International Journal of Speech Technology (2018) 21:319–332
https://doi.org/10.1007/s10772-018-9510-0

A multi-tier security system (SAIL) for protecting audio signals from malicious exploits

N. Sasikaladevi[1] · K. Geetha[1] · K. N. Venkata Srinivas[1]

Received: 2 November 2017 / Accepted: 9 April 2018 / Published online: 13 April 2018

Abstract
This paper proposes a multi-tier SegmEntation ECC Desegmentation (SEED) model to suit audio cryptosystem for Securing Audio sIgnal (SAIL) based on discrete wavelet transform and elliptic curve encryption. It is aimed with the prospect of enhancing the level of security in digital audio communication for unreliable public networks. The proposed SAIL system works as a multitier SEED model by performing segmentation, DWT compression, ECC encryption and desegmentation. In the reverse process, this multitier model proceeds with segmentation, decryption, decompression, and desegmentation. The novelty of this work relies on the adoption of ECC for encryption as it is first of its kind in audio streaming. The selection of appropriate ECC curve is a real challenge, and complex multiplication method has been applied. ECC has been chosen for encryption as it has been identified as a discrete logarithm problem which is resistant to be attacked by quantum computers. The performance of the recommended SAIL cryptosystem has been tested using different audio samples characterizing human voice, animal voice and Instrumental music. Analysis of the proposed model shows the effectiveness for fast audio encryption as it works on compressed data and also computationally simple. Various statistical analysis have been done on the proposed model, and the obtained result ratifies better level protection of audio signals from different security threats and can be recommended for multi channel audio processing.

Keywords Discrete wavelet transform · Elliptic curve cryptography · Audio encryption

1 Introduction

The rapid growth of the Internet and its application drastically increases the need for securing the transmission of multimedia data in the public network. The Internet is used in most of the entire domain including education, government, military, banking, and commerce, etc. Growing popularity of the Internet and use of electronic gadgets increases the transmission of multimedia data in the network. Mobile phones are widely used for transmission of audio, images, video and text data over the Internet. The public network is open to all. It is vulnerable to attacks by hackers and intruders. Protecting the multimedia data against unauthorized access is the demand today. There is a need to provide confidentiality and integrity for multimedia content. Different techniques have been proposed to secure the multimedia content, and it includes encryption, watermarking and steganography. Protection of multimedia content is different from regular text data. In specific, most of the traffic generated in the mobile is audio traffic. Protecting the audio signal against the attackers is the fundamental requirement today. Enforcing confidentiality against the audio signal is needed now.

Predominantly used information representation structure today is the audio signals, which are widely used by the modern community for different types of the communication. In recent days, the secret sharing in the form of audio is happening publicly. Audio is accepted as the evidence in the court cases. Digital audio needs to be protected against unauthorized access. But audio signals are entirely different types of signals as compared to text and images. Audio signals are represented as wave signals, and it has been

✉ N. Sasikaladevi
sasikalade@gmail.com

K. Geetha
geethavalavan@gmail.com

K. N. Venkata Srinivas
knvs97@gmail.com

[1] Department of Computer Science and Engineering, School of Computing, SASTRA Deemed University, Thanjavur, TN, India

characterized by various metrics such as frequency, amplitude, and phase. Most of the existing cryptographic algorithms are best suitable for text data. It cannot be used for audio signals directly due to its representation and in specific audio signals are high volume and highly redundant data. Hence, efficient cryptographic algorithms are required to secure the sensitive audio signals before transmitting the signal over the public network in specific Internet and mobile network. Designing the audio encryption algorithm is the challenging task today.

Recently researchers have studied this problem, and they proposed different kinds of the algorithms for protecting audio against the unauthorized access (Al Saad and Hato 2014; Li et al. 2009; Kohad et al. 2012; Sharma 2012; Zeng et al. 2012; Sheu 2011; Elshamy et al. 2013; Zhao et al. 2014; Mermoul and Belouchrani 2010; Al-Karim et al. 2013). Several image and video-based encryption, watermarking and stenography algorithms are available in the research (Zhang et al. 2008; Lin and Chang 2001; Petitcolas et al. 1999; Langelaar et al. 2000; Chen and Lin 2003; Barni et al. 2001; Refregier and Javidi 1995; Hedelin et al. 1999; Yang et al. 1998; Kim et al. 2004; Kwon et al. 2006; Wu and Ng 2002; Wang and Fan 2010), but, the audio protection methods are relatively very low. In the audio scrambling techniques, the audio signal is rearranged to remove the correlation between the audio samples. Most of the audio scrambling techniques are based on 1D linear mapping (Zeng et al. 2012). But, this types of algorithms are vulnerable to attack. Because, the audio signal has small a variation concerning time, the adjacent samples have similar signals. Therefore, the audio encryption is one of the challenging tasks and limited techniques are proposed to protect the audio (Al Saad and Hato 2014; Li et al. 2009; Kohad et al. 2012; Sharma 2012; Zeng et al. 2012; Sheu 2011; Elshamy et al. 2013; Zhao et al. 2014; Mermoul and Belouchrani 2010; Al-Karim et al. 2013). This problem is addressed in this paper. This paper proposes the efficient audio encryption scheme to provide confidentiality for the sensitive audio signals.

The rest of the paper is arranged as follows. The subsequent section illustrates the existing models for audio encryptions. In Sect. 3, the proposed SEED encrypt algorithm and SEED decrypt algorithm is described with the diagram. Section 4 is structured for the experimental result, performance analysis, and security analysis. Security analysis illustrates that the proposed algorithms are highly sensitive to a minor alteration of the keys. Statistical analysis demonstrates that mean square error (MSE), peak signal to noise ratio (PSNR), correlation analysis and Histogram analysis. And it is proved that the proposed algorithms resist all statistical attacks. Experimental results express the usefulness of the audio encryption scheme. A brief conclusion is given in Sect. 5.

1.1 Motivation

The secret spy the microphone prevails everywhere and paves the way to hackers. Hackers find their entry by remote access trojans into government and corporate sectors. They can acquire audio information through these microphones and will transmit as compressed audio files via email, for illegal uses. Many computing systems can be compromised if its audio and microphone channels are not physically partitioned. Risks will explode exponentially in Voice over Internet Protocol (VoIP) phone systems.

Intruders can have remote access to microphones and can easily escape from security software, and their activity cannot be trapped. Usage of memory buffers and other types of storage devices can still raise the danger of misuse. Malicious software can easily manipulate these technologies when users switch between systems. Since many VoIP networks transmit data between networks of different security policies. That increases the danger of electromagnetic interference leakage. As the switch logic in firmware is reprogrammable, it can be tampered with, and hence difficult to identify whether it has been used or compromised. But a new variety of innovation is budding to avoid malicious audio signal interference. By maintaining audio signals physically divided from the microphone or speaker signals, the likelihood of leakage between signals on either side can be removed. As a result, organizations can avoid signals from being oppressed or manipulated by malicious software, thereby preserving the integrity of the signal when users switch between computers. Further, the use of a microphone mute button, which can physically control microphone when not in method, and cannot be manipulated by software or drivers, thereby assuring extend system security.

2 Related works

The audio encryption based on one and two-dimensional discrete time a chaotic system was proposed by Akgül and Kaçar (2015). In this model, the audio samples of type both mono and stereo are scrambled, and security of the algorithm is increased by non-linear models. Sadkhan and Mohammed (2015) proposed the pseudo-random bit generator for the audio encryption, which is based on the chaotic map. Tamimi and Abdalla (2014) demonstrated a scrambling process to protect audio with traditional block cipher algorithms; the secret key was designed in such a manner that it is audio signal dependent and the public key reliant. Lima and Silva Neto (2016) proposed an audio scrambling method by using cosine number transform (CNT), CNT

is structured based on finite fields, and is repeatedly pertained to range of audio sequences of raw uncompressed data, the blocks are preferred using a overlying rule, that yields confusion and diffusion in the encrypted data of different blocks of audio signals.

