Audio Encryption Scheme Using Self-Adaptive Bit Scrambling and Two Multi Chaotic-Based Dynamic DNA Computations

ROAYAT ISMAIL ABDELFATAH
Electronics and Electrical Communications Engineering Department, Tanta University, Tanta 31527, Egypt
e-mail: royat_esmaeel@f-eng.tanta.edu.eg

ABSTRACT In this paper, a novel scheme for secure audio transmission is introduced. The novelty of this scheme is that it combines four different techniques for audio encryption in the same scheme which makes it more secure: self-adaptive scrambling, multi chaotic maps, dynamic DNA encoding and cipher feedback encryption. Also, it introduces two new designed multi chaotic maps as pseudo random generators that combine five different chaotic maps with eight control parameters. The scheme consists of three phases with three secret keys. The first phase is a self-adaptive bit scrambling where SHA512 of the input audio is computed to be used as a first secret key for cyclic shifting the input audio binary stream which efficiently reduces the strong correlation between neighboring audio samples. The second phase is a dynamic DNA encoding of the scrambled audio by using a second secret key obtained from a pseudo random generator (LCS) using a novel design of multi chaotic maps including Sine, Chebyshev and Logistic map with three control parameters. As DNA encoding provides fast computations and large capacity for data transmission, the third phase consists of two DNA algebraic operations “AND” and “XOR” using a pseudorandom DNA sequence as a third secret key generated by another new designed multi chaotic HLG map which combines Henon, Logistic and Gaussian chaotic maps and has five control parameters. The third phase is structured in a cipher feedback mode and so it achieves strong diffusion and confusion in the encrypted audio. Also combining five different chaotic maps in the proposed scheme increases the number of secret control parameters which achieves a huge key space and so robust strength against brute force attacks. The scheme is evaluated by using different measurements including signal to noise ratio (SNR), peak signal to noise ratio (PSNR), number of samples change rate (NSCR), unified average changing intensity (UACI), root mean square (RMS), crest factor (CF), correlation coefficient, histogram, key sensitivity and key space. From results, it is clear that the scheme is highly secure and stronger than many recent similar audio encryption schemes against different types of attacks.

INDEX TERMS Audio encryption, DNA encoding, chaotic maps, security analysis.

I. INTRODUCTION Multimedia data security is a big challenge for the modern open communication systems which are threatened by different types of attacks. The main tool for achieving this security is the encryption. Audio files usually have larger size compared to the other multimedia files as images or texts. So the conventional encryption schemes as DES [1], AES [2] and RSA [3] are not suitable for audio data encryption as it will require long computational time, high power consumption and so can’t provide real time communication. So introducing new audio encryption schemes with high security and high speed has become the main concern of the researchers working in multimedia security area [4]–[9]. In [10], a perception based method for partially encrypting telephonic speech has been introduced. In [11], [12], two selective audio encryption schemes have been proposed. In [13], the authors proposed a progressive audio scrambling with compression algorithm. Most of the modern encryption schemes are based on two main principles; confusion and diffusion [14]. Confusion means that the relation between the ciphertext and the key is as complex as possible, while diffusion means that the
plaintext statistics are completely dissipated in the ciphertext statistics. Chaos theory has been discovered as a source of confusion and diffusion in cryptography [15]. Chaotic map has a dynamic behavior and is very sensitive to control parameter and initial conditions. This means that a tiny change in the initial conditions produces a remarkable change in the output. This matter makes the system more random. Chaotic maps have been used for many image encryption schemes as in [16]–[23]. In [24], a Baker chaotic map based speech segment permutation has been used for audio encryption.

The authors in [25] used 2-D cellular automata for audio scrambling and encryption.

The same matter for audio encryption. In [26], the authors introduced a chaotic encryption scheme for speech signals in transform domains. A voice encryption scheme based on off-line ica has been proposed in [27]. An audio shuffling encryption scheme has been introduced in [28]. In [29], compound chaotic mapping has been used for voice encryption scheme. An audio encryption scheme with single and double dimension discrete time chaotic systems has been proposed in [30]. Compressive sensing and Arnold transform has been used for audio encryption in [31].

The authors in [32] have proposed random unified chaotic maps as PRBG and used it for voice encryption in wireless communication. In [33] an audio encryption scheme based on the cosine number transform has been introduced. An audio encryption scheme using confusion and diffusion based on multi-scroll chaotic system and one time key has been presented in [34]. In [35], the authors proposed an approach for speech encryption using Zaslavsky map for pseudo random numbers generation.

A mixture of chaos functions has been used for audio encryption in [36]. A novel approach based on stream cipher for selective speech encryption has been introduced in [37]. A chaotic shift keying using multiple permutations has been introduced in [38]. In [39], an audio encryption scheme using delayed uncertainty of hybrid bidirectional associative memory and fuzzy neural networks has been proposed.

In [40] an audio encryption algorithm using permutation-substitution architecture with chaotic circle map and modified rotation equations has been proposed. An audio encryption scheme based on fast Walsh Hadamard Transform and mixed chaotic keystreams has been introduced in [41].

