work is focused on protecting speech information of a speaker. The proposed system is an encryption and authentication mechanism to increase the immunity against various attacks like eavesdropping, Brute-force attack and statistical attack etc. The system is used to increase the immunity using AES with 128 bit key for data protection and authentication of identity. Various quality metrics like SNR, PSNR, Correlation Coefficient and Spectral Distortion are performed to evaluate the robustness of the proposed system.

ACKNOWLEDGMENT

I am grateful to all kinds of support provided by Prof. Dr. E. Chandra Eswaran for guiding me for my research work. Thanks are also extended to all the higher authorities of Bharathiar University for giving me opportunity for doing my research work.

Funding: The research work is supported by Department of Science and Technology-PURSE (Phase – II).This work has been submitted for Indian Intellectual property with Patent Application Number 201841032393

REFERENCES

1. D. Reynolds, "An overview of automatic speaker recognition technology," in Proc. IEEE Int. Conf. Acoustics Speech Signal Processing (ICASSP), vol. 4, pp. 4072–4075,2002.
2. H. Gish and M. Schmidt, "Text-independent speaker identification," IEEE Signal Processing Mag., vol. 11, no. 4, pp. 18–32, 1994.
3. M. Senoussaoui, P. Kenny, N. Dehak, and P. Dumouchel, "An i-vector extractor suitable for speaker recognition with both microphone and telephone speech," in Proc. IEEE Odyssey, Brno, Czech Republic, 2010.
4. Luanlan," The AES Encryption and Decryption Realization Based on FPGA," In Seventh international conference on computational intelligence and security (CIS) pp. 603–607, 2010.
5. Evans, N. W. D., & Mason, J. S. D, "An assessment on the fundamental limitations of spectral subtraction," In IEEE international conference on Acoustic, speech and signal processing pp. 1–1, 2006.
6. P. Kenny, P. Ouellet, N. Dehak, V. Gupta, and P. Dumouchel, "A study of inter-speaker variability in speaker verification," IEEE Trans. Audio, Speech, Lanġ. Process. vol. 16, no. 5, pp. 980–988, Jul. 2008.
7. Emad Mossa, "Security enhancement for AES encrypted speech in communications," International Journal of Speech Technology, International Journal of Speech Technology archive Volume 20 Issue 1, pp. 163-169, March 2017.
8. L.A. Khan a , M.S. Baig b , Amr M. Youssef, Speaker recognition from encrypted VoIP communications, Journal of Digital Investigation: The International Journal of Digital Forensics & Incident Response archive, Volume 7 Issue 1-2, pp: 65-73, October 2010 .
9. Manas A. Pathak and Bhiksha Raj, Privacy-Preserving Speaker Verification and Identification Using Gaussian Mixture Models, IEEE Transactions On Audio, Speech, And Language Processing, Vol. 21, No. 2,pp. 397-406, February 2013
10. T. El-Gamal, "A public-key cryptosystem and a signature scheme based on discrete logarithms," IEEE Trans. Inf. Theory, vol. IT-31, no. 4, pp. 469–472, Jul. 1985
11. T. Bianchi, A. Piva, and M. Barni, "Composite signal representation for fast and storage-efficient processing of encrypted signals," IEEE Trans. Inf. Forensics Security, vol. 5, no. 1, pp. 180–187, March 2010.

AUTHORS PROFILE

S.Sujiya is a Ph.D. Research Scholar at Bharathiar University Coimbatore. She received her BCA Degree in Providence College for women, niligires in 2008, and MCA degree from Avinashilingam University, Coimbatore in 2011. She completed her MPhil Degree in SNS College of Arts and Science, Coimbatore in 2015. Her research interests include Speech and speaker recognition.

E. Chandra Professor and Head in the Department of Computer Science, Bharathiar University, Coimbatore, Tamil Nadu, India. She has more than 24 years of teaching experience and 20 years of Research Experience. Her Area of Specialization includes Neural Networks, Speech Recognition System. She has produced 7 Ph.D scholars. She has authored more than 77 papers published in refereed International journals, presented 46 papers in National and International Conferences and published 2 books. She has obtained funding projects from UGC in the field of speech signal processing. She is an active member of CSI, Life member of Society of Statistics and Computer Applications, Senior Member of ACM with Digital Library, Life Member of International Association of Engineers, Life Member of International Association of Computer Science and Information Technology. She is the Board of Studies Member for various affiliated Institutions, Member of various Professional Societies, Inspection Commission member for various colleges and selection committee member of different institutions. She was a Coordinator for RUSA scheme from Bharathiar University and Reviewer for International Journals.