mathematics

Article

An Efficient Audio Encryption Scheme Based on Elliptic Curve over Finite Fields

Hafeez Ur Rehman [1,*], Mohammad Mazyad Hazzazi [2], Tariq Shah [1], Zaid Bassfar [3] and Dawood Shah [1]

[1] Department of Mathematics, Quaid-i-Azam University, Islamabad 45320, Pakistan; stariqshah@gmail.com (T.S.); dawoodshah254@gmail.com (D.S.)
[2] Department of Mathematics, College of Science, King Khalid University, Abha 61421, Saudi Arabia; mmhazzazi@kku.edu.sa
[3] Department of Information Technology, University of Tabuk, Tabuk 71491, Saudi Arabia; zbassfar@ut.edu.sa
* Correspondence: hrehman@math.qau.edu.pk

Abstract: Elliptic curve (EC) based cryptographic systems are more trustworthy than the currently used cryptographic approaches since they require less computational work while providing good security. This paper shows how to use an EC to make a good cryptosystem for encrypting digital audio. As a preliminary step, the system uses an EC of a particular type over a binary extension field to distort the digital audio pixel position. It reduces the inter-correlation between pixels in the original audio, making the system resistant to statistical attacks. In creating confusion in the data, an EC over a binary extension field is used to make a different number of substitution boxes (S-boxes). The suggested design employs a unique curve that relies on efficient EC arithmetic operations in the diffusion module. As a result, it generates high-quality pseudo-random numbers (PRNs) and achieves optimal diffusion in encrypted audio files with less processing work. Audio files of various sizes and kinds can all be encrypted using the provided algorithm. Moreover, the results show that this method effectively protects many kinds of audio recordings and is more resistant to statistical and differential attacks.

Keywords: Mordell elliptic curve; Galois field; substitution box; pseudo-random numbers generator; audio encryption

MSC: 12F12

Citation: Ur Rehman, H.; Hazzazi, M.M.; Shah, T.; Bassfar, Z.; Shah, D. An Efficient Audio Encryption Scheme Based on Elliptic Curve over Finite Fields. *Mathematics* **2023**, *11*, 3824. https://doi.org/10.3390/math11183824

Academic Editor: Alla Levina

Received: 20 July 2023
Revised: 28 August 2023
Accepted: 30 August 2023
Published: 6 September 2023

1. Introduction

The relevance of multimedia data in social life has expanded in recent decades because of the rapid development of digital technology. As a result, a wide range of industries relies on multimedia data. Since data are digitally stored as bits in computer networks worldwide, recent technology has altered how information is transmitted. Also, as digital information demands newer cryptographic algorithms, the remaining ones must be updated and adapted to function. The field of multimedia data security is receiving a lot of attention because of these dangers. Numerous techniques have been developed to secure against unauthorized access to private data transmitted over public networks. Several prevalent algorithms, such as data encryption standard (DES), advanced encryption standard (AES), Rivest-Shamir-Adlemen (RSA), and Rivest cipher 4 (RC4), are frequently used for security objectives and are highly secure and dependable. These techniques rely on mathematical calculations and iterative procedures, making them very protective for confidential transmissions. Also, multimedia data are differentiated by their multidimensional nature due to the massive amount of data and other distinguishing properties, such as solid pixel correlation, large storage capacity, and high redundancy between pixels. Thus, security could be insufficient depending on algorithms such as AES, RSA, and DES for multi-media data. There are numerous encryption approaches for the safety of multimedia data in the

literature. A digital image encryption technique based on chaotic systems and Galois fields is presented in [1–4]. However, because of the enormous computational complexity of the high-dimensional, chaotic system, its application to the device is costly.

In contrast, the EC group structures are more sensitive to input parameters than chaos-based structures, providing better protection than chaos. In [5], the author developed a hybrid cryptosystem based on AES and the EC. The schemes based on an EC over finite fields using different techniques are presented in [6–8]. To increase the security of the encryption algorithm, an iterative structure made up of the random binary encoding technique is devised. Also, a concept for image encryption based on mixed chaos and hash operations is presented in [9–12]. The audio tracks have a large amount of storage space and are distinct from other multimedia content. As a result, digital audio should be protected using a specific method.