DNA cryptography is a recent security tool which has many advantages [42]. Among these advantages is large parallelism which provides faster computations. Also, it has large capacity for data storage and transmission by DNA molecules and it has a good power saving. This matter makes DNA cryptography better compared to conventional schemes as DES and AES. DNA computations have been firstly used for encryption in [43], [44] in 1994. Combining both DNA cryptography and chaotic maps will produce a very secure cryptosystem. This matter has been done in many image encryption schemes as in [45]–[47]. For audio encryption, a cryptosystem using chaotic maps and DNA encoding has been proposed in [48]. A novel fast and secure approach for voice encryption based on DNA computing has been introduced in [49].

Recently in [50], an encryption scheme for audio encryption using DNA encoding with Logistic map and channel shuffling has been introduced. In this scheme, the key generation is completely independent of the input audio blocks and the encryption of each block is independent of the previously encrypted blocks which may cause week confusion and diffusion and hence make the scheme not robust enough against chosen ciphertext and plaintext attacks.

This paper proposes an improved audio encryption scheme consists of three phases with three secret keys and has the following advantages:

(a) Scrambling is a good tool to hide the plaintext statistics and especially, bit scrambling achieves confusion and diffusion that audio samples scrambling cannot achieve and destroys the correlation between the adjacent samples. So the first phase of the proposed scheme is a scrambling of the input audio binary bit stream by using the hash function SHA512 of the input audio file as a first secret key. With this input audio key-dependency, the scheme is more random and stronger against statistical and differential attacks.

(b) In the second phase, a pseudo random generator (LCS) combines multi chaotic maps: Sine, Chebyshev and Logistic map is used for generating a pseudo random sequence as a second secret key for dynamic DNA encoding of the scrambled audio.

Using dynamic DNA encoding to change the audio samples and spread the original input audio effect into the cipher audio instead of using fixed DNA encoding achieves security enhancement as eight encoding rules are used instead of only one rule. DNA encoding provides faster computations and larger capacity for information storage and transmission.

(c) A second pseudo random generator which combines multi chaotic maps: Henon, Logistic and Gaussian chaotic maps as a new combination is used for generating a pseudo random binary sequence that is DNA encoded to be used as a third secret key with two DNA algebraic operations “AND” and “XOR” in the third phase of the proposed scheme.

The structure of this phase is in a cipher feedback mode that the current cipher audio file depends not only on the original input audio file and the key but also on the previous cipher audio. This matter spreads quickly any tiny change in the original input audio into the whole cipher audio and so achieves strong diffusion and confusion in the encrypted audio. The combination of a five different chaotic maps in the proposed scheme makes it more random and has a larger key space than a simple single chaotic system.

The remainder of the paper is organized as follows: Section II introduces the preliminary studies. Section III introduces the proposed scheme. In Section IV, the simulation results are presented. Section V presents the security analysis; Section VI gives comparative study of the proposed scheme. Conclusions are introduced in Section VII.
II. PRELIMINARIES

A. CHAOTIC MAP

Chaotic maps behavior changes rapidly with any slight change in the control parameters. The main idea in using chaotic maps for encryption is using its control parameters and initial conditions as secret keys. The chaotic maps are subdivided into one-dimensional and high dimensional chaotic maps. The one-dimensional maps have little parameters as Sine Map (SM), Logistic Map (LM), Chebyshev Map (CM), and Gaussian Map (GM) are defined as following as:

- **Sine map**:  
  \[ SM(n + 1) = r \times [\sin(pi \times SM(n))] \]  

- **Logistic map**:  
  \[ LM(n + 1) = r \times LM(n)[1 - LM(n)] \]  

- **Chebyshev Map (CM)**:  
  \[ CM(n + 1) = \cos(r \times \cos^{-1}(LM(n))] \]  

- **Gaussian Map (GM)**:  
  \[ GM(n + 1) = \exp(-r \times GM(n)^2) \]  

- **Henon Map (HM)**:  
  \[ HM(n + 1) = 1 - uHM(n)^2 + \beta HM(n - 1) \]  

The proposed LCS & HLG will be discussed:

1) FIRST PROPOSED MULTI CHAOTIC MAP (LCS)

The first proposed multi chaotic map is LCS which merges Sine, Chebyshev and Logistic map and has three control parameters as shown in Fig. The Mathematical function of the proposed LCS is described as following:

\[
LCS(n + 1) = (r \times \sin(1.5 \times \pi \times LCS(n))) \\
\times (1 - \cos(u \times \cos^{-1}(LCS(n)))) - 0.287
\]

where \( r, u \) are the chaotic system parameter and LCS(0) is initial value.