Ciptasari et al. (2014) demonstrated the encryption techniques by the hybrid combination of the Discrete Wavelet Transform (DWT) and Discrete Cosine Transform (DCT) to design the resilient of the audio. It is used to provide the visual cryptography and time stamping and watermarking on the digital data. In this paper, watermarking is not embedded in the plain audio, it is utilized to create the secret image and a public image that is used to protect the audio signal. In the visual cryptographic techniques, chaotic maps like Chebyshev map (Liu and Wang 2010), Tent map, and disordered schemes, like Lorenz system (Anees 2015) and Chen scheme (Tong et al. 2015), are regularly used to create the random sequence such as key flow, for minor modification of one of the premier parameters can lead to utterly dissimilar path. Augustine et al. (2015) proposed an audio scrambling technique based on compressive sensing (CS) and Arnold transform (AT). The scrambling and compressive sensing are carried out by means of a key-based depth matrix, and the encryption is performed by the use of an Arnold matrix in which the first condition is created by using a Piecewise Linear Chaotic map (PWLCM). Audio encryption algorithms proposed to handle outmoded, and strong audio signals are the chaos-based and double random phase encoding (DRPE) methods (Al-Karim et al. 2013).

A chaotic map based audio encryption algorithm is proposed by (Eldin et al. 2015; Elkholy et al. 2015; Mostafa et al. 2015; Alwahbani and Bashier 2013). An audio encryption based on LFSR is proposed in James et al. (2014). Dengre and Gawande (2015) proposed an audio encryption for uncompressed data. Selective audio data encryption for multimodal surveillance system is proposed in Cichowski and Czyzewski (2012). Datta and Gupta (2013) proposed a fractional encryption and watermarking methods for audio signals with the reduction of quality. Rashidi and Rashidi (2013) proposed an FPGA based AES encryption algorithm for an audio signal. Voice authentication and real-time audio encryption are proposed in Nguyen et al. (2013). Kulkarni and Patil (2015) proposed a strong encryption technique for audio data hiding in digital images for better security. Ashok et al. (2013) proposed a secure cryptographic scheme for audio signals. Iyer et al. (2016) proposed a multimedia encryption based on hybrid approach. Context-aware multimedia encryption is proposed in Fazeen and Bajwa (2014). Washio and Watanabe (2014) proposed an audio secret sharing scheme. Zhao et al. (2014) proposed a dual key speech encryption algorithm based underdetermined BSS. Scrambling based speech encryption via compressed sensing is proposed in Zeng et al. (2012). Lu et al. (2012) demonstrated an audio data hiding based on AT and double random phase encoding methods.

In modern society, numerous secret commercial talks need to be protected. In many real-time situations, digital audio needs to be protected from malicious exploits, and this alertness of privacy protection provokes the rapid development of protection mechanism. Audio encryption has invited a great deal of interest from researchers.

Audio is considered as one of the essential representation types; it has been broadly used in present society. In some cases such as sensitive business conversation, an audio proof is acceptable in court. Hence, the digital audio need be concealed as secret information. In specific, more and more consciousness of individual privacy protection triggers the instant design of audio encryption techniques. Hence, audio encryption has gained a great deal of attention from researchers.

3 Proposed multi-tier seed model

Figure 1 depicts the multi-tier SEED model for the proposed SAIL cryptosystem. Various activities carried out in each tier during the encryption phase referred as the forward process, and with that of the decryption, phase referred as the reverse process is shown as four tiers. The input audio signal is digitized by performing analog to digital conversion. In the first tier, the input audio signal is segmented and then compressed in the second tier by applying discrete wavelet transformation (DWT) finally; the compressed audio signal is encrypted using ECC in the third tier. The final tier performs desegmentation to construct the digital audio information.

DWT has been preferred over alternative transformations for the following reasons:

1. It can offer best audio quality than DCT with increased compression ratio.
2. DWT performs compression for the whole file rather than block by block, and hence the compression errors will be distributed across the entire file.

ECC has been preferred over RSA for the following reasons:

1. It extends the same level of security with a just 160-bit key size equivalent to a 1024 bit key size required for RSA as per the recommendation of National Institute of Standards and Techniques (NIST), and key generation is also faster in ECC.
2. It is not vulnerable to timing attack as that of RSA.

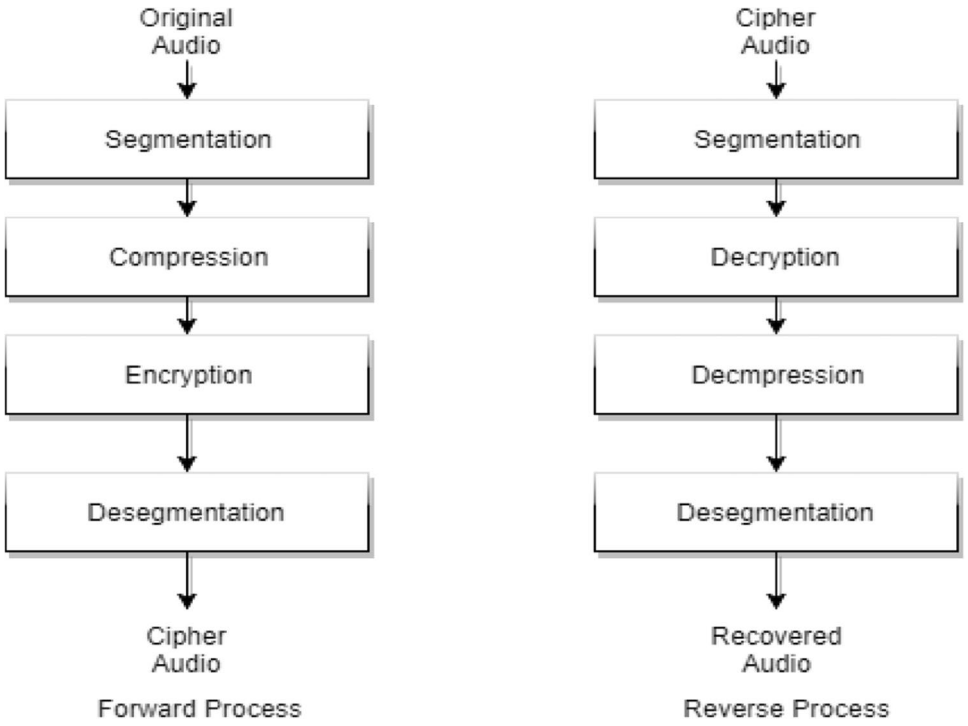

Fig. 1 Multi-tier SEED model for SAIL cryptosystem

3. Brute force attack and Pollard's who attack are computationally expensive or infeasible as it involves exponential running time.
4. Computational complexity and overhead are very minimal in ECC when compared with RSA as the former is based on additive group whereas the latter belongs to a multiplicative group.
5. ECC involves point operation which is less complicated than exponentiation operation performed in RSA.
6. More suitable for power constrained devices as it requires less computing power.

The novelty of the proposed SEED system entirely relies on the selection of appropriate Elliptic curve over prime field.

3.1 Elliptic curve cryptography

Elliptic curve cryptography (ECC) is an asymmetric cryptosystem standardized by IEEE P1363. It offers the equal level of security offered by Rivest Shamir and Adleman (RSA) but with lesser key size. Hence it reduces the processing overhead. Elliptic curve is based on the Weierstrass equation of the form (1)

$$y^2 + axy + by = x^3 + cx^2 + dx + e \tag{1}$$

where a, b, c, d, and e are real numbers and x and y take n values in the real numbers. Simplified form of the Eq. (1) is,

$$y^2 = x^3 + ax + b \tag{2}$$

Equation (2) is the cubic equation of degree 3 where a and b are coefficients, and x and y are variables. An elliptic curve over finite fields uses either prime curve or binary curve. The prime curve is based on GF(p), the coefficients and values took n values in the set of integers from 0 to p − 1 and represented as $E_p(a,b)$. The binary curve is based on $GF(p^m)$, the variables and coefficients of the cubic equation take values in $GF(p^m)$, and it is represented as $E_p^m(a,b)$.

