Multimedia Systems (2020) 26:301–311
https://doi.org/10.1007/s00530-019-00644-6

REGULAR PAPER

Con(dif)fused voice to convey secret: a dual-domain approach

C. Lakshmi[1] · Vasudharini Moranam Ravi[1] · K. Thenmozhi[1] · John Bosco Balaguru Rayappan[1] · Rengarajan Amirtharajan[1]

Received: 17 May 2019 / Accepted: 30 December 2019 / Published online: 10 January 2020

Abstract
Owing to the rapid development and advancements in the field of networks and communication, sharing of multimedia contents over insecure networks has become vital. The confidentiality of audio signals is predominantly needed in military and intelligence bureau applications. The proposed algorithm addresses this issue by encrypting audio signal using chaotic maps in spatial and transform domain. Discrete Fourier transform (DFT), discrete cosine transform (DCT) and integer wavelet transform (IWT) approaches are considered for the experiment. The algorithm involves three-layer security of confusion and diffusion in the spatial domain, and confusion in the transform domain. The confusion in the transform domain is equivalent to diffusion in the spatial domain. Different sizes of audio samples are considered to validate the effectiveness of the proposed scheme. Experimental results prove that the DFT-assisted encryption scheme is more efficient than the DCT- and IWT-based methods because the DFT scheme employs effective diffusion through reversible phase coding. Effectiveness of the proposed method is substantiated using various metrics. Correlation coefficients arrive significantly closer to zero; number of samples changes rate (NSCR) value is at 100% and scrambling degree close to 1. Besides, the proposed scheme has a larger keyspace higher than 2^{128}. Thus, the proposed algorithm has the potency to withstand the statistical, differential and brute force attacks.

Keywords Audio encryption · Tent map · Logistic map · Discrete Fourier transform · Discrete cosine transform · Integer wavelet transform

1 Introduction

In recent times, multimedia data security has gained a lot of attention, as sharing multimedia data, namely text, image, audio or video [1, 2] in which audio data are focused more due to field of audio conferencing, audio processing, military transmission and biometric audio authentication. Information security is classified as encryption Steganography and watermarking.

Steganography and watermarking methods [3–7] involve information hiding policy to address copyright protection. Multimedia encryption employs mathematical tools and strategies to map original data to an unrecognizable format to address confidentiality [8–12]. Due to their properties, every multimedia datum requires a unique method to acquire security. In audio encryption, researchers incorporated traditional methods such as AES [13] and DES for an extended period owing to their strong protection, However, due to their limited keyspace, they are prone to brute force attacks. Due to drawbacks found in traditional methods, chaos-based encryption methodologies are in practice [14–16].

2 Related works

During the diffusion phase, the position of the instantaneous value of audio is modified which results in meaningless audio in turn to maximize the entropy. Chaotic maps are used to generate deterministic random numbers based on their initial parameters. They are subdivided into two, one-dimensional chaotic maps [11, 16, 17] and multidimensional maps. One-dimensional chaotic maps depends on single control parameter and hence simpler compared

Communicated by F. Wu.

✉ Rengarajan Amirtharajan
amir@ece.sastra.edu

[1] School of Electrical and Electronics Engineering, SASTRA Deemed University, Thanjavur 613 401, India

to higher-dimensional chaotic maps [18] which include more parameters and are more complicated. Compared to other methods, Different techniques can be applied to various applications to obtain encrypted audio. The algorithm proposed in [18] utilizes higher dimension chaos maps to enhance its security and keyspace, as the chaotic map has significant merit such as key sensitivity due to their real-valued key.

Ghasemzadeh [19] proposed an audio encryption method which employs combined chaos to get a reversible and flexible encryption scheme. Wang suggested a pseudo-random number generation method to enhance the quality of the encryption [17]. Belmeguenai [20] proposed an encryption scheme that uses a pseudo-random number generator to increase its keyspace. Zaslavsky map as a pseudo-random generator was proposed for speech encryption by Farsana [21]. Eldin proposed an encryption scheme for cloud computing, using chaotic maps and multi-key algorithms along with discrete transforms [22]. Rao proposed modulus multiplicative method for audio encryption which is suitable for the internet application [23]. An audio cryptosystem is proposed in [24] which incorporates deoxyribonucleic acid (DNA) encoding techniques along with chaos map and hybrid chaotic shift transform (HCST). A selective audio encryption scheme was proposed for sensor networks [25], and a partial encryption scheme for mp3 files was proposed using watermarking and shuffling in [26]. In [27], Farsana proposed an audio encryption method based on Fast Walsh Hadamard Transform to remove residual intelligibility in the transform domain.

The spatial domain approaches offer higher keyspace and key sensitivity. However, spatial domain approaches required logical XOR operation to employ the diffusion phase which is vulnerable to the chosen plain text attack. Transform domain approaches overcome this limitation, and the transform domain approaches are more suitable for the real-valued audio samples. There are different transforms available such as discrete Fourier transform, discrete cosine transform (DCT), and discrete sine transform (DST) which transform the time domain content into combination of real and complex forms, real and complex form, respectively[16], quantum fourier transform (QFT) which applies concepts of quantum mechanics to encryption [28]. In number transform, transformation is employed using number theory [29].

Confusion and diffusion are the phases involved in the encryption process [30–32]; confusion, often called a permutation, incorporates a shuffling process to break the correlation between the samples. Fourier transform is segregated into its frequency coefficients and results in real and complex values which contain low- and high-frequency components, respectively. DFT has lesser computational time. Also, DFT-based encryption resists against the channel noise due to its inherent properties of scaling invariant [24]. Belazi [33] proposed a permutation–substitution-based encryption scheme that incorporates chaos systems.

From this literature survey, the robustness of the transform domain approach is better than the spatial domain. Keyspace and key sensitivity are the advantages of chaos approaches.