1.1. Related Work

There are various algorithms for encrypting and decrypting digital audio in the literature. But there is no single method for all types of audios. In 2002, encryption of telephonic oration based on the perception technique was proposed by Servetti and De Martin [13], in which they presented two different designs. The first approach has a high bit rate but a low-security scope according to the specifications. In contrast, the second approach aims to encrypt more and more data while preserving high-security levels. A simplistic compression and selection encryption-based perceptual audio coding method is provided [14]. The evaluation of encrypted MP3 encoded data was the focus of this research. A selective partial encryption approach for MP3 audio was later developed by Servetti et al. [15]. Unfortunately, to protect the perceptual data, the suggested process degrades the original audio sequence quality while maintaining the audio information's content. Following that, Bhargava et al. presented many MPEG-compatible digital video encryption techniques [16]. He used a secret key in this algorithm to randomize discrete cosine transform coefficients. Consequently, Grangetto et al. [17] proposed a novel multimedia data protection system assembled on an arithmetic coding foundation. Mathematical code randomization was used in the suggested work to achieve the objective of multimedia data encryption. Similarly, the scrambling of digital audio data in the compressed field was designed by Yan et al. [18]. Before transmission, the proposed approach successfully scrambled the personal audio data using a key. However, the brute force attack uncovered a susceptibility at work. Neto and Lima [19] proposed an audio encryption method utilizing cosine number transformation. Chunks of audio data are encrypted using the appropriate form. The ciphered data were confused and diffused due to the block selection's overlapping sample criteria. Refs. [20,21] provide digital network architecture (DNA) encoding, channel scrambling, and a new approach to audio encryption using predefined synchronization of a fractional order chaotic system. Next, the chaotic maps given in [22] are used to devise a lossless audio encryption technique. In this study, they used three-dimensional (3-D) chaotic systems for encrypting various types of audio files. In [23], the author divided the audio data into two blocks, such as eight bits and seven bits. Further, they used substitution and permutation techniques to encrypt the data. Although, they have taken data in which some values are in the range of 16-bits, which are kept fixed, and while decrypting, those values were found missing. Moreover, the utilization of $GF(2^7)$ or $GF(2^8)$ requires larger memory and storage.

1.2. Motivation

The bulk of audio encoding schemes lack sufficient cryptanalysis and security assessments to validate the stability of these cryptosystems against adversarial attacks. As a result, a robust algorithm is demanded to protect audio data from various threats. Due to less computational action and solid security, EC-based cryptographic architectures are more steadfast than existing cryptographic techniques. In addition, the features of finite fields compel us to create a novel digital audio data security algorithm based on an EC over

a Galois field. Moreover, using Galois fields in hardware cryptography would increase performance and reduce cost.

1.3. Our Contribution

The primary goal of developing this technique is to provide integrity and authenticity for encrypting multiple audio files. In reference [23], the author segmented the audio data into two sections, one comprising eight bits and the other consisting of seven bits. Subsequently, they employed substitution and permutation approaches for data encryption. However, it is worth noting that the data contained specific values within the 16-bit range, which remained fixed during encryption but were mysteriously missing during decryption. Moreover, $GF(2^7)$ or $GF(2^8)$ also demanded more extensive memory and storage resources. In this manuscript, we have divided data into four blocks and utilized an EC over the binary extension field of order 16 to address the issue of missing elements, which is in the range of 16-bits. Moreover, the use of smaller fields leads to potentially faster computations. The points mentioned above are the main findings of this scheme. Additionally, the use of EC points in the binary extension field provides greater strength to the proposed algorithm. Accordingly, we organize our work so that Section 2 begins with a brief introduction to the EC and Galois fields. Thorough explanations of the proposed encryption technique are exemplified in Section 3. Likewise, Section 4 examines the suggested encryption algorithm's security analysis and computer simulation results. In the Section 5, we put together some concluding reflections.

2. Preliminaries

In this part, several fundamental and crucial notions, like ECs, Galois fields, Euler's phi function, and primitive polynomials are presented.