3.1.1 Arithmetic operations on Ep(a,b)

$P + 0 = P$ where, $Q \epsilon E_p(a,b)$

If $P = (x_p, y_p)$ then $-P = (x_p, -y_p)$

If $P = (x_p, y_p)$ and $Q = (x_Q, y_Q)$ with $P \neq -Q$ then $R = P + Q = (x_R, y_R)$ is based on the formula given below,

$$x_R = \left(\lambda^2 - xp - xq\right) mod\ p \tag{3}$$

$$y_R = (\lambda(xp - xR) - yp)\, mod\, p \tag{4}$$

$$\text{where, } \lambda = \begin{cases} \frac{y_Q - y_p}{x_q - x_p}\ mod\ p\ , if\ P \neq Q \\ \frac{3x^2 p + a}{2yp}\ mod\ p,\ if\ P = Q \end{cases}$$

Elliptic curve encryption and decryption need a point G and an elliptic curve $E_p(a,b)$. B selects a private key d and determines the public key as $P_A = d^* G$. B transmits the pair (G, P_A) to A. A selects the message P_m and to select the secret key r. Then he encrypts the P_m as follows,

$$C1 = r * G,\ C2 = P_m + r P_A \tag{5}$$

The pair (C1, C2) transmitted across the network. B decrypts the message as follows,

$$P_m = C2 - d * C1 \tag{6}$$

3.2 Proposed SEED encryption algorithm

The SEED encryption algorithm is given in Algorithm 1 and shown in Fig. 2. The original audio is digitized and segmented with the segment size of 8 bits. The first 44 bytes represent the wave description and the remaining deals with the scanned payload data. Again it is segmented into 8 bits data. The audio payload is compressed by the ID DWT. Elliptic curve encryption is applied on each pair of 8 bytes of data. Encrypted data is prefixed with 44 bytes of wave description to form the cipher audio signal. The result of the SAIL cryptosystem is the compressed and encoded audio signals. This data is transmitted across the network.

Elliptic curve selection for audio encryption based on complex multiplication (CM) method

1. Given prime number p, estimate the minimum Determinant D with torsion value t based on Eq. (7).

$$4p = t^2 + Ds^2, \text{ where } t.s \in Z \tag{7}$$

$$\#E(F^p) = p + 1 - t, \text{ where } |t| \le 2\sqrt{p} \tag{8}$$

2. Check if the order of $E(F^p)$ has admissible factorization. Otherwise choose different D and t. Dot step 2 until an order with acceptable factorization is found.
3. Create the class polynomial $H_D(x)$.
4. Find the root j_0 of $H_D(x)$, where j_0 is the j-invariant of the curve.
5. Set $k = {j_0}/{(1728 - j_0)} \ (mod\ p)$ and the curve
$$E(F^p) : y^2 = x^3 + 3kx + 2k \tag{9}$$
6. Verify the order of the curve. If it is not equal to $p + 1 - t$, then create the twist using randomly chosen nonsquare $C \in F_p$.

In this audio encryption algorithm, every 8 bytes form one point with 4-byte coordinates. To cover all the points 32 bit largest prime number is used $p = 2^{32} - 1 = 214748364$ .CM method is applied to compute the elliptic curve over $F_{214748364}$, the constructed curve is,

$$E\left(F_{214748364}\right) : y^2 = x^3 + 390064447, \ where\ a = 0,\ b = 390064447 \tag{10}$$

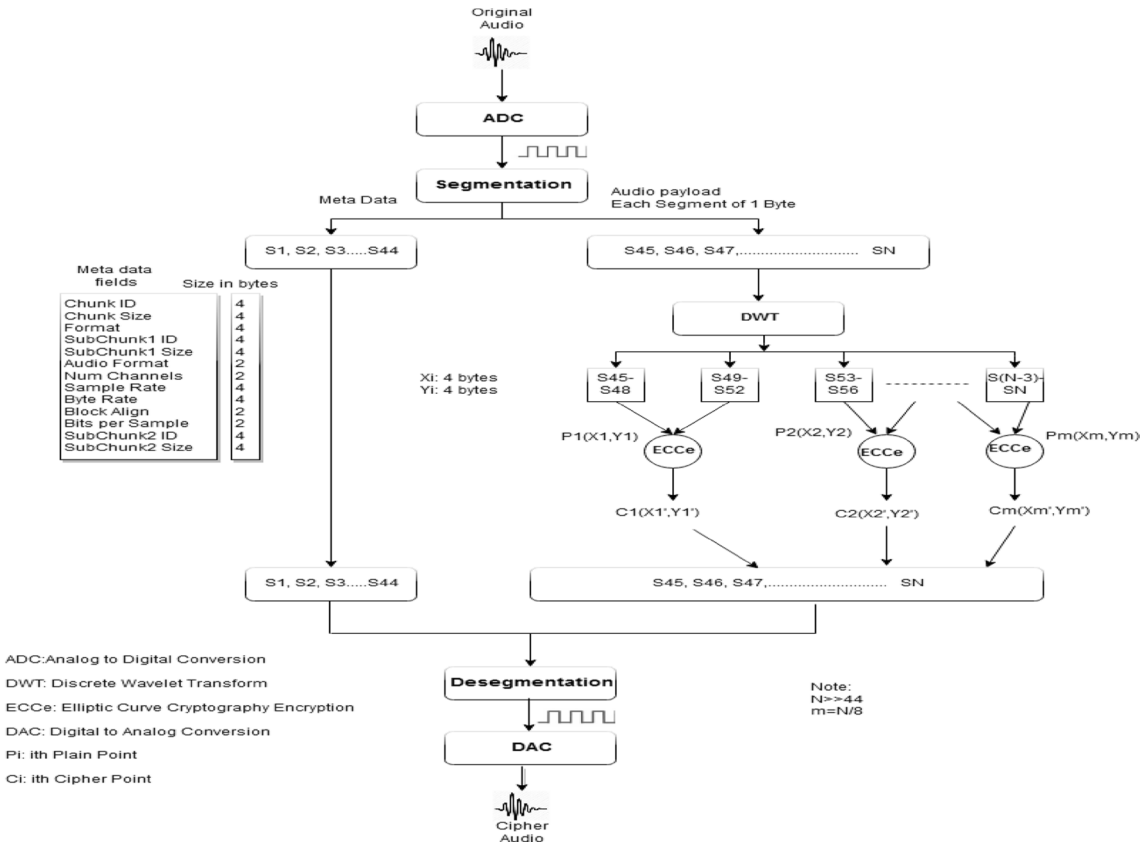

Fig. 2 SAIL cryptosystem forward process

Selected curve is well suitable for encryption of all three different categories of audio such as human voice, animal sound, and instrumental music.

Algorithm 1: SEED Encryption

Input: Digitized plain audio signal A_p

Output: Cipher audio signal A_c

Procedure

Step 1. Digital segmentation tier

Digitized audio signal A_p is fragmented into segments of size 1 byte $A_p = \left\{ S_p^1, S_p^2, S_p^3, \ldots S_p^N \right\}$

First 44 bytes $A_{sw} = \left\{ S_p^1, S_p^2, S_p^3, \ldots S_p^{44} \right\}$ contains wave data and the remaining $A_{spl} = \left\{ S_p^{44}, S_p^{45}, S_p^{47}, \ldots S_p^N \right\}$ is the audio payload.

Step 2. One dimensional compression tier

$A_{spl} = \left\{ S_p^{44}, S_p^{45}, S_p^{47}, \ldots S_p^N \right\}$Compute the length of the audio payload Vector: $N = \left| A_{spl} \right|$

Haar scaling function is described in Eq. (11)

$$\phi(x) = \begin{cases} 1, \ if \ 0 \ \leq x < 1 \\ 0, \ Otherwise \end{cases} \tag{11}$$

Haar wavelet mother function is described in Eq. (12)

$$\psi(x) = \phi(2x) - \phi(2x-1) \tag{12}$$

$$\psi(x) = \begin{cases} -1, \ {}^1\!/_2 \ \leq x < 1 \\ 1, \ 0 \leq x < {}^1\!/_2 \\ 0, \ Otherwise \end{cases}$$

Consider the audio paylad as a vector of length $N = 2^n$

1-level Haar transform for $f = (x1, \ x2, \ldots .xn)$

$$f \underset{H_1}{\rightarrow} \left(a^1 \middle| d^1 \right) \tag{13}$$

where,

$$a^1 = \left(\frac{x_1 + x_2}{\sqrt{2}}, \frac{x_3 + x_4}{\sqrt{2}}, \ldots, \frac{x_{N-1} + x_N}{\sqrt{2}} \right) \tag{14}$$

$$d^1 = \left(\frac{x_1 - x_2}{\sqrt{2}}, \frac{x_3 - x_4}{\sqrt{2}}, \ldots, \frac{x_{N-1} - x_N}{\sqrt{2}} \right) \tag{15}$$

1-level Haar wavelets:

$$W_1^1 = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0, 0, \ldots 0 \right)$$

$$W_2^1 = \left(0, 0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \ldots 0 \right)$$

$$W_{N/2}^1 = \left(0, 0, 0, 0, .. \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right)$$

Therefore, d1 is represented in Eq. (16)

$$d^1 = (fW_1^1, fW_2^1, \ldots fW_{N/2}^1) \tag{16}$$

1-level Haar scaling functions:

$$V_1^1 = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0, 0, \ldots 0 \right)$$

$$V_2^1 = \left(0, 0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \ldots 0 \right) \ldots$$

$$V_{N/2}^1 = \left(0, 0, 0, 0, .. \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right)$$

Therefore, a1 is represented in Eq. (17)

$$a^1 = (fV_1^1, fV_2^1, \ldots fV_{N/2}^1) \tag{17}$$

$V_1^1, V_2^1, \ldots .V_{N/2}^1, W_1^1, W_2^1, \ldots W_{N/2}^1$ construct an orthonormal basis in an N-dimensional space.

$$V_i^1 . V_j^1 = 0, \ W_i^1 . W_j^1 = 0, \ i \neq j, \ V_i^1 . W_i^1 = 0 \tag{18}$$

$\left| V_i^1 \right| = \left| W_i^1 \right| = 1$ They form a new coordinate system.

