An Audio Encryption Scheme Based on Empirical Mode Decomposition and 2D Cosine Logistic Map

Alenrex Maity, and Bibhas Chandra Dhara

***Abstract*—In this work, we have proposed an audio encryption method. The proposed method is audio signal sensitive as the hash value of the given signal is computed using SHA3-512, which returns a significantly large key of size 512-bit. This key is used to set the different parameters. This work suggests a 2D Cosine Logistic Map (2DCLM) by fusing the Cosine map with the Logistic map. The proposed 2DCLM functions well under chaos. The given audio signal is scrambled with the help of the hash value. The scrambled signal is decomposed by Empirical Mode Decomposition (EMD); before using the EMD, the signal is segmented into a 2D signal to reduce the time complexity of the EMD method. The residuals given by EMD and the stream generated by 2DCLM are XOR-ed to encrypt the signal. Finally, the 2D encrypted signal is transformed into a 1D encrypted audio signal. The efficacy of the present method is evaluated with the help of different audio streams. The findings of the simulation and comparison indicate that the suggested technique may deliver effective encryption results while thwarting cryptographic assaults.**

Link to graphical and video abstracts, and to code: https://latamt.ieeer9.org/index.php/transactions/article/view/8454

***Index Terms*—Encryption, Audio encryption, Security, Chaotic map, 2D chaotic map, Empirical Mode Decomposition**

I. Introduction

Multimedia data security is now becoming more significant for various applications, including pay-per-view TV, industrial or military imaging systems, medical imaging, confidential audio/video conferencing, online transactions, passwords, and digital signatures. Audio data also plays a significant and prevalent role in multimedia systems. In several fields, including confidential voice conferencing, military systems, education, and telephone banking, digital audio communications are seeing a remarkable upswing and importance [1]. The security and resilience of those applications would be greatly influenced by the weaknesses in audio data security, causing a significant impact. Several techniques have been developed to protect the information. Audio signals have a larger size and a stronger correlation among samples. Therefore, conventional encryption techniques are not appropriate for encrypting audio signals since they have a lot of power consumption and complex computations. As a result, one significant emphasis of multimedia security researchers is developing audio encryption techniques that are both secure and fast. The chaos-based approaches are the most efficient encryption methods for tackling redundant and large amounts of audio data due to their highly secure and quick encryption [2].

In this study, we have proposed an audio encryption technique using Empirical Mode Decomposition (EMD) and a hybrid chaotic map. In this method, from the original audio signal, a hash value is calculated by using SHA3-512 [3]. From this hash value, the initial parameters of the proposed hybrid chaotic map, a 2D Cosine Logistic Map (2DCLM), are defined, and the hash value is also used to scramble the audio signal. Then, a 2D signal (represented by a 2D matrix) is computed from the scrambled audio signal. Each row of the matrix (considered a signal) is undergone through the EMD, resulting in some Intrinsic Mode Functions (IMFs) and one residue signal. The stream generated by 2DCLM encrypts the residue signal. The sum of all the IMFs and the encrypted residue gives the encrypted signal. The major contributions of this work are highlighted below:

1) The suggested approach is sensitive to plaintext.
2) A hybrid chaotic system 2DCLM is suggested, and its dynamic behavior is also examined.
3) EMD and 2DCML-based audio encryption technique is proposed.
4) The proposed approach performs well, according to experimental results.

The remaining part of this article is structured as follows: The existing works related to audio encryption are discussed in Section II. The background of the 2DCLM and EMD are demonstrated in Section III. The proposed method is presented in Section IV. Section V presents the experimental findings and performance analyses. At the end, in Section VI, this paper is concluded.

II. Related Works

A very brief introduction to the state-of-the-art (SoA) of audio encryption is studied in this section.

Albin *et al.* [4] proposed an audio encryption technique based on a discrete wavelet transform and a multi-scroll chaotic system, whereby the initial parameters of the chaotic system were generated using SHA-256. In this method, only the low-frequency component of the DWT is encrypted by the chaotic system. A speech encryption method based on an encoding scheme, DNA encryption approach, and a permutation function was suggested by Kate *et al.* [5] to offer a quick and secure audio encryption solution. Naskar *et al.* [6] presented a block cipher technique that encrypts audio using DNA encoding and a logistic-chaotic map. This method also included channel shuffling to improve the security of the

Alenrex Maity and Bibhas Chandra Dhara are with the Department of Information Technology, Jadavpur University, Kolkata, West Bengal - 700106, India (e-mails: alenrex8@gmail.com and bcdhara@gmail.com).