2) SECOND PROPOSED MULTI CHAOTIC MAP (HLG)

A second new designed multi chaotic map HLG which combines Henon, Logistic and Gaussian chaotic maps and shown in Fig. 2 is used for generating a pseudo random binary sequence used in the third phase as a key. The Mathematical function of the proposed HLG is described as following:

\[
HLG(n + 1) = 1 - \left[a \times \left(\exp(-4.9 \times HLG(n)^2)\right)^2\right] + b \times (\beta \times HLG(n - 1) \times (1 - HLG(n - 1)))
\]

where \( a, b \) and \( \beta \) are the chaotic system parameter and HLG(0) and HLG(1) are the initial values.

3) STATISTICAL TESTS AND CHAOTIC BEHAVIOR OF THE PROPOSED MULTI CHAOTIC MAPS

The random behavior of the two proposed multi chaotic maps LCS and HLG is tested by the NIST test suit which includes 16 statistical tests. These tests measure the randomness of the generated sequence. The test depends on the probability value (p-value). A comparison between the p-value and a significance level \( \alpha = 0.01 \); which is the threshold between rejection and non-rejection region; is done. If \( p \)-value < 0.01, the sequence is not random and rejected while if \( p \)-value > 0.01, the sequence is random and accepted. 10⁶ bit binary sequences generated from the two proposed LCS and HLG are tested by SP800-22 with 10³ iterations. Table 1 lists the results. All the obtained p-values are > 0.01 and the generated sequence pass all SP800-22 tests so the generated sequences from the two proposed multi chaotic LCS and HLG have a good randomness.

The Lyapunov Exponent (LE) is another tool for evaluating the performance of chaotic map which measures how the chaotic map is sensitive to the initial values. It can be thought of as the average logarithmic rate of separation or convergence of two nearby points of two time series \( A_t \) and
Bt separated by an initial distance \( \Delta D_0 = \|A_0 - B_0\|_2 \)

\[
\delta = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \ln \left| \frac{\Delta D_i}{\Delta D_0} \right|
\]

(8)

Stable system has negative LE, while positive values refer to exponential divergence from the initial value. As the maximum LE increases, the chaotic map random behavior increases. The LE of the proposed LCS, Logistic, Chebyshev and Sine chaotic maps are shown in Fig. 3. While the LE of the proposed HLG, Logistic and Gaussian chaotic maps are shown in Fig. 4. It is obvious that the LE of LCS and HLG is positive for all values of \( u \in (1, 4) \), while the other chaos have negative or less positive values and so not chaotic for some intervals of \( u \). So the generated sequences by LCS and HLG are more chaotic.

### B. DNA ENCODING

DNA is Deoxyribonucleic acid. It is a nucleic acid which carries the genetic information of living organisms. It has four different nucleic acid bases adenine (A), thymine (T), cytosine (C), guanine (G), where C and G, T and A are pairs complementary. If four bases A, T, C, G are encoding to 00, 11, 10, 01 we will get 24 types of encoding schemes. Due to the complementary relationship between DNA bases, there are eight types of coding groups that meet the complementary rule which are shown in Table 2. The permutation and

| Test                                | P-value for LCS | P-value for HLG | Result  |
|-------------------------------------|-----------------|-----------------|---------|
| Monobit frequency                   | 0.1158          | 0.6469          | Passed  |
| Block frequency                     | 0.0440          | 0.6326          | Passed  |
| Runs                                | 0.5421          | 0.7011          | Passed  |
| Longest-run-of-ones in a block      | 0.1767          | 0.4278          | Passed  |
| Binary matrix rank                  | 0.6891          | 0.2372          | Passed  |
| Discrete Fourier transform (spectral)| 0.3127          | 0.8543          | Passed  |
| Non-overlapping template matching   | 0.5389          | 0.6987          | Passed  |
| Overlapping template matching       | 0.4648          | 0.4075          | Passed  |
| Maurer’s universal statistical      | 0.4771          | 0.6842          | Passed  |
| Linear complexity                   | 0.7995          | 0.3149          | Passed  |
| Serial test                         | \{ 0.1104, 0.0049 \} | \{0.7981, 0.8090\} | Passed  |
| Approximate entropy                 | 0.4166          | 0.5978          | Passed  |
| Cumulative sums                     | 0.2143          | 0.6011          | Passed  |
| Random excursion                    | \{ 0.5742, 0.7832, 0.6642, 0.1402 \} | \{0.4115, 0.2620, 0.6029, 0.4604, 0.1460 \} | Passed  |
| Random excursion variant            | \{ 0.6517, 0.7453, 0.8901, 0.8434, 0.7978, 0.8662, 0.7965, 0.3801, 0.4165, 0.6942, 0.5551, 0.4186, 0.6992, 0.7932, 0.9558, 0.5806, 0.4167, 0.5250 \} | \{0.8505, 0.6583, 0.3402, 0.5894, 0.2195, 0.5030, 0.5519, 0.5474, 0.8095, 0.9024, 0.7463, 1, 0.7083, 0.5167, 0.11, 0.0594, 0.0389, 0.0354, 0.0747, 0.09832, 0.2898 \} | Passed  |
| Cumulative sums test reverse         | 0.1835          | 0.6829          | Passed  |
| Lempel-Ziv compression              | 0.2658          | 0.6724          | Passed  |