Step 3. Encryption

Elliptic curve chosen for the sound encryption is $E_{2147483647}$ $(0, 390064447)$ based on CM method presented above.

The generator point M is $(1027045486, 1393612238)$ is selected

Select random seed 'k' from $[1..(n-1)]$, $j = 1$

For every 8 bytes in audio payload repeat the following

Form the point $X_i$by considering the first 4 bytes as x coordinate and remaining 4 bytes as the y coordinate

Cipher pints will be generated by Eq. (20)

$$C_{i1} = k * X_i,\ C_{i2} = M + k * QQ \tag{19}$$

$$Y_j = C_{i1}, C_{i2} \tag{20}$$

Cipher audio signal $A_c = A_{sw}$ padded with Y where A_{sw} is the wave data; and is the compressed, encrypted audio

3.3 Proposed SEED decryption algorithm

The audio decryption algorithm is shown in Fig. 3, and its details are given in Algorithm 2. The cipher audio is digitized and segmented with the segment size of 8 bytes. The first 44 bytes represent the wave description, and the remaining is the sequence of 8 bytes data. Each pair of 8 bytes data is decrypted by using Elliptic curve decryption algorithm. Finally, ID Inverse Wavelet Transform is applied to recover the original signals.

Algorithm 2: **SEED Decryption**
Input: Cipher audio signal A_c
Output: Digitized plain audio signal A_p
Procedure:

Step 1. Digital Segmentation

Digitized audio signal A_c is fragmented into segments of size 8 bits $A_c = \left\{S_c^1, S_c^2, S_c^3 \ldots S_c^N\right\}$ $A_c = \{S_{p1}, S_{p2}, S_{p3} \ldots S_{pn}\}$.

First 44 bytes $A_{sw} = \left\{S_p^1, S_p^2, S_p^3, \ldots S_p^{44}\right\}$ contains wave data and the remaining $A_{sc} = \left\{S_c^{44}, S_c^{45}, S_c^{46}, \ldots S_c^N\right\}$ is the audio payload.

Step 2. Decryption

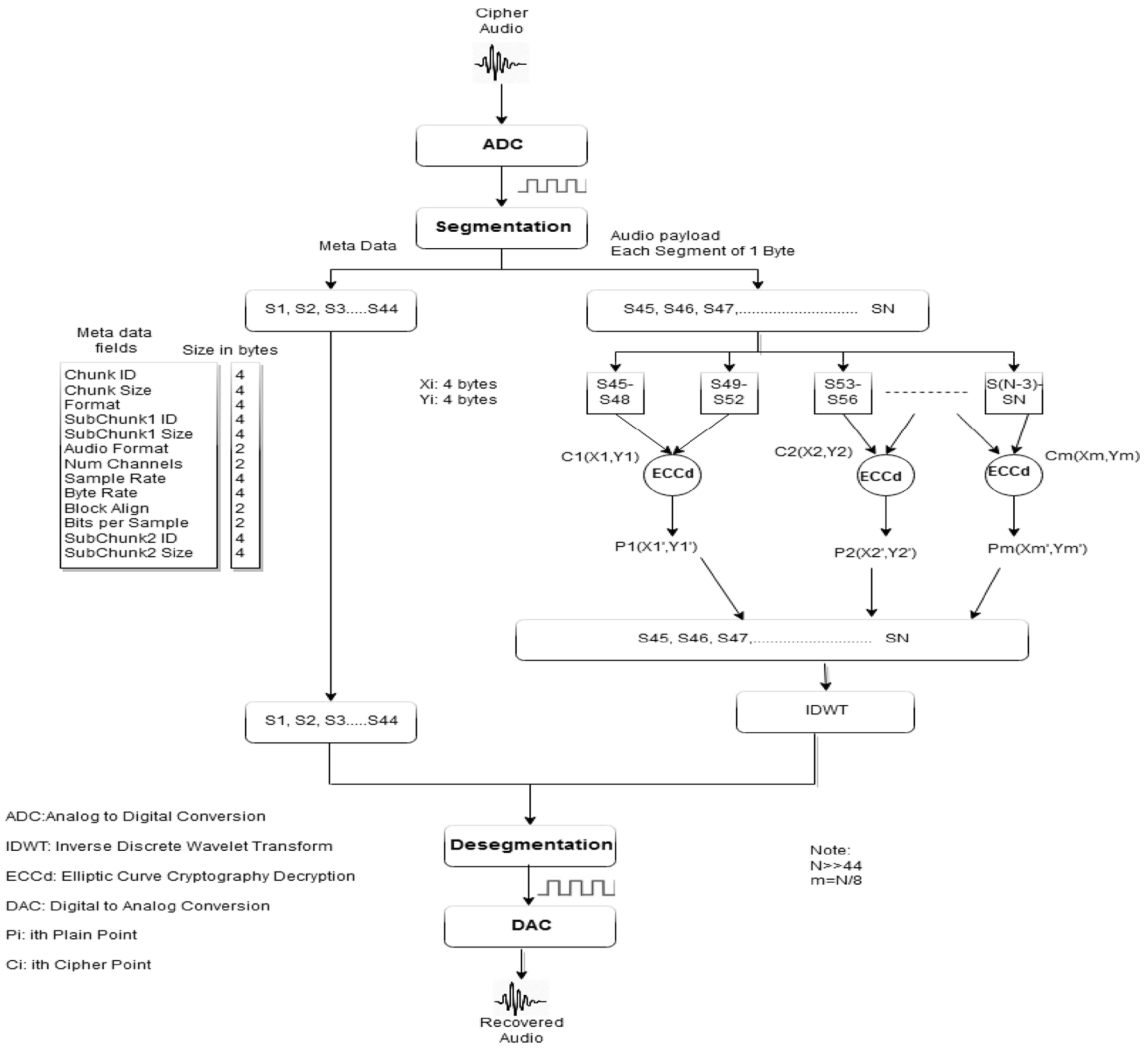

Fig. 3 SAIL cryptosystem reverse process

Elliptic curve chosen for the audio encryption is $E_{2147483647}$ $(0, 390064447)$

Generator point M is (1027045486, 1393612238)

Select the random seed' from $[1..(n-1)]$

For each Cipher pair in A_{sc} repeat the following $X_i = C_{i2} - d * C_{i1}$

Step 3. ID Decompression

Haar wavelet defined in defined in (5) used here for Inverse Discrete Wavelet Transform.

The transformation H_i is reversible. That means, f is reconstructed via (a^1, d^1)

$$a^1 = (a_1, \ldots a_{N/2}) \tag{21}$$

$$d^1 = (d_1, \ldots d_{N/2}) \tag{22}$$

$$f = \left(\frac{a_1 + d_1}{\sqrt{2}}, \frac{a_1 - d_1}{\sqrt{2}}, \ldots, \frac{a_{N/2} + d_{N/2}}{\sqrt{2}}, \frac{a_{N/2} - d_{N/2}}{\sqrt{2}} \right) \tag{23}$$

Reconstruction from 1-level Haar transform Eq. (24)

$$f = \left(\frac{a_1 + d_1}{\sqrt{2}}, \frac{a_1 - d_1}{\sqrt{2}}, \ldots . \frac{a_{N/2} + d_{N/2}}{\sqrt{2}}, \frac{a_{N/2} - d_{N/2}}{\sqrt{2}} \right) = A^1 + D^1 \tag{24}$$

$A_p = A_{sw}$ padded with X; where A_{sw} is the wave data, and X is the decrypted and decompressed audio

4 Experimental results

The designed SAIL system has been implemented in python, statistical and security analyses have been performed in Matlab. Audio Signals with the sampling rate of 8 kHz is used for human voice and animal voice. The sampling rate of 48 kHz used for instrumental music. All the audio signals are initially fed in the uncompressed form. Samples are taken from three different categories namely human voice, animal voice and instrumental music. These signals are encoded into binary using quantization. Normally, the "dense" visual feature of the waveform replicates the quick differences rising from the encryption process.