This paper proposes a tri-layer audio encryption technique which integrates the benefits of chaos and transforms domain by employing DFT, logistic and tent maps. The diffusion process is carried out in the frequency coefficients by means of interchanging the complex part, which is named as phase coding. Thus, DFT aided diffusion destructs the audio details as noisy audio. This diffusion scheme replaces the simple XOR-based diffusion which resists against chosen plain text attack, but at the same time, logistic and tent maps are incorporated along with DFT to enhance the keyspace and key sensitivity to retain the rewards of chaos. The logistic map is implemented to generate the confusion index to perform the confusion process in the time domain, and the tent map is used to generate the scrambling index for employing the confusion in transform domain which is equivalent to the diffusion in the spatial domain. Besides, DCT- and IWT-based encryption is also implemented along with DFT.

Significant contributions of the proposed work as follows,

- This scheme employs tri-level audio encryption through Spatial and Transform domain fusion.
- Confusion and diffusion in the spatial domain increase the keyspace and key sensitivity to improve the level of security.
- Confusion in the transform domain is equivalent to the diffusion in the spatial domain.
- Reversible phase coding through the confusion in the Fourier coefficients.
- Reversible phase coding replaces the conventional diffusion such as XOR which is vulnerable to the chosen plaintext attack.
- Audio input is not digitized to employ the XOR diffusion.
- No loss of data due to digitization and quantization.
- Transform domain approach increases the complexity of the security level.
- Flexibility with the choice of transform.
- The system is simple yet robust with high yielding encryption for audio files.

The proposed algorithm complies with all the required security parameters and integrates merits of the chaos along with the transform domain approaches to offer a triple-layered (Confusion—Diffusion—Confusion) security to the audio files. In the following sections, chaotic maps are defined, followed by the proposed methodology.

3 Preliminaries

Tent map is a linear map, and Logistic map is a non-linear map which provides high computational speed, complexity and security. Chaotic maps are sensitive to the initial conditions which can lead to different results based on the initial value [34, 35]. The maps are employed to generate cyclic random sequences to perform confusion and diffusion [24].

Mathematical model for the tent map is defined as,

$$X_{n+1} = r \times \left[1 - 2\left|X_n - 0.5\right|\right], \tag{1}$$

where $X_n \in [0, 1]$ is the initial parameter, r is a constant

Mathematical model for the logistic Map is defined as,

$$X_{n+1} = r \times X_n\left[1 - X_n\right], \tag{2}$$

where $X_n \in [0, 1]$ is the initial parameter and $r \in [0, 4]$ is the constant parameter which affects the randomness of the system.

4 Proposed methodology

In original audio data, neighboring instants are highly correlated and its amplitude is closely ordered and well organized. To achieve desire encrypted audio, the data should be highly uncorrelated and more randomized. Confusion–diffusion processes are involved in attaining the desire of encrypted audio. This paper proposes a three-layer audio encryption scheme. Initially, confusion is employed in the spatial domain to scramble the organized data, followed by indirect diffusion which marks the second layer of protection, and this is achieved by converting data to transform domain by applying DFT, consequently ending with the third layer of security by implementing a confusion algorithm on the data. Figure 1 illustrates the flow diagram of the proposed scheme.

The proposed algorithm as follows:

Input	Audio file of two channels
Output	Ciphered audio file of two channels
Step 1	Read the dual-channeled audio with samples S and sampling rate, F in Hertz

S is stored as an m-by-n matrix, where m is the number of audio samples and n is the number of channels, where n is 2 for the dual channels. The values stored in S are normalized to the range [− 1.0, 1.0] of type double.

$$S = \begin{cases} S1 = \left[s_{11}, s_{12}, s_{13}, \ldots s_{1m}\right] \\ S2 = \left[s_{21}, s_{22}, s_{23}, \ldots s_{2n}\right] \end{cases} \tag{3}$$

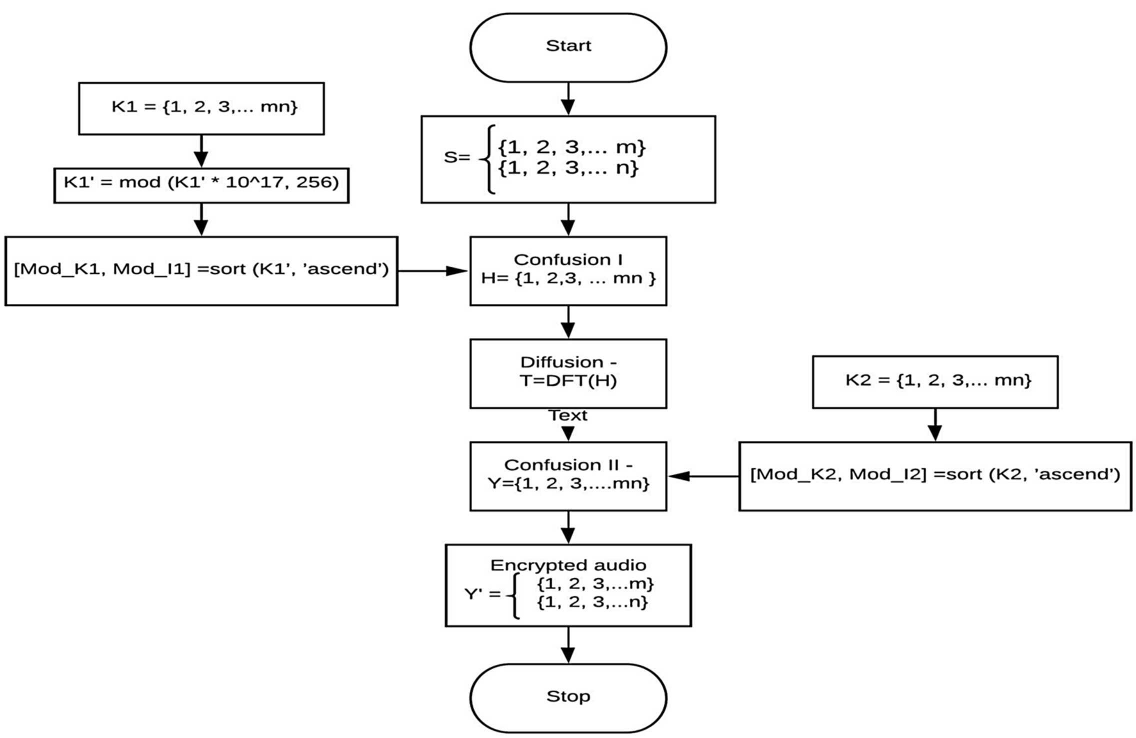


Fig. 1 Flow diagram for the proposed scheme

Step 2 Generate key $K1$ using logistics map given in Eq. (2) using an initial value to generate $m*n$ values using an iterative process

$$K1_i = [k_1, k_2, k_3, \dots k_{mn}], \tag{4}$$

where i represents the ith element in $K1$

Step 3 To Key $K1$, Eq. (5) is applied to get modified key $K1'$

$$K1'_i = \text{mod}\left(K1_i * 10^{17}, 256\right), \tag{5}$$

where mod() is the modulo function, i represents the ith element in $K1$.