**FIGURE 3.** Lyapunov exponent (LE) comparisons of LCS.

**FIGURE 4.** Lyapunov exponent (LE) comparisons of HLG.
III. PROPOSED SCHEME
The proposed scheme audio encryption consists of three phases that will be discussed in the following section. The block diagram of the proposed scheme is shown in Fig. 5.

combinations of bases can be used for storing and calculation of information and hence its encryption. If we consider the encoding rule 1 then the XOR, addition and subtraction for DNA sequences are shown in Tables 3, Table 4 and Table 5 respectively.
A. THE FIRST PHASE (SELF-ADAPTIVE BIT SCRAMBLING)
The first phase is a self-adaptive scrambling as it uses a key which is a function of the input audio itself. Scrambling means changing the bit position by cyclic right shifting the input audio binary bit stream as following:

1. The input audio is sampled and the hash function SHA-512 of the sampled audio is computed where

\[ H = \text{SHA}_{512} \text{ (input audio)} = h_1h_2h_3 \ldots h_{512} \]  

2. The sampled audio is converted into binary vector, \( I = [i_1, i_2, \ldots, i_N] \), where \( N \) is the number of bits of the input audio.

   \( I \) is reshaped into a two dimensions matrix \( M \) (512, \( P \)), where \( N=512 \times P \). If \( N \) is not multiple of 512 then it is zero padded.

   The cyclic right shift for the rows of \( M \) using the bits \( h_i \) of the hash \( H \) computed in Eqn. (8) is done as following:

   \[ h_i = \begin{cases} 1, & \text{Cyclic right shift } R_i \text{ by } (i + 100) \\ 0, & \text{No change in } R_i \end{cases} \]

B. THE SECOND PHASE (_DYNAMIC DNA ENCODING)  
The second phase of the proposed scheme is a dynamic DNA encoding using a pseudo random sequence generated by the first proposed multi chaotic pseudorandom generator LCS, which merges Logistic, Chebyshev and Sine map as following:

1. Choose appropriate values for \( r \), \( u \) and LCS(0) for the proposed multi chaotic generator LCS to generate a chaotic sequence \( y=[y_1, y_2, \ldots, y_{N/2}] \) of length \( N/2 \) as in Eqn. (6) then convert this floating point sequence \( y \) into an integer sequence in the range \([1, 8]\), \( w=[w_1, w_2, \ldots, w_{N/2}] \) using Eqn. (9):

   \[ w_i = \left\lfloor \frac{y_i \times 10^{14}}{\text{mod} 8} \right\rfloor + 1 \]  

2. Reshape the scrambled input audio matrix \( M \) (512, \( P \)) into a one dimension vector \( V = [V_1, V_2, \ldots, V_N] \) again then partition it into a sequence of two bits

\[ B = \left\{ B_1, B_2, \ldots, B_2 \right\} \]  

\[ B_i \in \{00, 01, 10, 11\}. \]

3. The sequence \( B \) is converted into a DNA sequence \( S \) of length \( N/2 \) where \( S = [S_1, S_2, \ldots, S_{N/2}] \) of four bases A, G, T and C using one of the eight DNA encoding rules listed in Table 2 depending on the sequence \( w \), where: The DNA encoding rule selected for encoding \( B_i \) into

\[ S_i = w_i \text{ and } 1 \leq i \leq \frac{N}{2} \]  

C. THE THIRD PHASE (DNA COMPUTATIONS)  
The third phase of the proposed scheme is DNA computations using “XOR” and “AND” operations with a key which is a pseudo random sequence generated by the second proposed multi chaotic pseudorandom generator \( HLG \), which merges Logistic, Gaussian and Henon map as following:

1. Select appropriate values \( c, \beta, HLG(0) \) and \( HLG(1) \) for the proposed multi chaotic generator \( HLG \) to generate a chaotic sequence \( Z=[Z_1, Z_2, \ldots, Z_{N/8}] \) of length \( N/8 \) using Eqn. (7) then convert this floating point sequence \( Z \) into an integer sequence in the range \([0, 255]\).

\[ L = \{L_1, L_2, \ldots, \ldots, L_{N/8}\} \]  

\[ L_i = \left\lfloor \frac{z_i \times 10^{14}}{\text{mod} 256} \right\rfloor \]  

2. Convert the sequence \( L \) into a binary bit stream of length \( N \) as each \( L_i \in [0, 255] \) is converted into 8 bits or one byte. again then partition it into a sequence of two bits:

\[ X = \{X_1, X_2, \ldots, X_{N/2}\} \]  

\[ X_i \in \{00, 01, 10, 11\}. \]

3. The sequence \( X \) is converted into a DNA sequence \( U=[U_1, U_2, \ldots, U_{N/2}] \) of four bases A, G, T and C using one of the eight DNA encoding rules listed in Table 2 depending on the sequence \( w \), where: The DNA encoding rule selected for encoding \( X_i \)

\[ = w_i \text{ and } 1 \leq i \leq \frac{N}{2} \]  