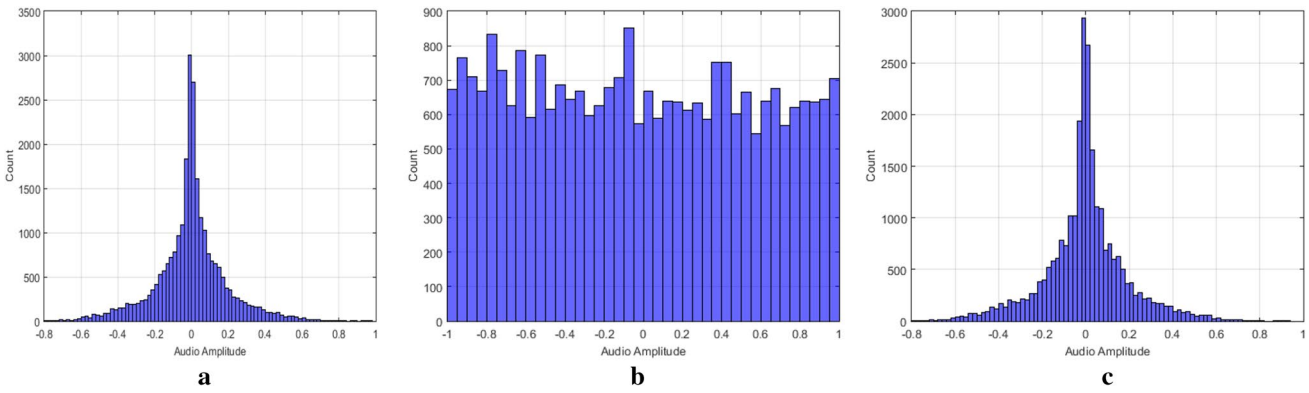

Fig. 4 **a** Dog barking sound, **b** encrypted sound, **c** decrypted sound

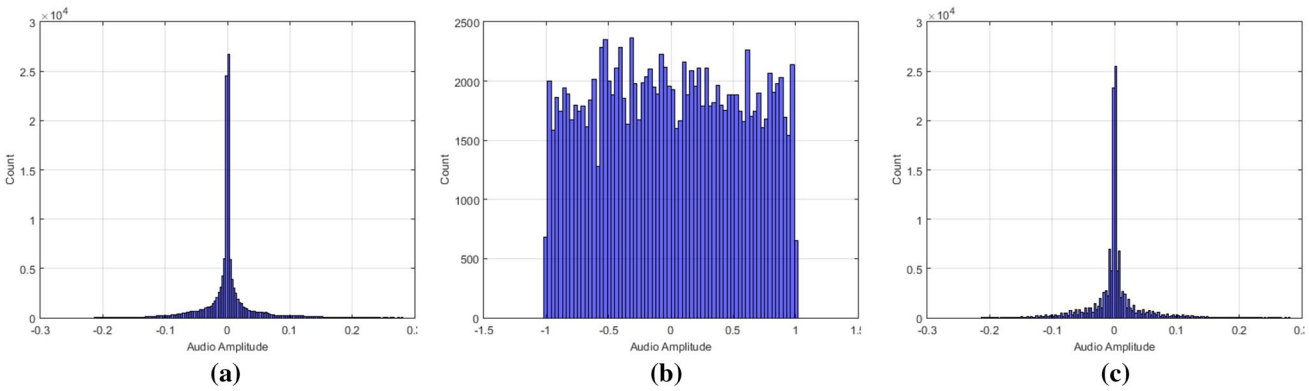

Fig. 5 **a** Human voice (hello), **b** encrypted voice, **c** decrypted voice

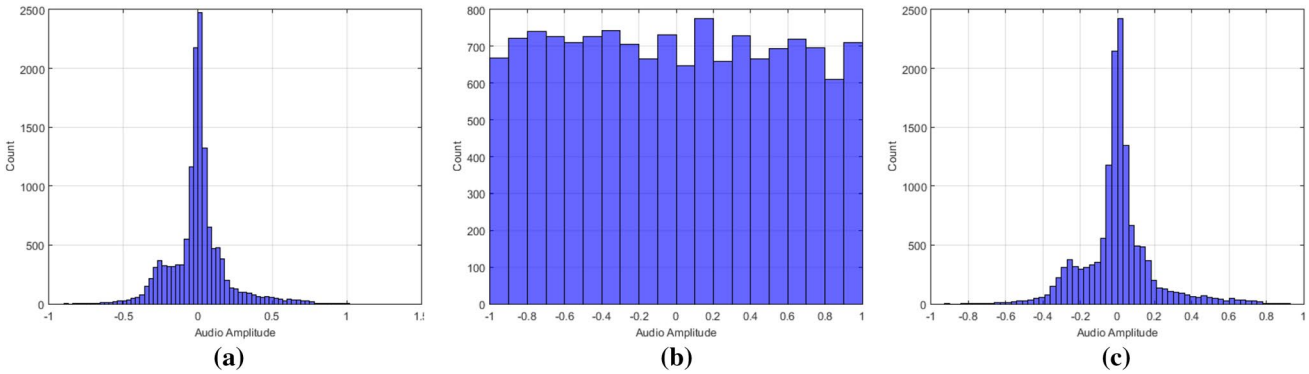

Fig. 6 **a** Music instrument sound (piano), **b** encrypted music, **c** decrypted music

4.1 Histogram analysis

Histogram analysis is performed on all three categories namely animal sound, human sound, and instrumental music and depicted in Figs. 4, 5 and 6. Figure 4a–c displays the original audio, encrypted version and its equivalent decrypted version all the three categories respectively. Figure 4b shows the histogram of an encrypted sound file. Figure 4c illustrates the decrypted sound. The stringent property of the encrypted part of the audio signals is also shown in their corresponding histograms. In Fig. 4c, the histogram of dog barking sound is shown; it followed a specific distribution model, which is alike to the distributions obtained for the other plain audio signals. Otherwise, the histogram of the encrypted version of audio Fig. 4b has a very flat structure. This response is also tested for the different audio signals such as human voice and instrumental music.

4.2 Time domain and frequency domain analysis

The time domain and frequency domain characteristics charts of the plain audios (human voice, animal sound, and instrumental sound) and its equivalent encrypted audio is shown in Figs. 7, 8 and 9. The figures indicate that the encrypted audio has no similarity to the plain audio and is full of noise and hence imperceptible. The decryption algorithm recovered the original sound successfully. Figures 7, 8 and 9 show that the decrypted audio resembles the original audio.

4.3 Correlation analysis

Statistical properties of the original signal and encrypted signal are analyzed by calculating the correlation coefficients. Equation (25) shows the correlation coefficient formula. It is computed on randomly selected P sample in the different categories of the audio signals such as human voice, animal sound, and instrumental music.