encrypted data. All ciphering and shuffling processes are based on the 32-byte key, which is updated during the encryption of each block of secret audio data. Wang and Su [7] proposed an audio encryption technique based on DNA coding and a chaotic system. First, chaotic sequences are created using the PWLCM technique. SHA-256 generates the initial value of the PWLCM and relies heavily on plaintext. Second, a cyclic shift is used for scrambling the audio. Lastly, the DNA matrix created by dynamic coding is finally XORed with the key to achieve diffusion. Shah *et al.* [8]. introduced an audio encryption technique based on substitution and permutation. The Mobius transformation is used in this approach to create S-boxes for the substitution, and the Henon chaotic map, which conducts pixel-wise permutation, is used for the permutation. Farsana and Devi [9] suggested an audio encryption method based on a discrete modified Henon map and a modified Lorenz-Hyperchaotic system; the modified Heanon map is used for the permutation operation, whereas the modified Lorenz-Hyperchaotic system is used for the substitution. Zoghdy *et al.* [10] introduced a communication model over Orthogonal frequency division multiplexing (OFDM) by using three different chaotic maps: a 2D Baker map, a Logistic map, and a Standard Chaotic map for encrypting and transferring audio signals. Kaur *et al.* [11] presented a technique that uses cryptographic protocols and numerous chaotic maps to encrypt the voice signal. The cubic map is used to split and jumble the input signal into four parts. The scrambled signal is processed through various one-dimensional chaotic maps, including cubic, logistic, skew-tent, and quadratic maps, to render it impervious to assaults. Based on the Chen memristor chaotic system, Dai *et al.* [12] suggested an audio encryption technique where the original chaotic system's resilience is increased by the Chen memristor chaotic system based on a magnetron titanium dioxide memristor, which also significantly extends the keyspace of the algorithm and effectively widens the system's parameter range. Adhikari and Karforma [13] presented an audio encryption scheme using a Henon map and a Tent map. A pseudo-random number sequence is generated as the secret key by XORing the sequence generated by the Henon map and tent map. XORing the secret key with the original audio file creates a cipher audio file. Khalee and Abduljaleel [14] suggested an algorithm to encrypt the speech. It depends on the k-means clustering and quantum chaotic map, which are used to produce keys. Additionally, two scrambling phases were employed: the first stage utilized bits and the suggested algorithm (binary representation scrambling, or BiRS), while the second stage used k-means and the suggested method (block representation scrambling, or BlRS). Nasreen and Muthukumar [15] proposed an audio encryption technique that relies on a dynamical system incorporating fractional derivatives, demonstrating chaotic behavior across a spectrum of fractional orders and parameter values. Hazaimeh *et al.* [16] proposed a block cipher technique to encrypt speech based on a Jacobian elliptic map. By merging an infinite collapse map (1D-ICM) with a logistic map, Wu *et al.* [17] suggested an audio encryption technique based on a 2D-logistic-nested-infinite-collapse (2D-LNIC). The keystream for the audio encryption technique is produced using 2D-LNIC, and the encryption procedure entails a simultaneous process of scrambling and diffusion. A chaos-based method for audio encryption utilizing the double DNA operation, the Sine-Cosine (SC) map, and the Logistic Sine-Cosine (LSC) map was presented by Kuma and Dua [18]. The suggested method is broken down into three sections: First, the audio is permuted using the Sine-Cosine map to execute the permutation. In the second phase, the double DNA addition encoding is used to jumble the permuted audio samples. Eventually, one of the chaotic maps from the SC Chaotic map and the LSC map is chosen to do dynamic diffusion.

Numerous chaotic systems, including the logistic map, have been widely used due to their high degree of efficiency and simplicity, according to the literature [6], [10], [11]. While there are certain benefits to these chaotic cryptosystems, there are also some drawbacks, such as inadequate security and a tiny keyspace. To address the issues mentioned above, this paper suggests a new cryptosystem. According to security analysis and experimental findings, the proposed encryption strategy based on the 2DCLM is advantageous regarding both keyspace and security. The 2DCLM, which is employed in this study, is briefly discussed in the following section.

III. Background

The proposed encryption technique for audio signals is based on EMD and 2DCLM. The 2DCLM and EMD are described in the following subsections.

A. 2D Cosine Logistic Map (2DCLM)

Chaotic methods are commonly used due to their superior performance and specific characteristics, including sensitivity to the initial condition, unpredictability, state of ergodicity, pseudorandomness, etc. [17], which perfectly fulfill the fundamental needs of cryptography. In this work, 2DCLM is proposed by combining the Cosine map and the Logistic map. The expressions of the Cosine map and Logistic map are given in Eq (1).

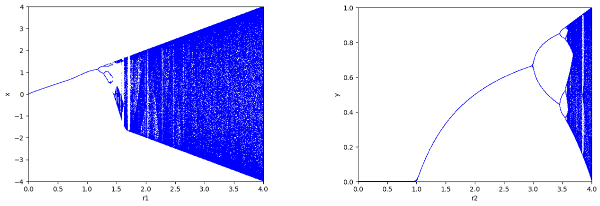
$$\begin{aligned} \text{Cosine map: } & f(r_1, x_n) = x_{n+1} = r_1\ cos(x_n(1 - x_n^2)) \\ \text{Logisitc map: } & g(r_2, y_n) = y_{n+1} = r_2\ y_n(1 - y_n) \end{aligned} \tag{1}$$

where r_1, r_2 are the control parameter, and $x_0, y_0 \in [0, 1]$ is the initial value. The value of r_1, r_2 significantly impacts how chaotically the map behaves. The Cosine map settles into a stable point for $0.0 \leq r_1 \leq 1.2$ as shown by the bifurcation diagram in Fig. 1(a). It exhibits its erratic behavior and chaotic character for $r_1 \geq 1.3$. The logistic map always collapses to zero for $0.0 < r_1 < 1.0$, as shown by the bifurcation diagram in Fig. 1(b). The logistic map settles into a stable point for $1.0 \leq r_1 \leq 3.0$. It exhibits its erratic behavior and chaotic character for $3.4 \leq r_1 \leq 4$.

Using the Cosine map and the Logistic map, the hybrid 2DCLM is defined as given in Eq. (2).

$$h(r_1, r_2, x_n, y_n) = \begin{cases} x_{n+1} = r_1\ cos(x_n(1 - x_n^2)) \\ y_{n+1} = r_2\ cos(y_n(1 - x_n)) \end{cases} \tag{2}$$

where $x_0, y_0 \in [0, 1]$ is the initial state and the control parameter $r_1, r_2 > 0$ those have a significant impact on



(a) (b)

Fig. 1. (a) Bifurcation diagram of Cosine map, (b) Bifurcation diagram of Logistic map.

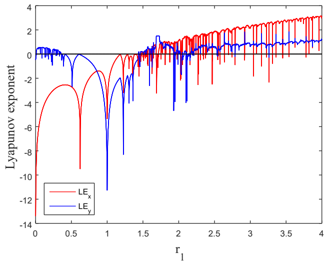
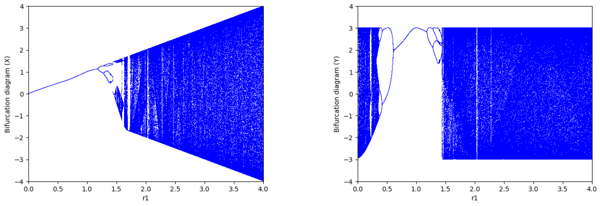


Fig. 2. Lyapunov exponents of 2DCLM at $r_2 = 3$



(a) (b)

Fig. 3. (a) Bifurcation diagram for sequence X at $r_2 = 3$, (b) Bifurcation diagram for sequence Y at $r_2 = 3$.

how chaotically the map behaves. The 2DCLM shows chaotic behavior for all values of the control parameter $r_1 \geq 1.482$, and $r_2 \geq 2.834$, whose chaotic range is superior to the Logistic map.