4. The sequence \( S=[S_1, S_2, \ldots, S_{N/2}] \) is algebraically manipulated with the sequence \( U=[U_1, U_2, \ldots, U_{N/2}] \) in a cipher feedback structure as following:

\[ C(i) = U(i) \oplus \{S(i) + C(i-1)\}, 2 \leq i \leq \frac{N}{2} \]  

\[ C(1) = U(1) \oplus S(1) \]  

where \( \oplus \) and \( + \) are “XOR” and “AND” DNA operations respectively and \( C = \{C_1, C_2, \ldots, C_{N/2}\} \) with

\[ 1 \leq i \leq \frac{N}{2} \]  

5. \( C \) is dynamically decoded into binary sequence using the chaotic sequence \( w \) and the encoding rules listed in Table 2 as in Eqn. (11) to produce the encrypted audio.
At the receiver and upon receiving the encrypted audio the encrypted image as following:

1. Firstly, the receiver has to share three secret key with the sender to be able to decrypt the encrypted audio. The three secret keys are:
   - the hash of the input audio $H$.
   - the three control parameters of the multi chaotic generator $LCS : r, u$ and $LCS(0)$.
   - the five control parameters of the multi chaotic $HLG$ generator: $a, b, \beta$, $HLG(0)$ and $HLG(1)$.

2. Secondly, the receiver uses the five control parameters of the multi chaotic $HLG$ generator: $a, b, \beta$, $HLG(0)$ and $HLG(1)$ to generate the same chaotic sequence $Z$ as in Eqn. (7) then convert it to the sequence $L$ as in Eqn. (12), and uses the three control parameters of the multi chaotic generator $LCS : r, u$ and $LCS(0)$ to generate the chaotic sequence $w$ as in Eqn. (10).

3. Both the received encrypted audio and the chaotic sequence $L$ are DNA encoded, using one of the eight DNA encoding rules listed in Table 2 and the chaotic sequence $w$ as in Eqn. (11), to the DNA sequences $C$ and $U$ respectively which are algebraically manipulated in a cipher feedback structure as following:

\[ S(i) = [U(i) \oplus C(i)] - C(i - 1) \mod N \]

and

\[ S(1) = U(1) \oplus C(1) \]

4. Thirdly, the DNA sequence $S$ is DNA decoded into binary bit stream using the chaotic sequence $w$. Then this binary bit stream of length $N$ is reshaped into two dimensions matrix $M^\prime$ (512, $P$). Then the rows of $M^\prime$ are cyclic left shifted using the bits $h_i$ of the hash $H$ is done as following:

\[ h_i = \begin{cases} 1, \text{Cyclic left shift } R_i \text{ by (i + 100)} \\ 0, \text{No change in } R_i \end{cases} \]

Finally, the shifted binary bit stream in converted back to the original sampled audio.

### IV. SIMULATION RESULTS

The laptop used is Intel(R) Core(TM) i5-6200U CPU@ 2.30GHz, 4GB RAM, Windows 10 (64-bit), Mathemtica version 11. The proposed scheme is applied to different audios with different sizes and several experiments were done to measure the performance of the proposed algorithm. The results are shown in Fig.6. It is clear that the scheme converts the original audio into a nearly noise-like encrypted audio.

### V. SECURITY ANALYSIS

A robust multimedia encryption scheme has to be strong enough against all types of attacks [51]–[54].

Security evaluation of the proposed scheme will be considered using some common measurements include key space and key sensitivity analysis, histogram, spectrogram, correlation coefficient, NPCR, UACI, MSE, SNR, PSNR, RMS and CF value.

#### A. KEY SPACE AND SENSITIVITY ANALYSIS

Key space is the set of all the keys used for audio encryption. It can be determined using two tests: the number of keys and the key sensitivity.

1) THE NUMBER OF KEYS

A good audio encryption scheme has to have a big key space to be strong enough against brute force attack.

The proposed scheme has three secret keys as previously mentioned. The three secret keys are:

- The hash SHA-512 of the input audio $H$.
- The three control parameters of the multi chaotic generator $LCS : r, u$ and $LCS(0)$.
- The five control parameters of the multi chaotic $HLG$ generator: $a, b, \beta$, $HLG(0)$ and $HLG(1)$.

If the computation precision is around $2^{52}$ as in [55] then the key space is $2^{512} \times (2^{52} \times 2^{52} \times 2^{52} \times 2^{52} \times 2^{52} \times 2^{52} \times 2^{52}) = 2^{928}$. Table 6 lists a comparison with other schemes. It is obvious that, the proposed scheme has a much larger key space and hence more strength against brute force attacks.