2.1. Galois Field GF

A field with finite elements is called a Galois field or a finite field, represented as $GF(p^m)$, where p denotes a prime number and m is any positive integer. Further, if $m = 1$, $GF(p)$ consists of the set of integers $Z_p = \{0, 1, 2, \ldots, p-1\}$ with arithmetic operation modulo prime p. A polynomial $h(x)$ of degree m in $F[x]$ is called an irreducible polynomial if it cannot be factorized into polynomials having a degree less than m. Accordingly, in every Galois field $GF(p^m)$ there is a polynomial called a primitive irreducible polynomial (PIP), which generates all the elements of that field, except zero. The number of different PIPs over the Galois field $GF(p^m)$ is calculated by $\frac{\varphi(p^m-1)}{m}$, where φ represents Euler's phi function.

2.2. Elliptic Curve

The Weierstrass form of an EC $E_{p,a,b}$ over a prime field p is a cubic equation whose coefficients are the two non-negative integers $a, b \leq p-1$, and defined as:

$$y^2 = x^3 + ax + b \tag{1}$$

where the points $(x, y) \in F_p \times F_p$ lying on the EC together with the point at infinity O forms a group. By Hasse formula, the approximate number of points $|E_{p,a,b}|$ lying on the EC is obtained by:

$$abs\ \left(\left|E_{p,a,b}\right| - p - 1\right) \leq 2\sqrt{p} \tag{2}$$

where *abs* denotes the absolute value; also, when we put $a = 0$ in Equation (1), we obtain a particular type of EC called a Mordell EC (MEC). The specialty of this curve is that when we take a prime number in the form $p = 2\ mod\ 3$, it has precisely $p + 1$ distinct points lying on that MEC.

3. Proposed Encryption Scheme

Arrays of digitally encoded audio signals are operated to store, modify, replicate, and generate sound using audio technology. In addition, "digital audio" can refer to the sample of discrete sequences selected from the audio wave format. Under the proposed technique, digital audio in the wav format will be encrypted before being sent across an insecure channel. A 16-bit signed integer class was first utilized to read the audio file, with a range value of $\left[-2^{15},\ 2^{15}-1\right]$. M rows and N columns are in these matrices, which we refer to as the original audio data matrix (G). The following part will go over the encryption scheme's step-by-step procedure.

3.1. Proposed Generation of Pseudo-Random Numbers (PRNs)

In multimedia data security, the production of random numbers plays a vital role. Many PRN-generating systems have been devised by multiple researchers. Moreover, ECs are frequently employed to generate random numbers. Generally, the generation of PRNs through ECs utilizes the EC's group law and arithmetic operations. In this part, we use specific EC and modulo operations to generate RNs that have enough length. In the beginning, to generate PRNs, choose the smallest prime number p such that $p > M \times N$ and $p \equiv 2\ (mod\ 3)$. Then choose an MEC and generate the curve $E_{p,b}: y^2 = x^3 + b\ mod\ p$. The specialty of this curve is that it has $p+1$ number of points (x^*, y^*) lying on that EC. Subsequently, the following map is used to obtain the required RNs $(x^*, y^*) \rightarrow y^*$. Afterwards, to obtain a new matrix G_s, this sequence is applied to shuffle the matrix G. To produce PRNs, we used the approach outlined above by fixing p and the parameter of MEC $b = 7$. The obtained sequence is then subjected to the NIST test, which is depicted in the analysis section.

3.2. Construction of S-Boxes Scheme

The primary aim of an S-box is to obscure the relationship between encrypted data and the encryption keys used. Consequently, creating effective S-boxes using secure methods in modern cryptographic systems is imperative. The S-box is typically engaged in the cryptosystem's confusion module to confound the cipher data. As an outcome, the quality of the S-box determines how much chaos the cryptosystem can create. Because audio contains a lot of information, as a result, multiple S-boxes are incorporated into the proposed cryptosystem to increase the level of randomness in the encrypted data. To achieve this goal, the security of EC structures has captured our attention, as they can be utilized to design S-boxes using their fundamental operations. To this end, this section presents a straightforward and rapid algorithm. A novel technique is used in this work to construct a different number of S-boxes by employing EC of the form $E^m_{2,b}: y^2 = x^3 + b$ over the Galois field $GF(2^m)$. In the suggested method for creating numerous S-boxes, arithmetic operations are used of Galois fields instead of EC. We chose an EC of the form $E^m_{2,b}: y^2 = x^3 + b$ over the Galois field $GF(2^m)$ and generated points (x^*, y^*) lying on that EC, where $b \in GF(2^m)\{0\}$. After that, we defined a bijective function such that $F: GF(2^m) \rightarrow GF(2^m)$ and defined by:

$$F(x^*) = \ell.x^* + m \tag{3}$$

where l, m be any elements of the Galois field except zero. Here, the operation multiplication and addition are taken over $GF(2^m)$. In addition, we applied an inverse function to the resulting elements of Equation (3) under the corresponding $GF(2^m)$ and obtained the required S-boxes. Also, we constructed 4×4 S-boxes that have entries consisting of four bits. So, we take $m = 4$ in this work to generate a various number of 4×4 S-boxes. The S-boxes created through this technique are presented in Table 1. The performance analysis of the proposed S-boxes is depicted in Table 2, exhibiting that the proposed scheme is better than other existing schemes.

Table 1. Proposed S-boxes.

1	8	12	6	2	11	15	5	5	12	8	2	7	14	10	0
3	4	5	9	0	7	6	10	7	0	1	13	5	2	3	15
0	15	10	7	3	12	9	4	4	11	14	3	6	9	12	1
14	13	11	2	13	14	8	1	10	9	15	6	8	11	13	4

Table 2. Analysis comparison of proposed S-boxes with some existing S-boxes.

S-Box	NL	SAC	LP	DP
Proposed_1	4	0.25	0.75	0.25
Proposed_2	4	0.25	0.75	0.25
Proposed_3	4	0.25	0.75	0.25
Proposed_4	4	0.25	0.75	0.25
Ref. [24]	3.5	0.125	0.25	0.4375
Ref. [25]	3.5	0.125	0.5	0.54469

3.3. Proposed Audio Encryption Algorithm

We debated the proposed encryption scheme in this subsection. The proposed technique keeps digital audio data in wav format safe when transmitted across insecure networks. Here is a step-by-step framework of the encryption method that is offered as follows:

Step 1. Since the technique initially reads the audio, which contains both negative and positive integers, it constructs a matrix $N_{i,j}$ consisting of 0 and 1 to identify the location of integers, defined as follows:

$$N_{i,j} = \begin{cases} 0, & if \quad G_{i,j} < 0 \\ 1, & if \quad G_{i,j} \geq 0 \end{cases} \tag{4}$$

where $G_{i,j}$ denotes the data matrix and (i, j) presents the location of an element in that data matrix.

Step 2. Next, apply an absolute function on the data matrix G to produce a new matrix G'' with integer values in the range of the Galois field $\left(2^{16}\right)$.

Step 3. Since there is a substantial correlation between adjoining integers, efficient data methods include several features to split this strong correlation among adjoining numbers. Generate the sequence α of RNs by selecting a prime $p > M \times N$ where $p \equiv 2 \ (mod\ 3)$ by the random number generator presented in Section 3.1. After that, shorten the sequence and employ the new one to scramble the actual audio matrix.

$$G^p(i, j) = G''(\alpha(i), \alpha(j)) \tag{5}$$

where i, j denotes the position of an element in the shuffled matrix G^p. The spectrogram graph and the scrambled audio waveform are displayed in Figures 2 and 3. Also, the figures indicate that the permutation step was the most disruptive in terms of audio distortion.

Step 4. Since any cryptosystem would be incomplete without a stage characterized by disorientation and uncertainty, the substitution process is employed to confuse the ciphered data in this step. To reduce the time complexity, we divided the permuted block into four subblocks, each of which contained 4-bit integers, using the following maps:

$$\pi : \ GF(2^{16}) \rightarrow GF(2^4) \times GF(2^4) \times GF(2^4) \times GF(2^4)$$

Defined by

$$\pi\left(\sum_{j=0}^{15} b_j x^j\right) = \left(\sum_{j=0}^{3} b_j x^j, \ \sum_{j=4}^{7} b_{j-4} x^{j-4}, \ \sum_{j=8}^{11} b_{j-8} x^{j-8}, \ \sum_{j=12}^{15} b_{j-12} x^{j-12}\right) \tag{6}$$

where the coefficients $b_j \in \{0, 1\}$. The above data are now split into the four subblocks G_1^p, G_2^p and G_3^p, G_4^p.