The Lyapunov exponent (LE) is a powerful tool for measuring the sensitivity of the chaos system to slight changes in the initial condition. A negative LE denotes that the dynamical system is stable, whereas a positive LE indicates the exponential divergence of neighboring initial points over a limited number of iterations. A chaotic system with two positive LEs is also regarded as a hyperchaotic system. Based on Fig. 2, it can be concluded that the proposed 2DCLM has a hyper-chaotic nature. The bifurcation diagram is also used to analyze the chaotic behavior of chaotic maps. From the bifurcation diagram of the proposed 2DCLM (Fig. 3 (a), (b)), it is clear that the new map contains a sizable chaotic zone with just a small number of periodic windows. The randomness of the proposed 2DCLM is evaluated concerning the NIST Randomness Test.

NIST Randomness Test: We used the SP800-22 test criteria to identify randomness, detailed explanation of the test is available in [19], [20]. The SP800-22 test consists of 188 minor tests and 15 significant tests. When the P-value > 0.01, the randomness test is passed; if not, the time series is not random. P-values are a crucial tool for evaluating the performance of time series. The conditions for sequence randomness are satisfied if all tests are passed. We have generated 100×10^6 points by the 2DCLM with initial conditions $x_0 = 0.5, y_0 = 0.8, r_1 \in [3.67, 4], r2 = 3$ for NIST test. Table I displays the result for every test in the NIST suite, along with the corresponding P-value. The average result is given for multiple occurrences of the same tests. The table shows that the sequences generated by 2DCLM passed all the tests, exhibiting good pseudo-randomness.

TABLE I
NIST RANDOMNESS TEST OF THE 2DCLM

	X sequence		Y sequence	
Statistical test	P-value	Result	P-value	Result
Frequency (Monobit)	0.6397	Passed	0.2661	Passed
Frequency Test within a Block	0.0901	Passed	0.7045	Passed
Runs	0.6099	Passed	0.6677	Passed
Longest-Run-of-Ones in a Block	0.6658	Passed	0.4106	Passed
Binary Matrix Rank	0.3282	Passed	0.1564	Passed
Spectral	0.0481	Passed	0.4744	Passed
Non-overlapping Template Matching	0.2485	Passed	0.8218	Passed
Overlapping Template Matching	0.5549	Passed	0.3976	Passed
Maurer's "Universal Statistical"	0.2730	Passed	0.1377	Passed
Linear Complexity	0.2793	Passed	0.8314	Passed
Statistical	0.7878	Passed	0.4917	Passed
	0.6114	Passed	0.6686	Passed
Approximate Entropy	0.1598	Passed	0.8069	Passed
Cumulative Sums (Cusums)	0.6811	Passed	0.4384	Passed
	0.6579	Passed	0.6627	Passed
Random Excursions	0.0537	Passed	0.0208	Passed
Random Excursions Variant	0.4482	Passed	0.0193	Passed

B. Empirical Mode Decomposition

EMD is a technique for decomposing a signal without leaving the time domain [21]. The signals may be decomposed into a collection of simple and intrinsic oscillations in a particular fashion on a dynamic feature time scale without the requirement of any prior system knowledge. This adaptive and effective decomposition based on the local characteristic time scale of the data is appropriate for nonlinear and non-stationary processes, resulting in a collection of Intrinsic Mode Functions (IMFs) [22] and a residue signal. An IMF is a signal that meets the following two criteria:

1) In the dataset, the difference between the total number of extrema (maxima and minima) and the number of zero crossings is one at most.
2) At any point, the average of the envelopes defined by local maxima and local minima is always zero.

Finding IMF is a time-consuming process. The EMD method, for a signal $Org(t)$, is as described below:

1) $s(t) = Org(t)$, $i = 1$
2) Determine all the local extrema (maxima and minima) of the input signal $s(t)$.
3) Apply cubic spline interpolation to the maxima and minima to generate the upper and lower envelopes $\{e_u, e_l\}$
4) The local mean $a_i(t)$ of the signal $s(t)$ is calculated as $a_i(t) = \frac{e_u+e_l}{2}$
5) The difference $d_i(t) = s(t) - a_i(t)$ may be regarded as the basic description of an IMF

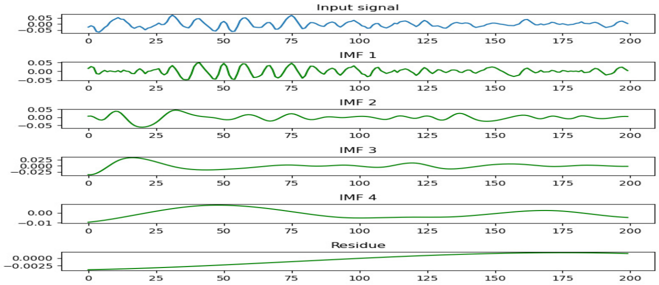


Fig. 4. EMD decomposition of an audio signal.

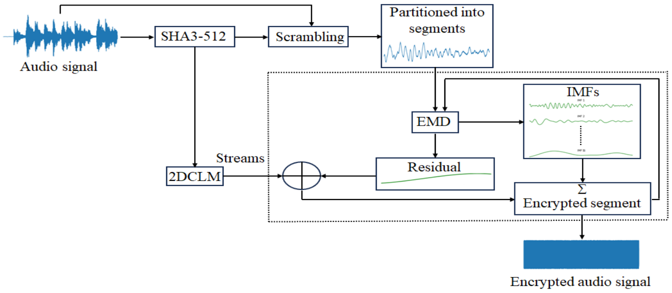


Fig. 5. Proposed audio encryption process.