$$r_{xy} = \frac{cov(x,\ y)}{\sqrt{D(x)D(y)}} \tag{25}$$

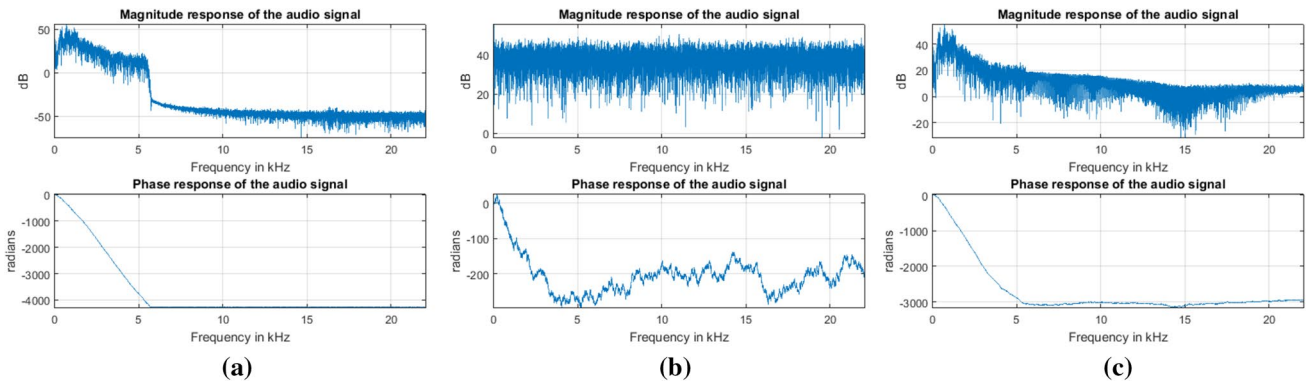

Fig. 7 **a** Dog barking sound, **b** encrypted audio, **c** decrypted sound

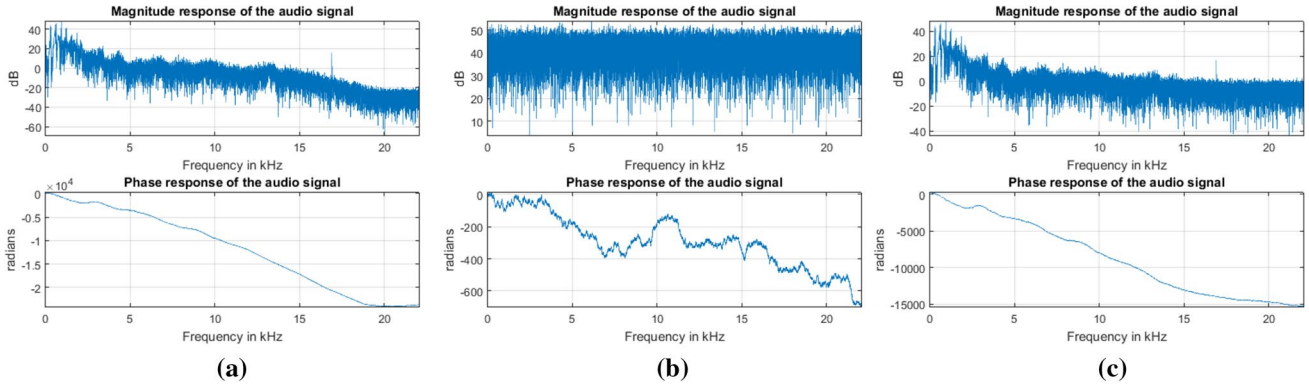

Fig. 8 **a** Human voice (hello), **b** encrypted voice, **c** decrypted voice

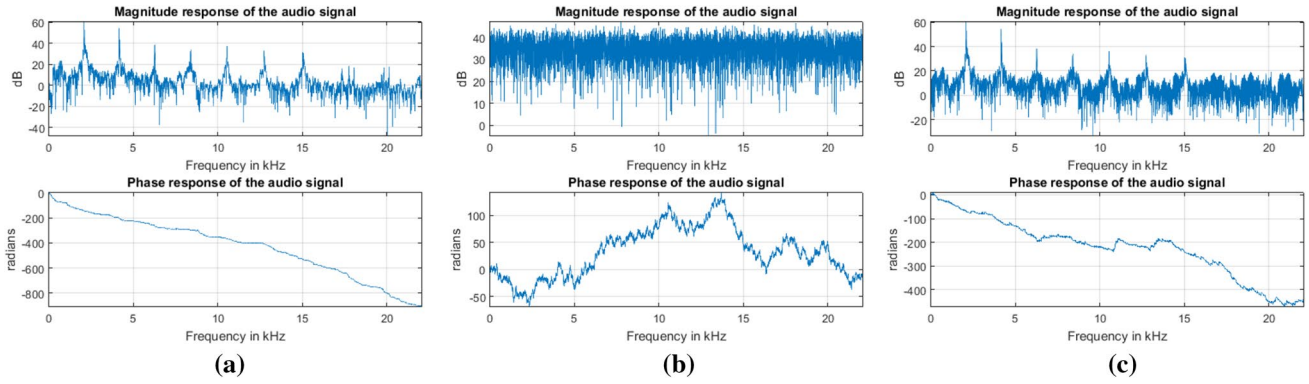

Fig. 9 **a** Music instrument sound (piano), **b** encrypted music, **c** decrypted music

Table 1 Correlation analysis of recovered and cipheredaudio w.r.t original audio

Audio	Correlation coefficient of recovered audio	Correlation coefficient of ciphered audio
Instrumental audio	0.997	0.0133
Human audio	0.9876	−0.0040
Animal audio	0.95497	−0.0011

Table 2 MSE of decrypted audio concerning original

Audio	MSE of recovered audio
Instrumental audio	5.7714e-04
Human audio	4.1163e-06
Animal audio	6.0708e-04

$$conv(ex, y) = \frac{1}{p}\sum_{i=1}^{p}\big(x_i - E(x)\big)\big(y_i - E(y)\big),$$

$$D(x) = \frac{1}{p}\sum_{i=1}^{p}(x_i - E(x))^2,\ E(x) = \frac{1}{p}\sum_{i=1}^{p}x_i \tag{26}$$

Xi is the value of the n-th chosen the audio sample, and you are the value of the equivalent adjoining audio sample. Original digital audio signals have correlation coefficients near to 1 exhibiting closer resemblance, whereas encrypted digital audio signals have correlation coefficients near to zero and hence claiming no resemblance as shown in Table 1.

This illustrates that the proposed method is not vulnerable to statistical attacks. In addition to this, the entropy of the encrypted digital audio signals has inherent values ranging from 15.7057 to 15.7117. Even though these ranges are larger than those usually observed for original 16-bit audio signal, they are not too near to 16. This is because of the association between the numbers of samples of the audio signals used in the experiments. The equivalent encrypted audio has entropy equal to 15.9735, which is considerably near to 16. The related performance is tested for all types of audio samples. This implies that the encrypted audio signals are near to a random basis and the proposed model is also secure against the entropy attack.

Table 3 PSNR of decrypted audio concerning the original audio

Audio	PSNR of recovered audio
Instrumental audio	1.6060e+03
Human audio	1.6490e+03
Animal audio	1.6056e+03

4.4 MSE and PSNR analysis

MSE and PSNR is calculated for all the audio samples taken for analysis.

$$MSE = \frac{1}{N * M} \sum_{n=1}^{N} \sum_{m=1}^{M} \left[f(i,j) - f_0(i,j)^2 \right] \tag{27}$$

where f and f_0 are the intensity functions of decrypted and original sounds. (i, j) is the position of the data. (N*M) is the size of the sound file. Table 2 shows the MSE of recovered sound. It shows that the MSE of the decrypted sound concerning its original image is closer to 0 which is desirable.

PSNR is the ratio of the mean square difference of two sounds to the maximum mean square differences that exist between two audio files. Larger the value of PSNR, greater the quality of the sound. PSNR value is tabulated in Table 3.

$$PSNR = 20 * \log \frac{255^2}{\sqrt{MSE}} \tag{28}$$

4.5 Power spectrum analysis

The power spectral density (PSD) is the distribution of power per unit frequency. It calculates the PSD of discrete time domain based audio signals using spectrum. The PSD is generalized to discrete time variables. Signals are sampled at discrete time intervals $x_n = x(n\Delta t)$ for a total measurement period of $T = N\Delta t$. Figures 10, 11 and 12 shows the PSD of the original sound, encrypted sound and the decrypted sound for the different categories of audio signals. From the figures, it can be inferred that the PSD of the original sound and encrypted sound has great variation but the PSD of the original sound and encrypted sound remains same.