2) KEY SENSITIVITY

A robust audio encryption scheme should be very sensitive to any tiny change in its keys. This means that a tiny change in the key value has to provide a totally different encrypted audio. In the proposed scheme, the correct keys are used for encrypting “Audio-1” file in Fig. 7 (a) to obtain the encrypted audio shown in Fig. 7 (b). Then slight changes in the keys values are made as shown in Table 7. The new encrypted audios are displayed in Fig. 7(c)-(j). The correlation and the number of signal change rate between the encrypted audios [Fig.7 (c)-(j)] and the encrypted audio Fig.7 (b) are listed in Table 7. These values show that there is no statistical similarity between the encrypted audio with the original keys and the encrypted audios with a slight change in the keys values. This matter proves that the proposed scheme is very sensitive to the secret keys.
FIGURE 6. Simulation results: [(a), (g), (m), (s), (y), (ee)] Original audio files. [(b), (h), (n), (t), (z), (ff)] Corresponding encrypted audio. [(c), (i), (o), (u), (aa), (gg)] Histogram of the original audio files [(a), (g), (m), (s), (y), (ee)] respectively. [(d), (j), (p), (v), (bb), (hh)] Histogram of encrypted audio [(b), (h), (n), (t), (z), (ff)] respectively. [(e), (k), (q), (w), (cc), (ii)] Spectrogram of original audio files [(a), (g), (m), (s), (y), (ee)] respectively. Spectrogram of encrypted audio files [(b), (h), (n), (t), (z), (ff)] respectively.
FIGURE 6. (Continued.) Simulation results: [(a), (g), (m), (s), (y), (ee)] Original audio files. [(b), (h), (n), (t), (z), (ff)] Corresponding encrypted audio. [(c), (i), (o), (u), (aa), (gg)] Histogram of the original audio files [(a), (g), (m), (s), (y), (ee)] respectively. [(d), (j), (p), (v), (bb), (hh)] Histogram of encrypted audio [(b), (h), (n), (t), (z), (ff)] respectively. [(e), (k), (q), (w), (cc), (ii)] Spectrogram of original audio files [(a), (g), (m), (s), (y), (ee)] respectively. Spectrogram of encrypted audio files [(b), (h), (n), (t), (z), (ff)] respectively.
and so it is robust against the brute force and statistical attacks.

**B. HISTOGRAM ANALYSIS**

Histogram analysis is an accurate metric for measuring the quality of encrypted audio signal. A good encryption scheme has to encrypt the original audio file into a random like noise with sample values distribution nearly flat. So it can face statistical attacks. Histogram of the original tested audio files is shown in Fig. 6 [(c), (i), (o), (u), (aa), (gg)] while histogram of the corresponding encrypted audio is shown in Fig. 6 [(d), (j), (p), (v), (z), (hh)].

**C. SPECTROGRAM ANALYSIS**

A spectrogram is a visual representation of an audio file frequency spectrum varying with time. It is obtained with Fourier transform. It is represented in two dimensions diagram: time and frequency. The audio samples are subdivided into chunks, and then Fourier transform is used to compute the frequency spectrum magnitude for each chunk. The spectrogram of the original tested audio files is shown in Fig. 6 [(e), (k), (q), (w), (cc), (ii)], while the spectrogram of the corresponding encrypted audio files is shown on Fig. 6 [(f), (l), (r), (x), (dd), (jj)].

This proves that for the proposed scheme, there is no similarity between the original tested audio files and their encrypted versions.

**D. CORRELATION ANALYSIS**

Correlation analysis is a statistical metric to measure the strength of an encryption scheme against various type of statistical attack [56]. Generally, it measure the mutual relationship between similar segments in the original audio and the encrypted audio files. A robust encryption scheme has to converts the audio file into random-like noisy signal with low correlation coefficient. Correlation coefficient can be

| Proposed method | Correct key | Modified key | Figure of modified key | Correlation coefficient | Number of signal change rate |
|-----------------|-------------|--------------|------------------------|------------------------|----------------------------|
| LCS initial value LCS(0). | LCS(0) | LCS(0) + 0.1 | Fig. 7 (c) | 0.0069 | 100% |
| The control parameter u | u | u + 0.1 | Fig. 7 (d) | 0.00097 | 99.999% |
| The control parameter r | r | r + 0.01 | Fig. 7 (e) | 0.00048 |
| HLG initial value HLG(0). | HLG(0) | HLG(0) + 0.1 | Fig. 7 (f) | 0.0032 | 100% |
| HLG initial value HLG(1). | HLG(1) | HLG(1) + 0.01 | Fig. 7 (g) | 0.0041 | 100% |
| The control parameter a | a | a + 0.01 | Fig. 7 (h) | -0.0013 | 100% |
| The control parameter b | b | b + 1 | Fig. 7 (i) | -0.001 | 100% |
| The control parameter β | β | β + 0.01 | Fig. 7 (j) | -0.00021 | 100% |
where \( N_s \) is the total number of samples, \( x_i \) and \( y_i \) are the sample values of the original and encrypted audio files, \( E(x) \) and \( E(y) \) are means values of samples, \( \sigma_x \), \( \sigma_y \) are the standard deviation of the encrypted audio file and original audio files, \( \text{cov}(x, y) \) is the covariance between both files. In Table 8, the size, duration and the correlation coefficient between the original audio files and the corresponding encrypted files are shown. The scattering plot diagram of the original audio file and the corresponding encrypted audio file is shown in Fig.8. The results show that the correlation coefficient values are close to zero which means that no similarity between the two files and reflects the randomness of the encrypted audio file. This matter proves the high quality of the proposed encryption scheme.