Step 5. Generate different 4×4 S-boxes by using the technique mentioned in Section 3.2. In this step, we performed the substitution by applying different 4×4 S-boxes to the subblocks G_1^p, G_2^p, G_3^p, and G_4^p, respectively, thus acquiring new subblocks G_1^s, G_2^s, G_3^s, and G_4^s. The performance index of these 4×4 S-boxes is given in Section 4.

Step 6. Combine the resultant matrices G_1^s, G_2^s, G_2^s, and G_4^s and apply an inverse map of the map given in Equation (6) to convert the subblocks of 4-bits to a single block of 16-bits.

$$\pi^{-1} \; : \; GF(2^4) \times GF(2^4) \times GF(2^4) \times GF(2^4) \rightarrow GF(2^{16})$$

Defined by

$$\begin{aligned} &\pi^{-1}\left(\sum_{j=0}^{3} b_j x^j \, , \, \sum_{j=4}^{7} b_{j-4} x^{j-4} \, , \, \sum_{j=8}^{11} b_{j-8} x^{j-8}, \, \sum_{j=12}^{15} b_{j-12} x^{j-12}\right) \\ &= \left(\sum_{j=0}^{3} b_j x^j + \sum_{j=0}^{3} b_{j+4} x^{j+4} + \sum_{j=0}^{3} b_{j+8} x^{j+8} + \sum_{j=0}^{3} b_{j+12} x^{j+12}\right) \end{aligned} \tag{7}$$

After applying that map, we receive a new matrix G^s in which each element lies in the range of $GF(2^{16})$.

Step 7. Next, generate a sequence of random numbers α by choosing $p > M \times N$ to produce diffusion in the encrypted data through the proposed technique mentioned in Section 3.1. Consequently, take $M \times N$ number of elements from that sequence and reduce the size of the elements of that sequence in the range of $GF(2^{16})$. As we are working in a binary field, the generated sequence is applied to the elements of G^s as such:

$$G^{s2}(i, j) = G^s(i, j) + \rho(i, j) \tag{8}$$

Here, addition is taken over the Galois field $GF(2^{16})$, and (i, j) represents the position of integers in the matrix.

Step 8. Accordingly, convert the matrix G^{s2} entries into the integers $[-2^{15}, \, 2^{15} - 1]$ by using the matrix $N_{i,j}$. The mathematical formula is given below:

$$G_E(i, j) = \begin{cases} G^{s2}(i, j), & if \quad N_{i,j} = 0 \\ -G^{s2}(i, j), & if \quad N_{i,j} = 1 \end{cases} \tag{9}$$

The resulting matrix is then used to create an audio file, subsequently ciphered. Our proposed technique is applied to different audio file sizes mentioned in the security analysis. The step-by-step procedure of the proposed encryption scheme is given in Figure 1. The spectrogram graph and the scrambled audio waveform are displayed in Figures 2 and 3. Also, the figures indicate that the permutation step was the most disruptive in terms of audio distortion. Additionally, the histogram analysis of differently encrypted files is given in Figure 4. One can see from Figure 4 that the figure is uniform. As the encrypted audio files are uniform, the proposed method is efficient in preserving the actual extent of the original audio.