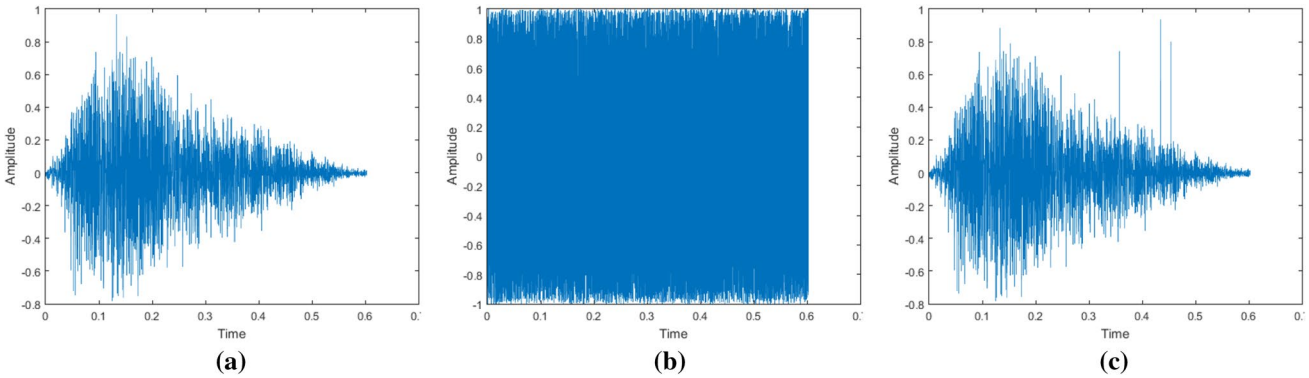

Fig. 10 **a** Dog barking sound, **b** encrypted sound, **c** decrypted sound

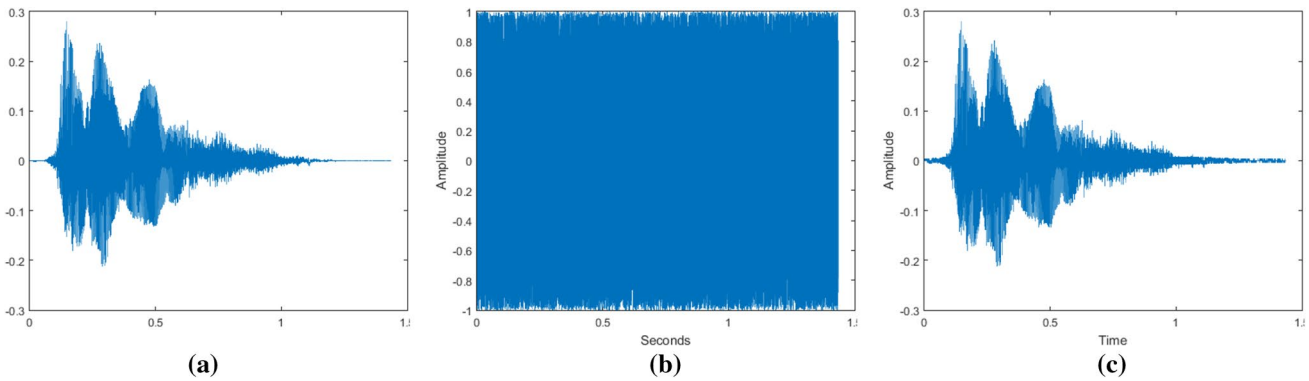

Fig. 11 **a** Human voice (hello), **b** encrypted voice, **c** decrypted voice

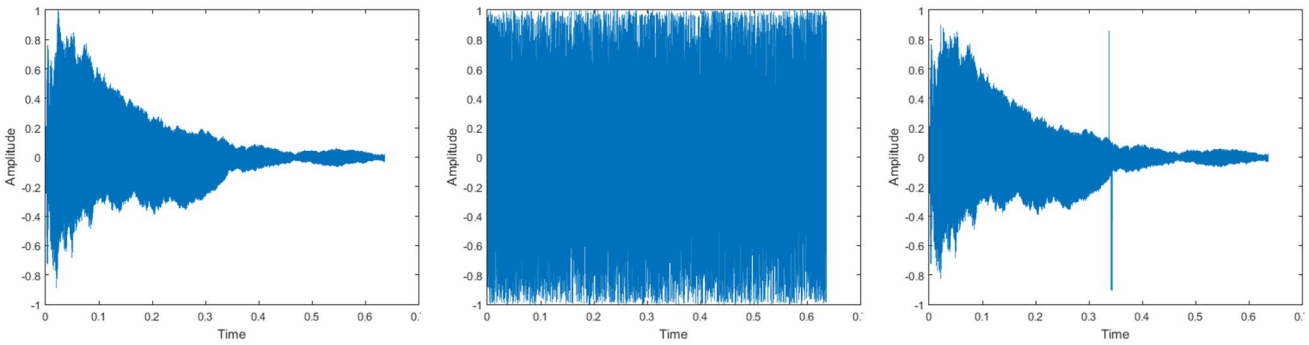

Fig. 12 **a** Music instrument sound (piano), **b** encrypted music, **c** decrypted music

4.6 Keyspace analysis

SEED encryption and decryption algorithms are based on DWT and ECC. ECC provides the equal level of security as compared to RSA with concise key length. As ECC is based on discrete logarithm problem (DLP), the brute force attack is impossible. Key size is chosen such a way that it should be best suited for RTP applications which are not tolerant to delays without compromise on the security level. The algorithm has been designed to scale up for larger key size. To preserve privacy, the random seed has been used as practiced in Diffie hellman key exchange. To break the cryptosystem adversary has to know the random seed which is not possible in the proposed system. Elliptic curve chosen for the sound encryption is $E_{2147483647}$ $(0, 390064447)$ and the generator point is (1027045486, 1393612238).

4.7 Key sensitivity analysis

The proposed algorithm is very much sensitive to the key, even one-bit change in the decryption key will provide nosiy audio and making it irrecoverable. This proposed cryptosystem is fully based on the random key. The usage of random keys provides different cipher audio for a given clear audio and hence making the known-plaintext attack and chosen plaintext attack ineffective. Key sensitivity test has been conducted by changing the initial parameters used for decryption which resulted in a completely different cipher audio.

4.8 Robustness to differential attacks

One sample is selected at random, to analyze the vulnerability of the proposed method against differential attack. The audio signal is modified by inverting the Least significant bit (LSB) of the sample. Modified and original audio is encrypted using the same key and evaluated by using the number of samples change rate (NSCR) and the unified average changing intensity (UACI) as given below:

$$NSCR = \frac{\sum_i D_i}{L} * 100\% \tag{29}$$

$$UACI = \frac{1}{L}\left[\sum_i \frac{\left|A_i - A_i'\right|}{65535}\right] \tag{30}$$

A and A_i' are the two encrypted audio data whose equivalent plain audio data have only single bit change in the sample; the values of the samples at location I of A and A_i' are correspondingly represented by Ai and A; L corresponds to the size of the audio vector; Di is calculated based on the rule,

$$D_i = \begin{cases} 1, & A_i \neq A_i' \\ 0, & Otherwise \end{cases} \tag{31}$$

The benchmark value for NSCR is 100% and for UACI is 33.3%. The minimum, the maximum and the average values of NSCR and UACI, calculated from the encryption of 100 different modified versions of each audio signal. Computed NSCR values are closer to 98%, and UACI is closer to 33%. The results are considerably closer to the ideal values and in depend on the position of the modified sample.

5 Conclusion

Audio security ensures the secrecy, integrity, accessibility and confidentially of the audio signal. This multi-tier SEED model performs DWT to compress the audio signal which can suit well for Real Time Protocol (RTP) based applications like VoIP, live audio streaming and video conferencing. Digital audio encryption is made as it can provide lower residual intelligibility and intensified cryptanalytic strength. Application of ECC claims this work to be unique of its kind as it is suited for digital encryption. Since larger key size

will be inappropriate for RTP and delay sensitive applications an optimal key size is chosen without compromising security. The SEED model provides the faster encryption as it performs the fixed-point operation that involves less computation time. This model is easy to implement in spite of its mathematical complexity but offers the higher degree of flexibility, as samples range can vary from 8 to 16 k. Various statistical analysis has been performed, and the results substantiate the higher level of security and ensure it is not vulnerable to any statistical attacks and hence more prudent for multi-channel audio processing.

Acknowledgements This part of this research work is supported by Department of Science and Technology (DST), Science and Engineering Board (SERB), Government of India under the ECR Grant (ECR/2017/000679/ES).

References

Akgül, A., & Kaçar, S. (2015). An audio data encryption with single and double dimension discrete-time chaotic systems. *Turkish Online Journal of Science & Technology, 5*(3), 14–23

Al Saad, S. N., & Hato, E. (2014). A speech encryption based on chaotic maps. *International Journal of Computer Applications, 93*(4), 19–28.

al-Karim, A., Al-Jalil, M. A., & Qays, I. (2013). Speech encryption using a chaotic map and blowfish algorithms. *Journal of Basrah Researches (Sciences), 39*(2), 68–76.

Alwahbani, S. M. H., & Bashier, E. B. M. (2013). Speech scrambling based on chaotic maps and one-time pad. *2013 International Conference on Computing, Electrical and Electronics Engineering* (*ICCEEE*), IEEE (pp. 128–133).