6) If $d_i(t)$ is an IMF then $D_i(t) = d_i(t)$, Goto Step 7
Else $s(t) = d_i(t)$ Goto Step 2.
7) Compute $Org(t) = Org(t) - D_i(t)$ and $s(t) = Org(t)$
If there is no discernible variation in $s(t)$, then final residue is $r(t) = s(t)$ and STOP
Else $i = i + 1$, Goto Step 2.

So, the original signal can be represented as given in Eq. (3), where n is the number of IMFs and $r(t)$ is the final residue.

$$Org(t) = \sum_{i=1}^{n} D_i(t) + r(t) \tag{3}$$

Fig. 4 depicts the decomposition of an audio sample using EMD. The figure shows that the input signal is decomposed into four IMFs and one residue. It is also shown that the frequency of IMF gradually decreases, and, finally, a smooth signal is found in the residue. We encrypt the residue signal, which is the skeleton of the original signal, $r(t)$ to $r'(t)$ in the encryption step. The encrypted audio signal is $Org'(t)$ is obtained by adding the signal $r'(t)$ with all IMFs (use Eq. (3)). Again, if $Org'(t)$ is decomposed, EMD returns the same set of IMFs (because each IMF has satisfied certain conditions and $Org'(t)$ is obtained by simple addition) and the residue $r'(t)$. So, during decryption, using the same parameters from $r'(t)$, the residue signal $r(t)$ can be obtained, and hence the original audio signal will be deciphered. The proposed encryption and decryption processes are discussed in the following section.

IV. Proposed Method

In this article, we have a symmetric audio encryption method. The proposed encryption and decryption techniques are described in the subsequent subsections.

$$\begin{cases} x_0 = (H[1:64] \oplus H[65:128])/2^{64} \\ y_0 = (H[129:192] \oplus H[193:256])/2^{64} \\ r_1 = (H[257:320] \oplus H[321:384])/2^{64} + 2.9 \\ r_2 = (H[385:448] \oplus H[449:512])/2^{64} + 2.9 \end{cases} \tag{4}$$

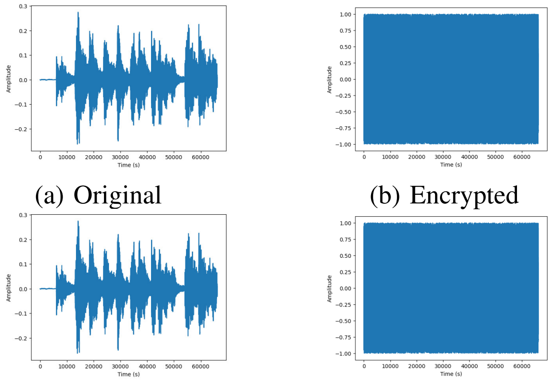
A. Encryption Algorithm

Fig. 5 shows a block schematic of the suggested encryption mechanism. First, a hash value H of 512 bits from the given signal is computed using SHA3-512. The hash value H is used to i) scramble the given audio signal; and ii) compute four parameters $\{x_0, y_0, r_1, r_2\}$ (using Eq. (4)), which will be used to define the initial state and parameters of 2DCLM. So, the key of the proposed method is the hash value H. Since the length of the audio signal is too high, to reduce the time complexity of the EMD, we have divided the signal into some segments and formed a 2D matrix in this experiment. Each row (or segment) of the 2D matrix passes through the EMD process, generating IMFs and residue r. The residue is XOR-ed with the stream generated by 2DCLM, giving the encrypted version of the residue. Next, the encrypted residue is added with corresponding IMFs. The above steps are iterated a number of times (say, itr) and result in an encrypted 2D version of the signal. Finally, from the encrypted 2D matrix, we compute the 1D encrypted audio signal. The algorithmic sketch of the proposed encryption method is given in **Algorithm 1: Encryption()**.

```
Algorithm 1 : Encryption $(S_{org}, S_{enc}, itr)$
Input: Audio signal $S_{org}$, number of iterations $itr$
Output: Encrypted audio signal $S_{enc}$
Let $N = |S_{org}|$
$H \leftarrow$ SHA3-512 $(S_{org})$
Compute $\{x_0, y_0, r_1, r_2\}$ using Eq. (4)
$\{X, Y\} \leftarrow$ 2DCLM using Eq. (2)
$K = mod(((X \oplus Y) \times 10^{16}), 256)$
$\Pi \leftarrow$ RandPermute $(H)$
$S' \leftarrow$ Scramble $(S_{org}, \Pi)$
$Mat_{M_1 \times M_2} \leftarrow S'_{1 \times N}$
$K'_{M_1 \times M_2} = K_{1 \times N}$
for $p = 1$ to $itr$ do
    for $i = 1$ to $M_1$ do // for each row of the Mat
        $a.$ $sig = Mat[i,:]$
        $b.$ $\{IMF_1, IMF_2, ..., IMF_n, r\} \leftarrow$ EMD $(sig)$
        $c.$ $r' = r \oplus K'[i,:]$
        $d.$ $sig' = \sum_{j=1}^{n} IMF_j + r'$
        $e.$ $Mat[i,:] = sig'$
$S'_{1 \times N} \leftarrow Mat_{M_1 \times M_2}$
$S_{enc} \leftarrow S'$
Return $S_{enc}$
```

B. Decryption Algorithm

As far as symmetric encryption is concerned, the decryption technique is the reverse encryption procedure. Here, the encrypted audio is handled using the reverse encryption process with the same initial conditions and the control parameters of the 2DCLM using the hash value H.



(a) Original (b) Encrypted

(c) Decrypted (correct key) (d) Decrypted (wrong key)

Fig. 6. Performance of the proposed method on ‘CantinaBand3.wav (CB)’ signal.