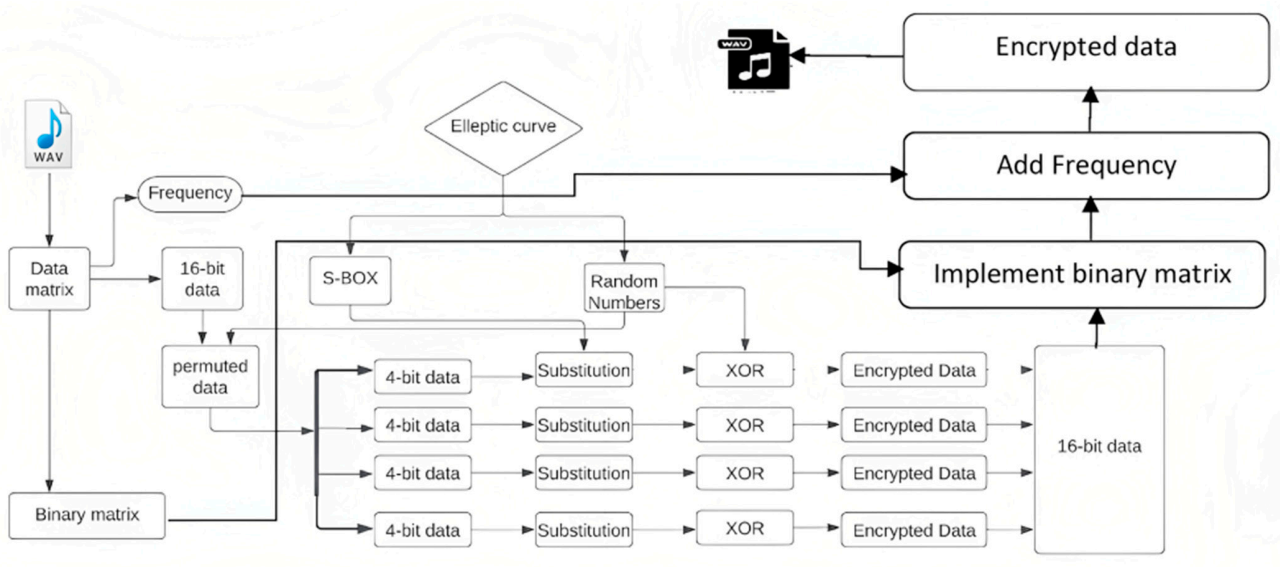

Figure 1. Flow chart of the proposed encryption scheme.

4. Security Analysis

A standard encryption scheme must defend against various attacks compromising confidentiality, integrity, non-repudiation, and data authentication. In this evaluation, we assess the significance and stability of the presented technique against various adversarial attacks. In the following sections, we subject the proposed scheme to several analyses outlined in the subsequent subsections.

4.1. Proposed S-Box Analysis

In this subsection, the proposed S-boxes robustness against various statistical and algebraic attacks is investigated by different tests, for instance, nonlinearity (NL), differential approximation probability (DP), strict avalanche criterion (SAC), and linear approximation probability (LP). The length between the Boolean functions of the S-box and the set of all affine functions is measured by the nonlinearity of the S-box. A Boolean vector function's upper bound can be computed using this formula: $2^{m-1} - 2^{(\frac{m}{2})-1}$. Thus, the maximum possible NL of any 4-bit S-box is six. Table 2 displays the average NL value of the proposed S-box, which stands at four, indicating its closeness to the optimal value. For an S-box to satisfy the SAC, a change in a single input bit must result in a change in half of the output bits. The proposed S-boxes successfully meet the SAC test, achieving a nearly optimal value of 0.5. The LP test of an S-box is used to calculate the highest value of coincident masked input bits with masked output bits [26]. A cryptographically strong S-box has the property that it attains a low score of LP. In Table 2, some of the newly constructed S-boxes by the proposed technique and their corresponding LP values are listed, which shows that the S-boxes based on the proposed scheme are suitable for secure communication against linear approximation attacks. Likewise, differential attacks are used to study the output differences for the corresponding input differences to obtain useful information [27]. For an S-box, the DP measures the strength of the S-box to prevent differential attackers. The additional evaluations presented in Table 2 demonstrate that the proposed S-boxes offer security against various attacks.

4.2. Proposed Audio Scheme Analysis

A standard encryption method must counter other attacks that try to thrash the data's confidentiality. Here, we analyze the strength and resilience of the suggested approach against diverse harmful attacks. MATLAB is used to perform all these calculations on a desktop computer. Using several keys, we encrypt and decrypt assorted audio models containing audio from multiple genres, such as music, speech, and other types of characters,

to test our algorithm. Figure 2 depicts the wave representation of plain and encrypted audio data. As the amplitude of the encrypted audio is uniform, as displayed in Figure 2, the original and encrypted audio bear no similarity. This indicates that audio encryption is working correctly. The following sections will perform different types of analysis on the proposed scheme, including histograms, entropy, complexity, key sensitivity, differential, correlations, and NIST analysis.