Anees, A. (2015). An image encryption scheme based on Lorenz system for low profile applications. *3D Research, 6*(3), 1–10.

Asok, S. B., et al. (2013). A secure cryptographic scheme for audio signals. *2013 International Conference on Communications and Signal Processing* (*ICCSP*), IEEE.

Augustine, N., George, S. N., & Pattathil, D. P. (2015). An audio encryption technique through compressive sensing and Arnold transform. *International Journal of Trust Management in Computing and Communications, 3*(1), 74–92.

Barni, M., Bartolini, F., & Piva, A. (2001). Improved wavelet-based watermarking through pixel-wise masking. *IEEE Transactions on Image Processing, 10*(5), 783–791.

Chen, L. H., & Lin, J. J. (2003). Mean quantization based image watermarking. *Image and Vision Computing, 21*(8), 717–727.

Cichowski, J., & Czyzewski, A. (2012). Sensitive audio data encryption for multimodal surveillance systems. In Audio Engineering Society Convention 132. Audio Engineering Society.

Ciptasari, R. W., Rhee, K. H., & Sakurai, K. (2014). An enhanced audio ownership protection scheme based on visual cryptography. *EURASIP Journal on Information Security, 1*, 2.

Datta, K., & Gupta, I. S. (2013). Partial encryption and watermarking scheme for audio files with controlled degradation of quality. *Multimedia Tools and Applications, 64*(3), 649–669.

Dengre, A., & Gawande, A. D. (2015). Audio encryption and digital image watermarking in an uncompress video. *International Journal of Advances in Applied Sciences, 4*(2), 66–72.

Eldin, S. M. S., et al. (2015). New audio encryption package for TV cloud computing. *International Journal of Speech Technology, 18*(1), 131–142.

Elkholy, M. M., Hennawy, H. M. E. L., & Elkouny, A. (2015). Design and implementation of hyperchaotic masking system for secured audio transmission. *2015 Tenth International Conference on Computer Engineering & Systems* (*ICCES*), IEEE.

Elshamy, A. M., et al. (2013). Optical image encryption based on chaotic baker map and double random phase encoding. *Journal of Lightwave Technology, 31*(15), 2533–2539.

Fazeen, M., Bajwa, G., & Dantu, R. (2014). Context-aware multimedia encryption in mobile platforms. *Proceedings of the 9th Annual Cyber and Information Security Research Conference*. ACM

Hedelin, P., Nordén, F., & Skoglund, J. (1999). SD optimization of spectral coders. *1999 IEEE Workshop on Speech Coding Proceedings*, IEEE.

Iyer, S. C., Sedamkar, R. R., & Gupta, S. (2016). A novel idea on multimedia encryption using hybrid crypto approach. *Procedia Computer Science, 79*, 293–298.

James, S. P., George, S. N., & Deepthi, P. P. (2014). An audio encryption technique based on LFSR based alternating step generator. *2014 IEEE International Conference on Electronics, Computing and Communication Technologies* (*IEEE CONECCT*), IEEE

Kim, S. W., Kim, Y. G., & Simon, M. K. (2004). Generalized selection combining based on the log-likelihood ratio. *IEEE Transactions on Communications, 52*(4), 521–524.

Kohad, H., Ingle, V. R., & Gaikwad, M. A. (2012). An overview of speech encryption techniques. *International Journal of Engineering Research and Development, 3*, 29–32.

Kulkarni, S. A., & Patil, S. B. (2015). A robust encryption method for speech data hiding in digital images for optimized security. *2015 International Conference on Pervasive Computing* (*ICPC*), IEEE.

Kwon, J. K., Park, S., & Sung, D. K. (2006). Collision mitigation by log-likelihood ratio (LLR) conversion in orthogonal code-hopping multiplexing. *IEEE Transactions on Vehicular Technology, 55*(2), 709–717.

Langelaar, G. C., Setyawan, I., & Lagendijk, R. L. (2000). Watermarking digital image and video data. A state-of-the-art overview. *IEEE Signal processing magazine, 17*(5), 20–46.

Li, H., et al. (2009). A novel audio scrambling algorithm in variable dimension space. *11th International Conference on Advanced Communication Technology, 2009, ICACT*. (Vol. 3, pp 1647–1651). IEEE.

Lima, J. B., & da Silva Neto, E. F. (2016). Audio encryption based on the cosine number transform. *Multimedia Tools and Applications, 75*(14), 8403–8418.

Lin, C. Y., & Chang, S. F. (2001). A robust image authentication method distinguishing JPEG compression from malicious manipulation. *IEEE Transactions on Circuits and Systems for Video Technology, 11*(2), 153–168.

Liu, H., & Wang, X. (2010). Color image encryption based on one-time keys and robust chaotic maps. *Computers & Mathematics with Applications, 59*(10), 3320–3327.

Lu, X., et al. (2012). Digital audio information hiding based on Arnold transformation and double random-phase encoding technique. *Optik-International Journal for Light and Electron Optics, 123*(8), 697–702.

Mermoul, A., & Belouchrani, A. (2010). A subspace-based method for speech encryption. *2010 10th International Conference on, Information Sciences Signal Processing and their Applications* (*ISSPA*). IEEE.

Mostafa, A., et al. (2015). Speech encryption using two-dimensional chaotic maps. *2015 11th International Computer Engineering Conference* (*ICENCO*), IEEE

Nguyen, H. H., Mehaoua, A., & Hong, J. W. K. (2013). Secure medical tele-consultation based on voice authentication and realtime audio/video encryption. *2013 First International Symposium on Future Information and Communication Technologies for Ubiquitous HealthCare* (*Ubi-HealthTech*), IEEE.

Petitcolas, Fabien, A. P., Ross, J.. Anderson, & Markus, G., & Kuhn (1999). Information hiding-a survey. *Proceedings of the IEEE 87*(7): 1062–1078.

Rashidi, B., & Rashidi, B. (2013). FPGA based A new low power and self-timed AES 128-bit encryption algorithm for encryption audio signal. *International Journal of Computer Network and Information Security, 5*(2), 10.

Refregier, P., & Javidi, B. (1995). Optical image encryption based on input plane and Fourier plane random encoding. *Optics Letters, 20*(7), 767–769.

Sadkhan, S. B., & Mohammed, R. S. (2015). Proposed random unified chaotic map as PRBG for voice encryption in wireless communication. *Procedia Computer Science, 65*, 314–323.

Sharma, D. (2012). Five level cryptography in speech processing using multi hash and repositioning of speech elements. *International Journal of Emerging Technology and Advanced Engineering, 2*(3), 21–26

Sheu, L. J. (2011). A speech encryption using fractional chaotic systems. *Nonlinear Dynamics, 65*(1), 103–108.

Tamimi, A. A., & Abdalla, A. M. (2014). An audio shuffle-encryption algorithm. *Proceedings of the World Congress on Engineering and Computer Science*. San Francisco, USA: WCECS

Tong, X. J., et al. (2015). A fast encryption algorithm of the color image based on the four-dimensional chaotic system. *Journal of Visual Communication and Image Representation, 33*, 219–234.

Wang, S., & Fan, Y. (2010). A watermarking algorithm of gray image based on histogram statistical characteristics. *Computer Technology and Development, 1*, 045.

Washio, S., & Watanabe, Y. (2014). Security of audio secret sharing scheme encrypting audio secrets with bounded shares. *2014 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*). IEEE.

Wu, Y., & Ng, B. P. (2002). Speech scrambling with Hadamard transform in the frequency domain. *Signal Processing, 2002 6th International Conference on*. Vol. 2. IEEE

Yang, W., Benbouchta, M., & Robert, Y. (1998). Performance of the modified bark spectral distortion as an objective speech quality measure. *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*. Vol. 1. IEEE.

Zeng, L., et al. (2012). Scrambling-based speech encryption via compressed sensing. *EURASIP Journal on Advances in Signal Processing, 2012*(1), 257.

Zhang, C., Wang, J., & Wang, X. (2008). Digital image watermarking algorithm with double encryption by Arnold transform and logistic. *NCM'08 Fourth International Conference on Networked Computing and Advanced Information Management*, Vol. 1. IEEE

Zhao, H., et al. (2014). Dual key speech encryption algorithm based underdetermined BSS. *The Scientific World Journal*. https://doi.org/10.1155/2014/974735