V. Experimental Results and Analysis

The suggested approach is implemented in the Python 3.10.9 software on a platform with an Intel(R) Core(TM) i5-10300H Processor, CPU @ 2.50GHz, and 8 GB of RAM. The testing was done by utilizing the most often used test audios: BabyElephantWalk60.wav (BE), CantinaBand3.wav (CB), Fanfare60.wav (FF), gettysburg10.wav (GB), Imperial-March60.wav (IM), PinkPanther60.wav (PP), preamble10.wav (PB), StarWars60.wav (SW) [23]. In our experiment, we set $iter = 1$, and the initial parameters of the 2DCLM, $x_0, y_0, r_1,$ and r_2 are determined by Eq. (4). An audio encryption method must be able to encrypt any audio into cipher audio and obtain the same plain audio through the decryption process. The outcome of the proposed methodology is presented in Fig. 6. Due to the space problem, we give the result on the signal CB. From this figure, it is clear that no one can guess anything about the original signal from the encrypted signal (see Fig. 6(b)). The audio signal can be recovered exactly (see Fig. 6(c)) by using the actual key. However, it cannot find the original signal using a slightly modified key. Therefore, the proposed method can properly encrypt and decrypt the audio signals. To analyze the performance of the proposed method, different performance metrics like key space, entropy, correlation coefficient, histogram analysis, plaintext and key sensitivity, robustness against the differential attack and noise attacks, and computational time analysis are used.

A. Key Space

A secure encryption method must have a large key space to resist brute-force attacks. The size of the key space directly impacts the security of a cryptographic technique. A reliable encryption technique should have a key space of at least 2^{100} [24]. In the suggested approach, the parameters and initial values of the 2DCLM and the permutation for the scrambling process are determined from the hash value, H, of the original audio signal using SHA3-512. Therefore, the key space of the suggested approach is defined by 512 bits, and as $2^{512} \gg 2^{100}$, our method can withstand all sorts of brute-force assaults.

TABLE II
Entropy Analysis of the Proposed Method

Audio	Entropy	
	Original	Encrypted
BE	7.021523	7.999134
CB	6.794367	7.998174
FF	6.597602	7.998144
GB	6.483891	7.994454
IM	7.248650	7.999515
PP	6.931921	7.999811
PB	6.271621	7.992942
SW	7.554577	7.999905
Avg.	6.863019	7.997760
Ref [1]	-	7.943833
Ref [8]	3.795475	7.997887

B. Entropy Analysis

Entropy [25] is the most crucial metric for measuring unpredictability or randomness. The mathematical expression for calculating the entropy is given in Eq. (5).

$$I(s) = \sum_{i=0}^{2^M-1} p(s_i) \log \frac{1}{p(s_i)} \tag{5}$$

where $p(s_i)$ is the probability of s_i and M is the number of bits used to represent s_i. If a source is truly random and generates 2^M symbols, then its entropy is M, i.e., if entropy is M, then one may assume that the source is random. However, it is not always true (for example, consider a signal of length $2^M \times k$ and there are the runs of length k for each value, then the signal is not random, but its entropy will be M). From Table II, we note that the entropy of the original audio is close to 7, whereas our method gives almost 8 (the ideal value). So, we may consider the encrypted signal is quite random (as the encrypted signal covers the entire domain, see Fig. 6(b)). Therefore, it is difficult for an attacker to decipher the original audio from the encrypted one. Compared with the existing method, we have similar performance.

C. Correlation Coefficient Analysis

The encryption strength of any cryptosystem may be evaluated using the correlation coefficient of the original and the cipher audio signal. The coefficient is computed by Eq. (6).

$$r_{\alpha\beta} = \frac{cov(\alpha,\beta)}{\sqrt{D(\alpha)}\sqrt{D(\beta)}},$$

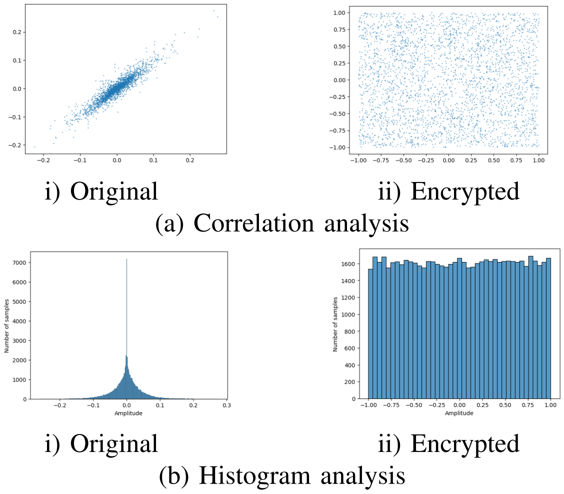
$$\begin{cases} cov(\alpha,\beta) = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}(\alpha_i - E(\alpha))(\beta_i - E(\beta)), \\ D(\alpha) = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}(\alpha_i - E(\alpha))^2, \quad E(\alpha) = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\alpha_i \end{cases} \tag{6}$$

where α and β are the adjacent sample values of the audio signal, N is the total number of duplets (α,β) retrieved from the audio signal, $E(\alpha)$ is the mean value, $D(\alpha)$ is the variance, $cov(\alpha,\beta)$ is the covariance and $r_{\alpha\beta}$ is the correlation.

In this study, we chose random 50000 pairs of adjacent samples from the original and encrypted audio to estimate the correlation between nearby samples. The correlation of the

TABLE III
CORRELATION ANALYSIS

Audio	Correlation	
	Original	Encrypted
BE	0.927073	0.000020
CB	0.934917	-0.000152
FF	0.973813	0.000006
GB	0.974584	-0.000424
IM	0.948969	0.000014
PP	0.963028	-0.000013
PB	0.955982	-0.000387
SW	0.934577	0.000431
Avg.	0.951618	0.0001809
Ref [1]	-	0.008733
Ref [5]	-	0.011358
Ref [6]	1.000000	0.035300
Ref [7]	0.888833	0.000766
Ref [8]	0.965575	0.002262
Ref [18]	0.971000	-0.002600
Ref [26]	0.998100	0.010400
Ref [27]	0.946400	0.003372
Ref [28]	-	0.000280



i) Original ii) Encrypted

(a) Correlation analysis

i) Original ii) Encrypted

(b) Histogram analysis

Fig. 7. Correlation and Histogram analysis.

plaintext and ciphertext is given in Table III. The table shows that the correlation of the original signals is strong (close to 1). For the encrypted signals, it is very low, close to zero (i.e., the correlation between the adjacent samples is almost negligible). The table also depicts that our approach performs like the SoA methods.