Next, rearrange S from m-by-n to a one-dimensional array to get S′.

Then, Eq. (6) is applied on $K1'$ to get Mod_$K1$ and Mod_$I1$, which gives the modified Key with the corresponding Index for each value in $K1$, in the ascending order

$$[\text{Mod\_}K1,\ \text{Mod\_}I1] = \text{sort}\left(K1',\ \text{'ascend'}\right), \tag{6}$$

where sort() algorithm arranges the matrix $K1$ in the ascending order and the rearranged matrix is stored in Mod_$K1$ with its corresponding indices stored in Mod_$I1$.

Then, a permutation algorithm is applied on S' using the modified Index, Mod_$I1$ to get H; this is the first layer of security.

$$H = [h_1, h_2, h_3 \dots h_{mn}]. \tag{7}$$

Step 4 Diffuse the elements of H by applying DFT to H, which gives F; this is the second layer of security.

Step 5 Generate key $K2$ using Tent map given in Eq. (1) using an initial value to generate $m*n$ values using an iterative process

$$K2_i = [k_1, k_2, k_3, \dots k_{mn}]. \tag{8}$$

Next, sort $K2$ using Eq. (6) to get Mod_$K2$ and Mod_$I2$, which gives the modified Key with the corresponding Index for each value in $K2$, in the ascending order

$$[\text{Mod\_}K2,\ \text{Mod\_}I2] = \text{sort}\ (K2,\ \text{'ascend'}), \tag{9}$$

where sort() algorithm arranges the matrix $K2$ in the ascending order and the rearranged matrix is stored in Mod_$K2$ with its corresponding original indices stored in Mod_$I2$.

Step 6 Permute T using modified Index, Mod_$I2$ to get T' Next; rearrange T' to Y, which is the ciphered audio of size m-by-n, where m is the number of samples, n is the number of channels, with the sampling frequency, F

Step 7 Calculate the metrics for the encrypted audio

Step 8 Repeat process from Step 4, replacing DFT with DCT and IWT for comparison

The decryption process involves performing reverse confusion using the tent map, followed by applying inverse Fourier transform and reverse confusion using the logistics map to get the recovered audio.

5 Results and discussion

The security level of the proposed algorithm is evaluated using various Security analyses such as correlation, entropy, NSCR, key sensitivity, scrambling degree and computational time analysis performed on the encrypted audio files. The obtained metrics are compared with the existing scheme and inferences about the results are discussed in this section. Nine test audio files of size 2.62×10^5 are considered for the security analyses.

5.1 Correlation analysis

Correlation between the samples (instantaneous value) is degree of closeness between two neighboring samples. For an original audio, the correlation coefficient is close to one, but the same time, for an ideal audio encryption, the correlation coefficient should be close to zero, signifying uncorrelation [29]. The correlation coefficient is calculated using Eqs. (10)–(10),

$$Q(x) = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{10}$$

$$R(x) = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - Q(x)\right)^2, \tag{11}$$

$$\text{Cov}(x, y) = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - Q(x)\right)\left(y_i - Q(y)\right), \tag{12}$$

$$\gamma_{xy} = \frac{\text{Cov}(x, y)}{\sqrt{R(x)\sqrt{R(y)}}}, \tag{13}$$

where $Q(x)$ is the mean of all x_n samples, $R(x)$ is the standard deviation of x_n samples, Cov(x, y) is the covariance between x and y samples, and γ gives the correlation coefficient for n samples. The correlation coefficient for nine test cases of audio files is given in Table 1 for the three different transform methods (DFT, DCT, and IWT).

The correlation values of adjacent samples of the original and encrypted audio are shown in Table 1. Samples

Table 1 Correlation analysis of the proposed algorithm

Test audio samples	DFT method	DCT method	IWT method
Case 1			
Original correlation	0.9951	0.9951	0.9951
Encrypted correlation	−0.0016	−0.0033	0.0041
Case 2			
Original correlation	0.8325	0.8325	0.8325
Encrypted correlation	0.0010	0.0040	−0.0053
Case 3			
Original correlation	0.5785	0.5785	0.5785
Encrypted correlation	0.0005	0.0009	0.0015
Case 4			
Original correlation	0.9855	0.9855	0.9855
Encrypted correlation	−0.0034	0.0043	−0.0080
Case 5			
Original correlation	0.9224	0.9224	0.9224
Encrypted correlation	0.0006	0.0007	−0.0043
Case 6			
Original correlation	0.9670	0.9670	0.9670
Encrypted correlation	0.0003	0.0008	0.0031
Case 7			
Original correlation	0.8489	0.8489	0.8489
Encrypted correlation	0.0002	−0.0037	−0.0020
Case 8			
Original correlation	−0.9389	−0.9389	−0.9389
Encrypted correlation	0.0001	−0.0017	−0.0018
Case 9			
Original correlation	0.9750	0.9750	0.9750
Encrypted correlation	0.0017	0.0052	−0.0022
Average			
Original correlation	0.7937	0.7937	0.7937
Encrypted correlation	0.0010	0.0027	0.0035

Original correlation, correlation between adjacent samples in original audio; encrypted correlation, correlation between adjacent samples in encrypted audio

in the original audio file are highly correlated. Thus, the correlation value of the original audio samples is closer to one. During the scrambling process, samples are randomly dislocated; hence, the correlation among the neighboring samples is broken, and so the correlation coefficients of the encrypted audio file should be closer to zero. Correlation coefficients are found for the nine test cases using the DFT-, DCT- and IWT-based encryption methods and those values are tabulated in Table 1; it is inferred that transforms have broken the correlation between the neighboring samples. Tabulated values show that the DFT-based scheme offers the least correlation than DCT and IWT schemes. Correlation coefficient denotes the degree of shuffling of the data. Better shuffling results in correlation coefficient with a value closer to 0; while poor shuffling results high correlation which is closer to 1. In the proposed work, two levels of confusion are applied. The first round of confusion is applied in the spatial domain and the second level of confusion is employed in the transform domain. Confusion implemented at these two points as confusion in the spatial domain is simple, and confusion in the transform domain increases the complexity of the system.