E. UACI AND NSCR ANALYSIS
Differential attacks resistance is a main measure for an encryption scheme strength. This resistance is measured through the NSCR (number of samples change rate) and UACI (unified average changing intensity) tests [57]. To apply these tests, two audio segments different only by one sample are encrypted with the same key. Then the two encrypted audio segments are compared by NSCR and UACI which are calculated as following:

\[
\text{NSCR} = \sum_i \frac{D_i}{N_s} \times 100\% \quad i = 1, 2, \ldots, N \tag{19}
\]

\[
\text{UACI} = \frac{1}{N_s} \left[ \sum_{i=1}^{N} \frac{(x_i - x'_i)^2}{65535} \right] \tag{20}
\]

where \( D_i \) is the number of samples change rate. The theoretical value for NPCR is 99.61\% and for UACI is 33.46\%. An audio encryption scheme with values of NPCR and UACI higher above the theoretical values will be more robust and more secure. For the proposed scheme, a sample is randomly selected then is slightly modified to produce a modified audio. Then the encrypted audio of the original audio \( x \) and the encrypted audio of the modified audio \( x' \) are used to compute NPCR and UACI for different audio files and are shown in Table 9. It is noted that the obtained values of the NPCR and UACI are higher than the theoretical values irrespective of the chosen sample position. This proves

FIGURE 8. Correlation between the samples in (a) original audio file, (b) encrypted audio file.

TABLE 8. SNR, PSNR and correlation coefficient between the original and encrypted audio files.

| Audio file | Size   | Duration (Sec.) | Correlation coefficient | SNR    | PSNR |
|------------|--------|-----------------|-------------------------|--------|------|
| Audio-1    | 156 KB | 1.8             | 0.00016                 | -28.53 | 4.25 |
| Audio-2    | 304 KB | 1.76            | -0.00028                | -28.14 | 4.31 |
| Audio-3    | 1.84 MB| 11              | 0.00052                 | -37.08 | 4.35 |
| Audio-4    | 1.02 MB| 33.52           | 0.00026                 | -32.96 | 4.12 |
| Audio-5    | 3.13 MB| 18.6            | 0.00041                 | 38.66  | 4.30 |
| Audio-6    | 543 KB | 3.15            | 0.00040                 | -29.96 | 4.34 |
| Audio-7    | 3.19 MB| 19              | 0.00053                 | -38.02 | 4.37 |
| Audio-8    | 984 KB | 5.7             | 0.00057                 | -32.95 | 4.33 |
that the proposed scheme has better resistance against the differential attack.

**F. SIGNAL TO NOISE RATIO (SNR)**

Signal to Noise Ratio (SNR) is used for measuring the signal quality [58], [59]. As the values be greater than 0 dB, as the signal be more than the noise. Given the original and the encrypted audio files, the SNR can be calculated as following:

$$ SNR = 10 \times \log_{10} \left( \frac{\sum_{i=1}^{N_s} x_i^2}{\sum_{i=1}^{N_s} (x_i - y_i)^2} \right) $$ (22)

where $x_i$ and $y_i$ are the samples of the original and encrypted samples and $N_s$ is the number of samples. For the proposed scheme, the values of SNR of the tested audio files are listed in Table 8. The more negative the SNR is, the more powerful is the scheme. From the table, the proposed scheme has better negative SNR and so it is stronger against attacks.

**G. PEAK SIGNAL TO NOISE RATIO (PSNR)**

The mean squared error for two streams stored in vectors X and Y is computed as following:

$$ MSE = \frac{1}{n} \sum_{i=1}^{n} (X[i] - Y[i])^2 $$ (23)

If X represents the original audio file and Y represents its encrypted audio, the PSNR can be computed as following:

$$ PSNR = 10 \times \log_{10} \left( \frac{\text{MAX}^2}{MSE} \right) $$ (24)

where MAX is the maximum value of the stream. The PSNR values for the encrypted tested audio files are computed and listed in Table 8. It is noted that the values are small. Lower values of PSNR is desired for encrypted audio files as it refers to high level of noise in the encrypted audio files and so strong resistance against attacks.