Fig. 7(a) depicts the scatter plot of the chosen sample pairs. For the original audio signal, it is obvious that the samples are grouped diagonally ($y = x$ direction), demonstrating a high degree of correlation. In the case of an encrypted signal, the samples are dispersed over the whole region, indicating a low correlation between them. Therefore, the encrypted audio produced by the proposed technique withstands statistical attacks based on the correlation coefficient analysis.

D. Histogram Analysis

A histogram plots data based on frequency. Good encryption should have evenly-spaced data frequencies. The histogram of audio signals is illustrated in Fig. 7(b). The histogram of the plain audio signal is confined within a narrow band and uneven distribution. On the other hand, the histogram of the cipher audio is uniformly dispersed and spread across the entire domain. Thus, our method can resist statistical attacks.

E. Plaintext Sensitivity

An encryption method is plaintext sensitive; if the plaintext is changed even a single bit, two entirely different ciphertexts will be generated, one from the original plaintext and the other from the modified plaintext. Here, a different hash value will be obtained if a single bit of the plaintext is complemented. So, the initial parameters of 2DCLM will be changed according to the hash value, and the streams generated by 2DCLM will be different. Two metrics, NSCR (Number of Samples Change Rate) and UACI (Unified Average Changing Intensity), are used to compare two encrypted audio files. The mathematical formula for the NSCR and UACI is given in Eq. (7).

$$NSCR = \frac{\sum_{i=1}^{N} D(i)}{N} \times 100\%$$

$$UACI = \frac{1}{N} \sum_{i=1}^{N} \frac{|E_1(i) - E_2(i)|}{2^M - 1} \times 100\% \tag{7}$$

where $D(i)$ is defined as

$$D(i) = \begin{cases} 0 & \text{if,} \quad E_1(i) = E_2(i) \\ 1 & \text{if,} \quad E_1(i) \neq E_2(i) \end{cases}$$

where N is the length of the audio signal. NSCR and UACI should be at 100% and 33.33%, respectively, as the ideal values [12]. Here, we execute the program for each signal ten times (by complementing different selected bits of plaintext). Their average value is reported in Table IV. This table shows that the value of the NSCR and UACI of our method is close to the ideal value, so the encryption method is plaintext-sensitive.

F. Key Sensitivity

Keys pace analysis is inadequate to ensure stability; hence, a resilient cryptosystem must achieve key sensitivity. Key sensitivity means that a minor change, say, only one bit is complemented, in the key should generate a completely different cipher signal, i.e., if a key K_2 is derived from key K_1. These two keys generate two cipher audio E_1 and E_2, and these two ciphers should be completely different. Here, we also execute the program for each signal ten times (by complementing different selected bits of the key (H)). The average value of NSCR and UACI is reported in Table IV. This table shows that our method returns values closer to the ideal values of NSCR and UACI than those produced by the method in [7]. So, the encryption method is key sensitive. We also observed that the decryption process is key sensitive; a noisy signal is given if a wrong key is used to decipher the signal (see Fig. 6(d)). So, in both ways, our method is key sensitive.

TABLE IV
Plaintext and Key Sensitivity Analysis

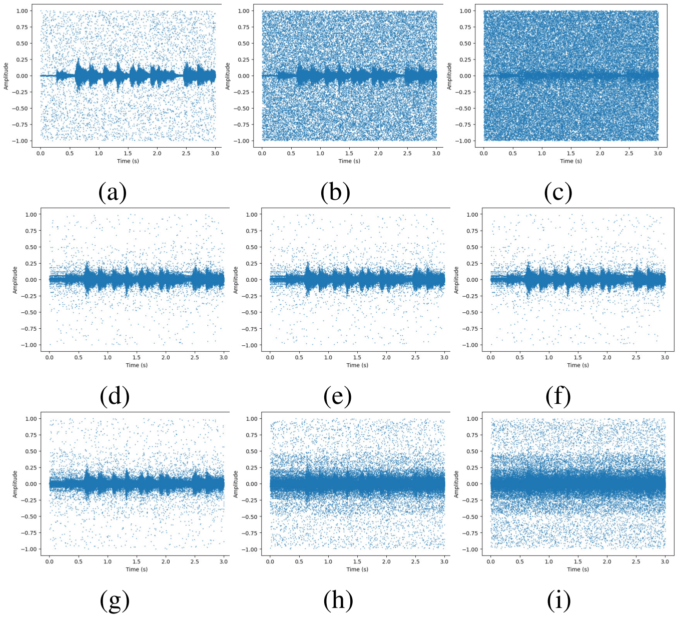
Audio	Plaintext Sensitivity		Key Sensitivity	
	NSCR	UACI	NSCR	UACI
BE	99.9990	33.4793	99.9927	33.4283
CB	99.9940	33.4365	99.9879	33.4180
FF	99.9982	32.8476	99.9865	33.5293
GB	99.9986	33.4425	99.9751	33.4767
IM	99.9982	33.4834	99.9956	33.3267
PP	99.9986	32.9418	99.9943	33.5358
PB	99.9981	33.4281	99.9711	33.6802
SW	99.9986	33.4495	99.9974	33.5460
Avg.	99.9979	33.3136	99.9876	33.4926
Ref [7]	-	-	98.6164	33.2352

TABLE V
Differential Attack Analysis

Audio	Differential Attack	
	NSCR	UACI
BE	99.9984	33.4673
CB	99.9985	33.4232
FF	99.9987	33.4400
GB	99.9968	33.4514
IM	99.9984	33.4694
PP	99.9981	33.4702
PB	99.9981	33.4291
SW	99.9986	33.4656
Avg.	99.9982	33.4520
Ref [1]	99.5835	25.5850
Ref [6]	99.9985	-
Ref [7]	100.0000	33.3578
Ref [8]	99.9893	30.3234
Ref [18]	99.9985	33.4000
Ref [26]	99.9494	33.1694
Ref [27]	99.9916	33.2057
Ref [28]	99.9948	35.5483

G. Differential Attack

The resilience of the proposed approach to diverse attacks is evaluated using differential test analysis. The selected plain text assault is a sort of differential attack that gauges the resistance of a system to its diffusion properties. Two encrypted audio signals are compared. Therefore, the original signal is modified a little, and then two encrypted signals are generated. The difference between the encrypted signals is measured in NSCR and UACI. For each test signal, we calculate the values of NSCR and UACI ten times (by complementing LSB of different samples, selected randomly), and their average is reported in Table V. NSCR and UACI values align with the ideal value, which suggests that the proposed method is sensitive to plaintext audio and can withstand differential attacks. The table also demonstrates that the proposed method is superior to the SoA methods except Ref [7].