5.2 Entropy analysis

Entropy represents the degree of randomness present in the data. In the proposed algorithm, it is found that the entropy values have increased compared to the original value. A higher value of entropy confirms more randomness in the data which resist against the statistical attack. Entropy can be mathematically defined as,

$$\mathrm{En} = -\sum_{i=1}^{n} v(x_i) \times \log_2(x_i), \tag{14}$$

where En is the entropy, x_i represents the ith sample, and v represents the probability of each sample.

Entropy values are entered in Table 2. In the proposed work, diffusion is performed in the transform domain. Entropy values depend on the number of samples present in the audio file. Digitization process is involved in the spatial domain encryption process, and then the maximum randomness will be represented as numerical value (equivalent to the number of bits used represent the sample). During the digitisation process, computations end up with quantisation process, which leads to data loss through round-off operation. Thus, the proposed method approaches the time domain samples directly; so, the maximum entropy is varied concerning the number of samples in the audio file.

5.3 NSCR analysis

The NSCR analysis is performed to find the percentage of change between the original and the encrypted audio. In an ideal encryption scheme, NSCR is 100% to prove that the proposed encryption process modifies the entire samples. It is an effective tool to validate the diffusion process. NSCR is calculated using Eqs. (15), (16),

$$\mathrm{NSCR} = \sum_{i=1}^{N} \frac{V_i \times 100\%}{N}, \tag{15}$$

$$V_i = \begin{cases} 0, & \text{if } C1_i = C2_i \\ 1, & \text{if } C1_i \neq C2_i \end{cases}, \tag{16}$$

where N is the number of samples per audio file, and $C1$ and $C2$ are the original and the encrypted audio data. Table 3 shows the NSCR values for nine test cases of audio files

Table 2 Entropy analysis

Test audio samples	DFT method	DCT method	IWT method
Case 1			
Original entropy	2.6188	2.6188	2.6188
Encrypted entropy	4.2152	2.7819	2.8266
Case 2			
Original entropy	3.2794	3.2794	3.2794
Encrypted entropy	3.7553	3.6778	3.0128
Case 3			
Original entropy	2.8829	2.8829	2.8829
Encrypted entropy	4.1723	3.0079	3.0428
Case 4			
Original entropy	2.5512	2.5512	2.5512
Encrypted entropy	3.1895	2.6576	3.1895
Case 5			
Original entropy	2.8415	2.8415	2.8415
Encrypted entropy	3.9020	3.1895	3.6949
Case 6			
Original entropy	3.3810	3.3810	3.3810
Encrypted entropy	4.0946	3.3020	4.0946
Case 7			
Original entropy	2.1329	2.1329	2.1329
Encrypted entropy	4.1474	4.0946	4.1474
Case 8			
Original entropy	3.0777	3.0777	3.0777
Encrypted entropy	4.3514	4.1474	4.3119
Case 9			
Original entropy	2.4576	2.4576	2.4576
Encrypted entropy	3.9734	4.2502	2.5579
Average			
Original entropy	2.8025	2.8025	2.8025
Encrypted entropy	3.9779	3.4565	3.4309

Original entropy is the entropy of the original audio and encrypted entropy is the entropy of the encrypted audio

for the three different transform methods (DFT, DCT, and IWT).

NSCR is a measure of diffusion. Table 3 proves that confusion in the transform domain offers the desired diffusion. For all three cases, the value obtained is almost 100%, which shows that the proposed diffusion scheme employs the encoding process adequately.

5.4 Error metric analysis

Mean square error (MSE) and peak signa–noise ratio (PSNR) can be used to find qualitative strength of encryption. The metrics are calculated using Eqs. (17) and (18):

$$\mathrm{MSE} = \frac{1}{MN} \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \left[E_{ij} - O_{ij}\right]^2, \tag{17}$$

$$\mathrm{PSNR} = 10 \log_{10} \frac{\mathrm{Max\ \ amplitude}^2}{\mathrm{MSE}}, \tag{18}$$

where E_{ij} is the encrypted data, O_{ij} is the original data; M and N are the total numbers of samples. Table 4 shows the PSNR values for nine test cases of audio files for the three different transform methods (DFT, DCT, and IWT).

The proposed algorithm intended to introduce random noise, which results in a reduction in PSNR. Thus, meaningful audio is converted as meaningless or noisy audio.

Inferences from Table 4 as follows:

1. As the proposed work deals with raw audio data of datatype double, its range is from -1 to $+1$. Hence, the maximum value given as max amplitude in Eq. (18) is taken as 1.
2. Equation (18) ascertains that whenever the MSE increases in denominator due to encryption, the value of PSNR reaches negative.

Table 3 NSCR of the encrypted audio

Test Audio samples	DFT method	DCT method	IWT method
Case 1	100	100	100
Case 2	100	100	99.9985
Case 3	100	100	99.9908
Case 4	100	100	99.8825
Case 5	100	100	99.9962
Case 6	100	100	99.9947
Case 7	100	100	99.9992
Case 8	100	100	99.9985
Case 9	100	100	99.9996
Average	100	100	99.9847

Table 4 Error metrics analysis: PSNR of the encrypted audio

Test audio samples	DFT method	DCT method	IWT method
Case 1	– 59.3124	52.2643	10.9737
Case 2	– 64.5123	45.9977	6.6721
Case 3	– 62.8085	48.1630	– 1.9394
Case 4	– 76.3297	40.6095	6.0354
Case 5	– 82.9409	34.3018	3.6949
Case 6	– 83.5385	27.9107	4.7732
Case 7	– 87.2841	26.9961	0.1851
Case 8	– 93.9774	24.0977	2.5922
Case 9	– 70.1253	19.7359	– 20.4083
Average	– 75.6476	35.5640	6.3638

3. The audio sensitive to human ear ranges from 60 to 120 dB; thus, the obtained PSNR is reduced from their critical decibel levels which ensure that the audio data are entirely encrypted and noisy.