**H. ROOT MEAN SQUARE (RMS) & CREST FACTOR (CF) VALUE**

Root mean square (RMS) value is determined by the average amplitude level of an audio signal. It is equivalent to standard deviation when the input signal has zero mean and is computed as following:

$$ RMS = \sqrt{\frac{1}{N} \sum_{i=1}^{N} |A_i|^2} $$ (25)

Crest factor (CF) is a waveform parameter which is determined as the ratio of the peak values to the effective value. It is a measure of the extremeness of the peaks in a waveform and this is the minimum possible value. CF of ratio 0 dB indicates no peaks as DC current. Higher CF means peaks. It is given as following:

$$ CF = 20 \log_{10} \frac{|V_{\text{Peak}}|}{V_{\text{RMS}}} $$ (26)

For the proposed scheme, RMS and CF values for the tested values are computed and listed in Table 10. As shown the values of RMS and CF of the encrypted audio files are close to 0.6 and 4.3 respectively. This proves that there no statistical relationship between the original audio files and the corresponding encrypted audio files.

**I. COMPUTATIONAL TIME ANALYSIS**

Table 11 lists the computational speed of the proposed scheme for the eight tested audio files. It takes approximately between 0.19 - 0.37 s to encrypt 1KB of data. This time increases as the audio file length increases. Comparison with the Ref. [49] proves that the proposed scheme is faster and so it can be used for real time communication. Compression of audio file before encryption will more reduce this running time.

**VI. COMPARATIVE STUDY**

The capabilities of the proposed scheme compared to other schemes will be evaluated in this section. The most common security parameters including correlation, key space, SNR, PSNR, UACI, NSCR, RMS and CF will be used for performance comparison. Table 12 shows this comparison with other schemes. From this result, it is found that the proposed scheme has the lowest value of correlation coefficient between the original audio file and its corresponding encrypted one which reflects the lowest similarity.
The proposed scheme is highly secure and stronger than many recent similar audio encryption schemes against different types of attacks.

**REFERENCES**

[1] Data Encryption Standard, Federal Inf. Process. Standards, Nat. Bureau Standards, Gaithersburg, MD, USA, 1999, vol. 46, no. 3, pp. 1–22.

[2] J. Daemen and V. Rijmen, “Rijndael: The advanced encryption standard,” *Dr. Dobb’s J.*, vol. 26, no. 3, pp. 137–139, Mar. 2001.

[3] R. Rivest, A. Shamir, and L. Adleman, “A method for obtaining digital signatures and public-key cryptosystems,” *Comm. ACM*, vol. 21, no. 2, pp. 120–126, 1978.

[4] B. Bhargava, C. Shi, and S.-Y. Wang, “MPEG video encryption algorithms,” *Multimedia Tools Appl.*, vol. 24, no. 1, pp. 57–79, Sep. 2004.

[5] A. A. A. El-Latif, L. Li, and X. Niu, “A new image encryption scheme based on cyclic elliptic curve and chaotic system,” *Multimedia Tools Appl.*, vol. 70, no. 3, pp. 1559–1584, Jun. 2014.

[6] H. R. Shakir, “An image encryption method based on selective AES coding of wavelet transform and chaotic pixel shuffling,” *Multimedia Tools Appl.*, vol. 78, no. 18, pp. 26073–26087, Sep. 2019.

[7] A. U. Rehman and X. Liao, “A novel robust dual diffusion/confusion encryption technique for color image based on chaos, DNA and SHA-2,” *Multimedia Tools Appl.*, vol. 78, no. 2, pp. 2105–2133, Jan. 2019.

[8] R. I. Abdelfatah, “Secure image transmission using chaotic-enhanced elliptic curve cryptography,” *IEEE Access*, vol. 8, pp. 3875–3890, 2020.

[9] R. I. Abdelfatah, “A new fast double-chaotic based image encryption scheme,” *Multimedia Tools Appl.*, vol. 79, nos. 1–2, pp. 1241–1259, Jan. 2020.

[10] A. Servetti and J. C. De Martin, “Perception-based partial encryption of compressed speech,” *IEEE Trans. Audio Speech Process.*, vol. 10, no. 8, pp. 637–643, Nov. 2002.

[11] N. J. Thorwirth, P. Horvatic, R. Weis, and J. Zhao, “Security methods for MP3 music delivery,” in *Proc. Conf. Rec. 34th Asilomar Conf. Signals, Syst. Comput.*, 2002, pp. 1831–1835.

[12] A. Servetti, C. Testa, and J. C. De Martin, “Frequency-selective partial encryption of compressed audio,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., ICASSP*, 2003, pp. 668–671.

[13] W.-Q. Yan, W.-G. Fu, and M. S. Kankanhalli, “Progressive audio scrambling in compressed domain,” *IEEE Trans. Multimedia*, vol. 10, no. 6, pp. 960–968, Oct. 2008.

[14] C. E. Shannon, “Communication theory of secrecy systems,” *Bell Syst. Tech. J.*, vol. 28, no. 4, pp. 656–715, Oct. 1949.

[15] G. Alvarez and S. Li, “Some basic cryptographic requirements for chaos-based cryptosystems,” *Int. J. Bifurcation Chaos*, vol. 16, no. 8, pp. 2129–2151, Aug. 2006.

[16] H. Huang and S. Yang, “Colour image encryption based on logistic mapping and double random-phase encoding,” *IET Image Process.*, vol. 11, no. 4, pp. 211–216, Apr. 2017.