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Fig. 8. Decryption results under different levels of salt and pepper (a)-(c), Gaussian (d)-(f), and speckle noise (g)-(i).

H. Robustness to Noise

Information transmitted via networks may be contaminated with noise. A reliable cryptography system should be able to decrypt the encrypted audio to some extent after noise assaults. This study adds salt and pepper, Gaussian, and speckle noise with 0.0001, 0.001, and 0.003 intensities to the cipher audio. The robustness of the suggested method is shown in Fig. 8, and it is clear that the decrypted audio signal is visible, and the performance becomes poor while increasing the noise intensity. Therefore, our suggested approach resists these kinds of noise attacks.

I. Computational Time and Speed Analysis

The encryption time and the running speed for various audio files are reported in Table VI. The table demonstrates that elongation in audio length corresponds to a proportional increase in encryption time. Based on this finding, it is inferred that our suggested approach encrypts 1 KB of data in around 0.1032 seconds. The encryption time of the proposed method is comparable to Ref [5], [6], [28]. Therefore, this technique can be employed in real-time applications. Finding IMFs is a time-consuming process. From our experiment, we note that EMD consumes around 93.86% of the total execution time.

VI. Conclusions

An audio encryption technique using 2DCLM and EMD is proposed in this work. This implementation is plaintext-sensitive. The proposed methodology works well, and the reliability of the proposed approach is verified against various statistical tests, demonstrating that our method is reliable against these attacks. Moreover, when contrasted with existing literature, the recommended procedures for encrypting audio show similar or even superior levels of security. However, due

TABLE VI
COMPUTATIONAL TIME ANALYSIS

	Size	Time	Speed	System
Audio	(KB)	(second)	(s/KB)	architecture
BE	2585	261.94	0.1013	Intel i5-10300H, CPU @ 2.50GHz, 8 GB RAM.
CB	130	16.06	0.1235	
FF	2585	267.34	0.1034	
GB	431	47.30	0.1097	
IM	2585	261.90	0.1013	
PP	2585	265.95	0.1028	
PB	414	44.43	0.1073	
SW	2585	269.50	0.1042	
Avg	1737.50	179.30	0.1031	
Ref [5]	141.66	32.59	0.2300	Intel i5, 4 GB RAM
Ref [6]	304	58.63	0.1928	-
Ref [7]	279.33	9.85	0.0352	-
Ref [18]	6477.50	239.65	0.0370	Intel i7, 16 GB RAM
Ref [28]	1395.80	372.4	0.2668	-

to the use of EMD, the time required for encryption in the proposed approach is longer than that required for specific SoA methods. In the future, we will suggest a faster EMD technique to enhance the audio time efficiency of the audio cryptosystem. Furthermore, there is potential to extend this research to encompass other forms of multimedia applications, such as images.

REFERENCES

[1] H. Aziz, S. M. M. Gilani, I. Hussain, A. K. Janjua, and S. Khurram, “A noise-tolerant audio encryption framework designed by the application of s8 symmetric group and chaotic systems,” *Mathematical Problems in Engineering*, vol. 2021, pp. 1–15, 2021. https://doi.org/10.1155/2021/5554707.

[2] E. A. Albahrani, T. K. Alshekly, and S. H. Lafta, “A review on audio encryption algorithms using chaos maps-based techniques,” *Journal of Cyber Security and Mobility*, pp. 53–82, 2022. https://doi.org/10.13052/jcsm2245-1439.1113.

[3] M. Karmani, N. Benhadjyoussef, B. Hamdi, and M. Machhout, “The sha3-512 cryptographic hash algorithm analysis and implementation on the leon3 processor,” https://doi.org/10.14445/22315381/IJETT-V69I6P210.

[4] C. Albin, D. Narayan, R. Varu, and V. Thanikaiselvan, “Dwt based audio encryption scheme,” in *2018 second international conference on electronics, communication and aerospace technology (ICECA)*, pp. 920–924, IEEE, 2018. https://doi.org/10.1109/ICECA.2018.8474602.

[5] H. Kakaei Kate, J. Razmara, and A. Isazadeh, “A novel fast and secure approach for voice encryption based on dna computing,” *3D Research*, vol. 9, pp. 1–11, 2018. https://doi.org/10.1007/s13319-018-0167-x.

[6] P. K. Naskar, S. Paul, D. Nandy, and A. Chaudhuri, “Dna encoding and channel shuffling for secured encryption of audio data,” *Multimedia Tools and Applications*, vol. 78, pp. 25019–25042, 2019. https://doi.org/10.1007/s11042-019-7696-z.

[7] X. Wang and Y. Su, “An audio encryption algorithm based on dna coding and chaotic system,” *IEEE Access*, vol. 8, pp. 9260–9270, 2019. https://doi.org/10.1109/ACCESS.2019.2963329.

[8] D. Shah, T. Shah, and S. S. Jamal, “Digital audio signals encryption by mobius transformation and hénon map,” *Multimedia systems*, vol. 26, pp. 235–245, 2020. https://doi.org/10.1007/s00530-019-00640-w.