5.5 Key sensitivity analysis

This test is carried out for different audio files, firstly the audio is encrypted with a key using initial values $(x1, y1)$ and then the same audio file is encrypted with the same key with slight modification. The analysis is made and the difference is observed using NSCR for four different audio files. Average NSCR value was found to be 99.8864, which proves that the algorithm is very much sensitive to the slight change in key-value and proposed scheme retains the merits of the chaos.

5.6 Scrambling degree analysis

Scrambling degree is defined as the amount of scrambling or confusion between the original and the encrypted audio. The following equations are used to calculate Scrambling degree,

$$S(j) = \frac{1}{4} * \sum_{i=4}^{N-2} \{4 * D(i) - [D(i-1) + D(i-2) + D(i+1) + D(i+2)]\}, \tag{19}$$

where S is the difference of the signal, $D(i)$ is the ith sample of the audio, N is the total number of samples. Then, the subtraction and the addition of the difference in the original and encrypted audio file are done,

$$V1 = S1 - S2, \tag{20}$$

$$V2 = S1 + S2, \tag{21}$$

where $V1$ is the subtraction of the original and encrypted audio files, and $V2$ is the addition of the original and encrypted file. To get the scrambling degree, Eq. (20) is divided by Eq. (21),

$$\text{Scrambling degree } = \frac{V_1}{V_2}, \tag{22}$$

Scrambling degree lies in the range of [0,1], with '0' being least scrambled and '1' being highly scrambled. A Scrambling degree of 1 denotes complete scrambling and alteration of the encrypted data compared to the original file; while, a Scrambling degree of 0 denotes no change to the encrypted data compared to the original. Table 5 gives the Scrambling degree for nine test cases of audio files for the three different transform methods (DFT, DCT, and IWT).

The average Scrambling degree between the original and encrypted audio of the DFT method is found to be 0.9997, while the average value for the DCT method is found to be 0.9398 and for IWT to be 0.9723. On comparing these values, it is seen that the better scrambling is obtained in the DFT method as shown in Fig. 2.

Table 5 Scrambling degree of the encrypted audio

Test audio samples	DFT method	DCT method	IWT Method
Case 1	0.9998	0. 9757	0.9492
Case 2	0.9997	0.9545	0.9867
Case 3	0.9999	0.9036	0.9979
Case 4	0.9996	0.9446	0.9662
Case 5	0.9997	0.9524	0.9985
Case 6	0.9999	0.9224	0.9478
Case 7	0.9995	0.9374	0.9357
Case 8	0.9999	0.9449	0.9844
Case 9	0.9999	0.9235	0.9758
Average	0.9997	0.9398	0.9723

Figure 3 expresses the distribution of the samples after the encryption process; it is evident that histogram of all the three schemes provides strong masking on the original audio file.

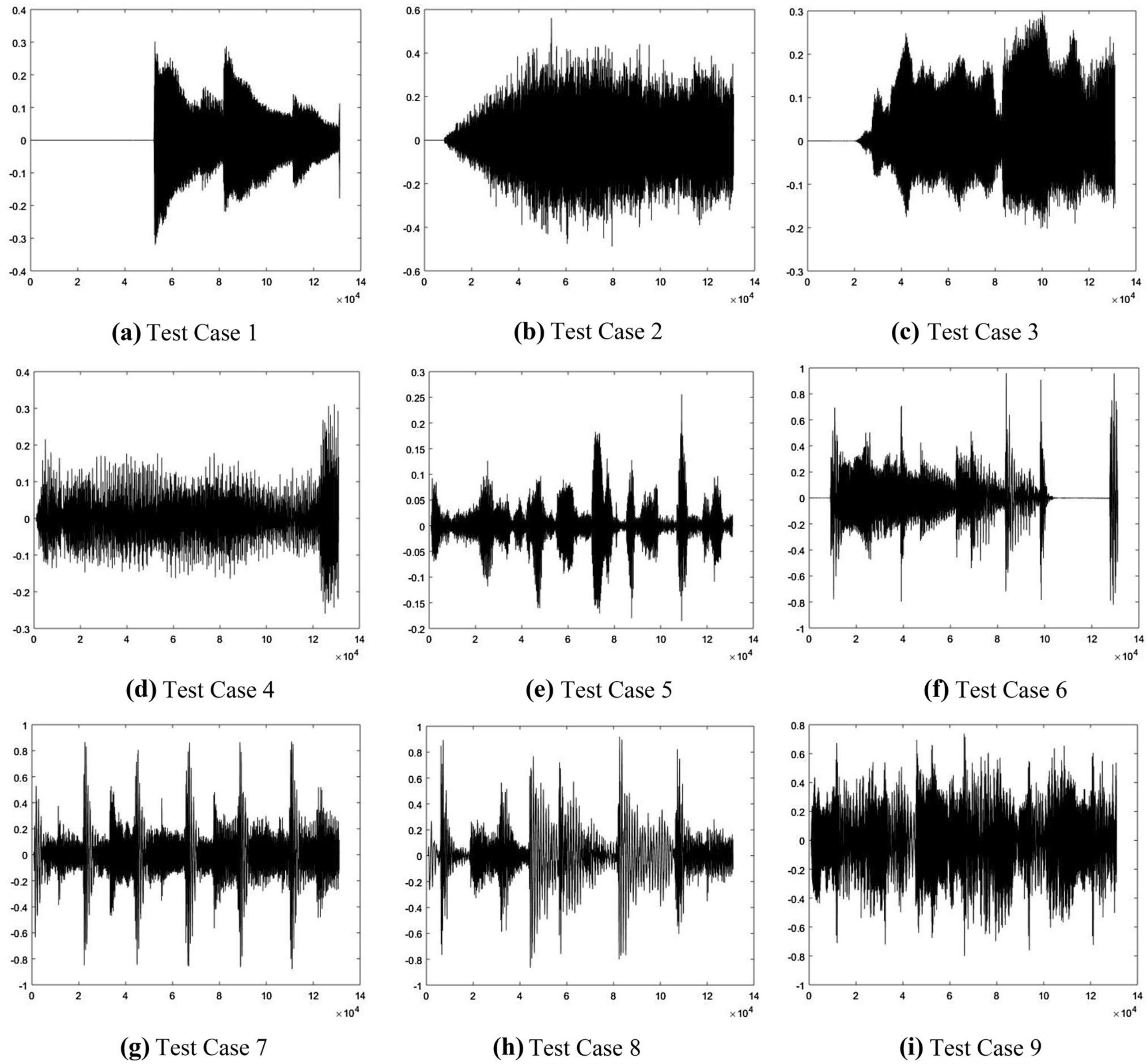
5.7 Computational time analysis

Computational time can be defined as the amount of time required to run an algorithm, and Table 6 gives the time taken to run the algorithm for DFT, DCT and IWT, respectively. Experiments are performed on a system with Intel(R) Core(TM) i3-4000M CPU @ 2.4 GHz, 4 GB RAM equipped with the MATLABR2015a environment.

It is observed that for all three cases, the computational time at an average is found to be 0.7877, 0.8375, 0.9238, respectively, for DFT, DCT and IWT. It can be noted that the algorithm is executed in less than a second, which shows that the speed at which the algorithm is applied is considerably high. Table 7 gives a comparison of the results obtained from the proposed algorithm and existing techniques [19–21, 27, 33]. The values considered for the proposed algorithm in the below table for the DFT, DCT and IWT methods are the best possible results found from the experiments.

5.8 Comparison analysis

On comparing the results from various techniques, it is found that the correlation coefficient of the proposed method with the DFT method found to be most efficient with a value of 10^{-4}. The average entropy of the nine audio files is 2.72 which increased to an average of 4.3045 after applying the encryption algorithm; this shows an increase in randomness caused by the diffusion used in the form of transform-assisted encryption algorithm.



(a) Test Case 1 **(b)** Test Case 2 **(c)** Test Case 3

(d) Test Case 4 **(e)** Test Case 5 **(f)** Test Case 6

(g) Test Case 7 **(h)** Test Case 8 **(i)** Test Case 9

Fig. 2 Time plot of nine audio test cases (**a**–**i**)

NSCR analysis gave a 100% result compared to the other techniques, and the PSNR values also gave batter results for the proposed algorithm. Most of the existing algorithms have employed the conventional XOR operation for diffusion which is vulnerable to the chosen plain text attack. In the proposed scheme, audio samples are confused in the transform domain which replaces the XOR operation. In addtion, confusion is employed in spatial domain using a chaotic map.

From the results, it can be evident that the proposed scheme offers better security through chaos-blend transform domain approach. In most of the work in the comparison, entropy is not analyzed which means that particular methods concentrated in the scrambling process instead of substitution process, but those methods are vulnerable to statistical attack. Some of the existing schemes focus the chaos on attaining keyspace and key sensitivity; few numbers of schemes employ the transform-based scheme but they are failed in the keyspace. Fortunately, the proposed scheme has integrated the merits of chaos and transform domain approach. As a result, correlation also reduced to 10^{-4} which is lesser than the existing schemes. Besides, proposed scheme attained all the significant metric in the desired range which is comparatively better than the existing methods listed in the comparison.

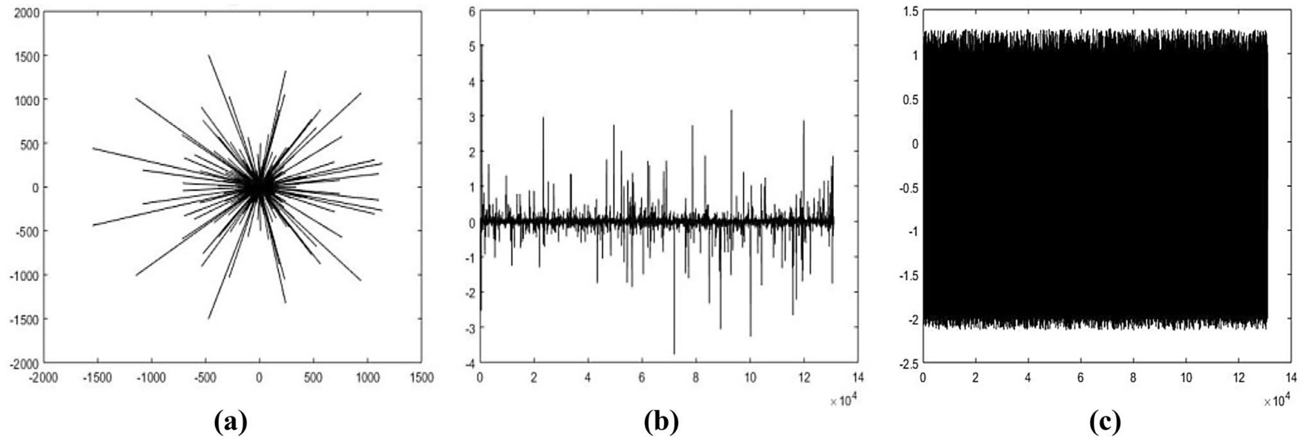


Fig. 3 The encrypted plot of the test case 1, **a** plot of the DFT-based method, **b** represents the DCT-based method, **c** represents the IWT-based method

Table 6 Computational time analysis

Test audio samples	DFT method (in s)	DCT method (in s)	IWT method (in s)
Case 1	0.762	0.799	0.817
Case 2	0.783	0.787	0.809
Case 3	0.788	0.789	0.783
Case 4	0.693	0.792	0.963
Case 5	0.792	0.872	0.997
Case 6	0.795	0.899	0.961
Case 7	0.859	0.769	0.945
Case 8	0.863	0.961	0.963
Case 9	0.755	0.870	1.077
Average	0.7877	0.8375	0.9238