[9] F. Farsana, V. Devi, and K. Gopakumar, “An audio encryption scheme based on fast walsh hadamard transform and mixed chaotic keystreams,” *Applied Computing and Informatics*, no. ahead-of-print, 2020. https://doi.org/10.1016/j.aci.2019.10.001.

[10] S. El-Zoghdy, H. S. El-sayed, and O. S. Faragallah, “Transmission of chaotic-based encrypted audio through ofdm,” *Wireless Personal Communications*, vol. 113, pp. 241–261, 2020. https://doi.org/10.1007/s11277-020-07187-4.

[11] G. Kaur, K. Singh, and H. S. Gill, “Chaos-based joint speech encryption scheme using sha-1,” *Multimedia tools and applications*, vol. 80, pp. 10927–10947, 2021. https://doi.org/10.1007/s11042-020-10223-x.

[12] W. Dai, X. Xu, X. Song, and G. Li, “Audio encryption algorithm based on chen memristor chaotic system,” *Symmetry*, vol. 14, no. 1, p. 17, 2021. https://doi.org/10.3390/sym14010017.

[13] S. Adhikari and S. Karforma, “A novel audio encryption method using henon–tent chaotic pseudo random number sequence,” *International Journal of Information Technology*, vol. 13, no. 4, pp. 1463–1471, 2021. https://doi.org/10.1007/s41870-021-00714-x.

[14] A. H. Khaleel and I. Q. Abduljaleel, “A novel technique for speech encryption based on k-means clustering and quantum chaotic map,” *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 1, pp. 160–170, 2021. https://doi.org/10.11591/eei.v10i1.2405.

[15] Nasreen and P. Muthukumar, “Secure audio signal encryption based on triple compound-combination synchronization of fractional-order dynamical systems,” *International Journal of Dynamics and Control*, vol. 10, no. 6, pp. 2053–2071, 2022. https://doi.org/10.1007/s40435-022-00942-4.

[16] O. M. Al-Hazaimeh, A. A. Abu-Ein, K. M. Nahar, and I. S. Al-Qasrawi, “Chaotic elliptic map for speech encryption,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 25, no. 2, pp. 1103–1114, 2022. https://doi.org/10.11591/ijeecs.v25.i2.pp1103-1114.

[17] R. Wu, S. Gao, X. Wang, S. Liu, Q. Li, U. Erkan, and X. Tang, “Aea-ncs: An audio encryption algorithm based on a nested chaotic system,” *Chaos, Solitons & Fractals*, vol. 165, p. 112770, 2022. https://doi.org/10.1016/j.chaos.2022.112770.

[18] A. Kumar and M. Dua, “Audio encryption using two chaotic map based dynamic diffusion and double dna encoding,” *Applied Acoustics*, vol. 203, p. 109196, 2023. https://doi.org/10.1016/j.apacoust.2022.109196.

[19] L. E. Bassham III, A. L. Rukhin, J. Soto, J. R. Nechvatal, M. E. Smid, E. B. Barker, S. D. Leigh, M. Levenson, M. Vangel, D. L. Banks, *et al.*, *Sp 800-22 rev. 1a. a statistical test suite for random and pseudorandom number generators for cryptographic applications*. National Institute of Standards & Technology, 2010. https://doi.org/10.6028/NIST.SP.800-22r1a.

[20] G. Li, X. Xu, and H. Zhong, “A image encryption algorithm based on coexisting multi-attractors in a spherical chaotic system,” *Multimedia Tools and Applications*, vol. 81, no. 22, pp. 32005–32031, 2022. https://doi.org/10.1007/s11042-022-12853-9.

[21] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, “The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, vol. 454, no. 1971, pp. 903–995, 1998. https://doi.org/10.1098/rspa.1998.0193.

[22] H. Ge, G. Chen, H. Yu, H. Chen, and F. An, “Theoretical analysis of empirical mode decomposition,” *Symmetry*, vol. 10, no. 11, p. 623, 2018. https://doi.org/10.3390/sym10110623.

[23] https://www2.cs.uic.edu/~i101/SoundFiles/.

[24] G. Alvarez and S. Li, “Cryptographic requirements for chaotic secure communications,” *arXiv preprint nlin/0311039*, 2003. https://doi.org/10.1142/S0218127406015970.

[25] E. M. Guizzo, *The essential message: Claude Shannon and the making of information theory*. PhD thesis, Massachusetts Institute of Technology, 2003. https://doi.org/1721.1/39429.

[26] P. Sathiyamurthi and S. Ramakrishnan, “Speech encryption using hybrid-hyper chaotic system and binary masking technique,” *Multimedia Tools and Applications*, vol. 81, no. 5, pp. 6331–6349, 2022. https://doi.org/10.1007/s11042-021-11757-4.

[27] D. Shah, T. Shah, M. M. Hazzazi, M. I. Haider, A. Aljaedi, and I. Hussain, “An efficient audio encryption scheme based on finite fields,” *IEEE Access*, vol. 9, pp. 144385–144394, 2021. https://doi.org/10.1109/ACCESS.2021.3119515.

[28] R. I. Abdelfatah, “Audio encryption scheme using self-adaptive bit scrambling and two multi chaotic-based dynamic dna computations,” *IEEE Access*, vol. 8, pp. 69894–69907, 2020. https://doi.org/10.1109/ACCESS.2020.2987197.

Alenrex Maity is currently a Research Scholar in the Department of IT at Jadavpur University, India. He received the B.Tech. degree in computer science and engineering from West Bengal University of Technology and an M.Tech degree in computer science and engineering from Kalinga Institute of Industrial Technology, India, in 2012 and 2015, respectively. He is researching image processing, computer vision, and information security.

Bibhas Chandra Dhara is a Professor in the Department of Information Technology at Jadavpur University, India. He received a B.Sc. degree (Hons.) in mathematics and the B.Tech. degree in computer science and engineering from the University of Calcutta, India, in 1997 and 2000, respectively, and the M.Tech. and Ph.D. degrees in computer science from the Indian Statistical Institute, in 2002 and 2008, respectively. He has published over 100 articles in peer-reviewed journals and proceedings. He is researching image processing, computer vision, pattern recognition, and information security.