Table 7 Comparison analysis

Scheme	Correlation	Entropy	NSCR	PSNR
DFT method	0.0001	4.3514	100	−59.3124
DCT method	0.0007	4.2502	100	52.2643
IWT method	0.0015	4.3119	100	−20.9737
Ref. [19]	0.0087	N/A	100	N/A
Ref. [20]	0.0011	N/A	99.6399	−10.6357
Ref. [21]	0.0029	N/A	99.8725	−23.89
Ref. [27]	0.0009	N/A	99.9989	−133
Ref. [29]	0.00415	N/A	99.9992	N/A
Ref. [30]	0.00232	N/A	N/A	−12.8495
Ref. [31]	0.00411	N/A	N/A	4.4364
Ref. [33]	0.0491	N/A	99.6399	−44.8
Ref. [34]	0.0223	N/A	99.9997	60.244
Ref. [35]	0.0026	N/A	99.9981	N/A

N/A not analyzed by the authors

5.9 Discussion about the results

From the Results given above, it can be observed that at an average, the correlation coefficient of the encrypted audio is found to be 0.0010, 0.0027 and 0.0035 for DFT, DCT and IWT, respectively, and for the references [19–21, 27, 33] to be 0.0087, 0.0011, 0.0029, 0.0009, 0.00415, 0.00232, 0.00411, 0.0491, 0.0223 and 0.0026, respectively. It can be noted that the values obtained from the proposed algorithm are more un-correlated as they are closer to zero; this is achieved due to the double confusion process applied in the algorithm.

The average entropy values for DFT, DCT and IWT are found to be 3.9779, 3.4565, and 3.4309, respectively, and it shows that the overall randomness increased when compared to the average original entropy of 2.8025. A higher value of entropy represents more randomness in data, which signifies a better diffusion process.

The proposed work utilizes transform domain as a mode of diffusion as other diffusion processes lead to quantization error. NSCR values were found to be at 100 per cent for DFT and DCT and at 99.9837 per cent for IWT and for [19–21, 27, 33] at 100, 99.6399, 99.8725, 99.9989 99.9992, 99.9997, 99.6399 and 99.9981, which depicts the change in one sample with respect to all the samples.

Peak signal–noise ratio value denotes the error metric, and a higher value denotes better encryption quality, while a lower value denotes poor quality as the original, and the encrypted data must be distinct. The obtained average PSNR through the proposed algorithm are − 75.6476, 35.5640, and 6.3638 respectively for DFT, DCT and IWT and references [20, 21, 27, 33] it is found to be − 10.6357, − 23.89, − 133, − 12.8495, 4.4364, − 44.8, and 60.244, it is observed that the values obtained through the proposed algorithm with DFT are much better. Scrambling degree denotes the confusion between the original and encrypted audio, a value of 1 denotes the ideal confusion value and 0 for the worst case. The proposed algorithm consisting of two confusion stages, in spatial and transform domain, which offers high results of 0.9997, 0.9398, and 0.9723 for DFT, DCT and IWT respectively. In [32], it is 0.9818, this shows that the confusion process applied in this work is highly efficient. Additionally, computational time marks the time taken to implement the algorithm on an audio sample; it shows that the algorithm takes less than a second to encrypt a file of 32 kb of data.

6 Conclusion

This paper proposes three layers of security which incorporate the merits of both spatial and transform domain approaches. In the first layer, audio samples are shuffled in the spatial domain, and then second layer security is achieved in transform domain through the confusion operation on transform coefficients which is equivalent to the diffusion in the spatial domain. The third and final layer of security involves the shuffling of the data in the spatial domain. The proposed work results in an average correlation coefficient in the range of 10^{-3}, an average entropy of 3.9779, NSCR of 100%, the Scrambling degree close to 1 and computational time less than one second for a 32 kb of data. From the experimentation, it is concluded that the algorithm which employs DFT is more efficient than the algorithms with DCT and IWT.

Acknowledgements Authors wish to acknowledge SASTRA Deemed University, Thanjavur, India for extending infrastructural support to carry out this work.

References

1. Engel, D., Stütz, T., Uhl, A.: A survey on JPEG2000 encryption. Multimed. Syst. **15**, 243–270 (2009). https://doi.org/10.1007/s00530-008-0150-0
2. Fallahpour, M.: Secure logarithmic audio watermarking scheme based on the human auditory system. Multimed. Syst. (2013). https://doi.org/10.1007/s00530-013-0325-1
3. Yan, D., Wang, R., Yu, X., Zhu, J.: Steganography for MP3 audio by exploiting the rule of window switching. Comput. Secur. **31**, 704–716 (2012). https://doi.org/10.1016/j.cose.2012.04.006
4. Sadek, M.M., Khalifa, A.S., Mostafa, M.G.M.: Video steganography: a comprehensive review. Multimed. Tools Appl. **74**, 7063–7094 (2015). https://doi.org/10.1007/s11042-014-1952-z
5. Fallahpour, M., Megias, D.: High capacity audio watermarking using FFT amplitude interpolation. IEICE Electron. Express. **6**, 1057–1063 (2009). https://doi.org/10.1587/elex.6.1057
6. Cheddad, A., Condell, J., Curran, K., Mc Kevitt, P.: Digital image steganography: survey and analysis of current methods. Signal Processing. **90**, 727–752 (2010). https://doi.org/10.1016/j.sigpro.2009.08.010
7. Cox, I.J., Miller, M.L., Bloom, J.A., Fridrich, J., Kalker, T. eds: Preface to the first edition. In: Digital watermarking and steganography (second edition), pp. xv–xviii. Morgan Kaufmann, Burlington (2008)
8. Lian, S.: Multimedia content encryption. Tech. Appl. (2008). https://doi.org/10.1201/9781420065282
9. Abuturab, M.R.: Color image security system based on discrete Hartley transform in gyrator transform domain. Opt. Lasers Eng. **51**, 317–324 (2013). https://doi.org/10.1016/j.optlaseng.2012.09.008
10. Madain, A., Abu Dalhoum, A.L., Hiary, H., Ortega, A., Alfonseca, M.: Audio scrambling technique based on cellular automata. Multimed. Tools Appl. **71**, 1803–1822 (2014). https://doi.org/10.1007/s11042-012-1306-7
11. Ye, G.: Image scrambling encryption algorithm of pixel bit based on chaos map. Pattern Recognit. Lett. **31**, 347–354 (2010). https://doi.org/10.1016/j.patrec.2009.11.008
12. Zhou, N., Zhang, A., Zheng, F., Gong, L.: Novel image compression-encryption hybrid algorithm based on key-controlled measurement matrix in compressive sensing. Opt. Laser Technol. **62**, 152–160 (2014). https://doi.org/10.1016/j.optlastec.2014.02.015
13. Dworkin, M.J., Barker, E.B., Nechvatal, J.R., Foti, J., Bassham, L.E., Roback, E., Dray Jr., J.F.: Announcing the advanced encryption standard (AES). Technol. Lab. Natl. Inst. Stand. **2009**, 8–12 (2001)
14. Socek, D., Magliveras, S., Ćulibrk, D., Marques, O., Kalva, H., Furht, B.: Digital video encryption algorithms based on correlation-preserving permutations. Eurasip J. Inf. Secur. (2007). https://doi.org/10.1155/2007/52965
15. McDevitt, T., Leap, T.: Multimedia cryptology. Cryptologia. **33**, 142–150 (2009). https://doi.org/10.1080/01611190802300408
16. Mosa, E., Messiha, N.W., Zahran, O., Abd El-Samie, F.E.: Chaotic encryption of speech signals. Int. J. Speech Technol. **14**, 285–296 (2011). https://doi.org/10.1007/s10772-011-9103-7
17. Wang, Y., Wong, K.-W., Liao, X., Chen, G.: A new chaos-based fast image encryption algorithm. Appl. Soft Comput. **11**, 514–522 (2011). https://doi.org/10.1016/j.asoc.2009.12.011
18. Rajaram, G.: Audio encryption using higher dimensional chaotic map. Int. J. Recent Trends Eng. **1**, 103–107 (2009)
19. Ghasemzadeh, A., Esmaeili, E.: A novel method in audio message encryption based on a mixture of chaos function. Int. J. Speech Technol. (2017). https://doi.org/10.1007/s10772-017-9452-y
20. Belmeguenai, A., Ahmida, Z., Ouchtati, S., Djemii, R.: A novel approach based on stream cipher for selective speech encryption.

Int. J. Speech Technol. (2017). https://doi.org/10.1007/s10772-017-9439-8
21. Farsana, F.J., Gopakumar, K.: A novel approach for speech encryption : Zaslavsky map as Pseudo random number generator. Procedia Procedia Comput. Sci. **93**, 816–823 (2016). https://doi.org/10.1016/j.procs.2016.07.302
22. Eldin, S.M.S., Khamis, S.A., Hassanin, A.-A.I.M., Alsharqawy, M.A.: New audio encryption package for TV cloud computing. Int. J. Speech Technol. **18**, 131–142 (2015). https://doi.org/10.1007/s10772-014-9253-5
23. Rao, R.: Efficient audio encryption algorithm for online applications using hybrid transposition and multiplicative non binary system. In: Presented at the (2013)
24. Sheela, J., Kaggere, S., Tandur, S.: A novel audio cryptosystem using chaotic maps and DNA encoding. J. Comput. Netw. Commun. **2017**, 1–12 (2017). https://doi.org/10.1155/2017/2721910
25. Wang, H., Hempel, M., Peng, D., Wang, W., Sharif, H., Member, S., Chen, H.: Index-based selective audio encryption for wireless multimedia sensor networks. IEEE Trans. Multimed. **12**, 215–223 (2010)
26. Kwon, G.-R., Wang, C., Lian, S., Hwang, S.: Advanced partial encryption using watermarking and scrambling in MP3. Multimed. Tools Appl. **59**, 885–895 (2012). https://doi.org/10.1007/s11042-011-0771-8
27. Farsana, F.J., Devi, V.R., Gopakumar, K.: Applied computing and informatics an audio encryption scheme based on fast walsh hadamard transform and mixed chaotic keystreams. Appl. Comput. Informatics. **2019**, 1–11 (2019). https://doi.org/10.1016/j.aci.2019.10.001
28. Yang, Y.G., Tian, J., Sun, S.J., Xu, P.: Quantum-assisted encryption for digital audio signals. Optik (Stuttg). **126**, 3221–3226 (2015). https://doi.org/10.1016/j.ijleo.2015.07.082
29. Lima, J.B., da Silva Neto, E.F.: Audio encryption based on the cosine number transform. Multimed. Tools Appl. **75**, 8403–8418 (2016). https://doi.org/10.1007/s11042-015-2755-6
30. Alwahbani, S., Bashier, E.: Speech scrambling based on chaotic maps and one time pad. In: Presented at the (2013)
31. Liu, H., Kadir, A., Li, Y.: Audio encryption scheme by confusion and diffusion based on multi-scroll chaotic system and one-time keys. Optik (Stuttg). **127**, 7431–7438 (2016). https://doi.org/10.1016/j.ijleo.2016.05.073
32. Ballesteros, D., Renza, D., Camacho, S.: High Scrambling degree in audio through imitation of an unintelligible signal. In: Presented at the (2016)
33. Belazi, A., Khan, M., Abd, A.A., Belghith, E.S.: Efficient cryptosystem approaches: S-boxes and permutation—substitution-based encryption. Nonlinear Dyn. (2016). https://doi.org/10.1007/s11071-016-3046-0
34. Sathiyamurthi, P., Ramakrishnan, S.: Speech encryption using chaotic shift keying for secured speech communication. Eurasip J. Audio Speech Music Process. (2017). https://doi.org/10.1186/s13636-017-0118-0
35. Chang, D., Li, Z., Wang, M., Zeng, Y.: A novel digital programmable multi-scroll chaotic system and its application in FPGA-based audio secure communication. AEU Int. J. Electron. Commun. **88**, 20–29 (2018). https://doi.org/10.1016/j.aeue.2018.03.007