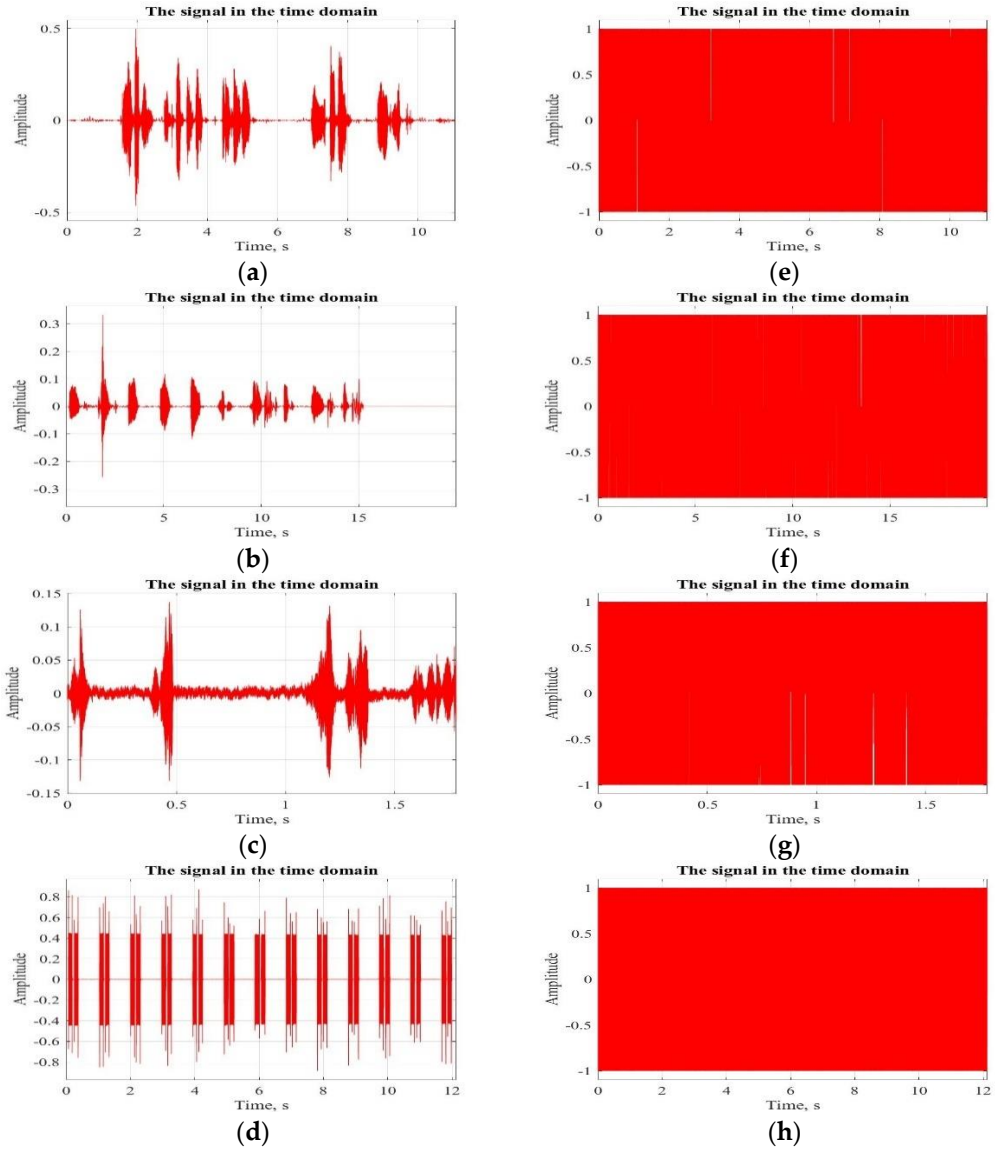

Figure 2. Waveforms of male, female, bird, and alarm (**a**–**d**) original audio (**e**–**h**) corresponding encrypted audio.

4.2.1. Spectrogram Analysis

Spectrum analysis is the fundamental instrument for analyzing sound. In terms of audio, the spectrogram is a two-dimensional graph in which a variety of hues represent the third dimension. As the name indicates, it is a description of frequency that changes over time. The hue in the third dimension indicates the sound's volume at a specific moment in time. Red and blue are used to specify low amplitude, whereas bright colors are used to signify maximum amplitude. Figure 3 shows the findings of our spectrogram study of our encryption method. Figure 3a,b depicts the spectrogram graphs of the original and encrypted audio files. The uniformity observed in the spectrogram graph of the encrypted

audio files proves that the audio file has been successfully encrypted. The spectrogram of this encrypted audio file differs significantly from that of the original in terms of amplitude and spectral shape.

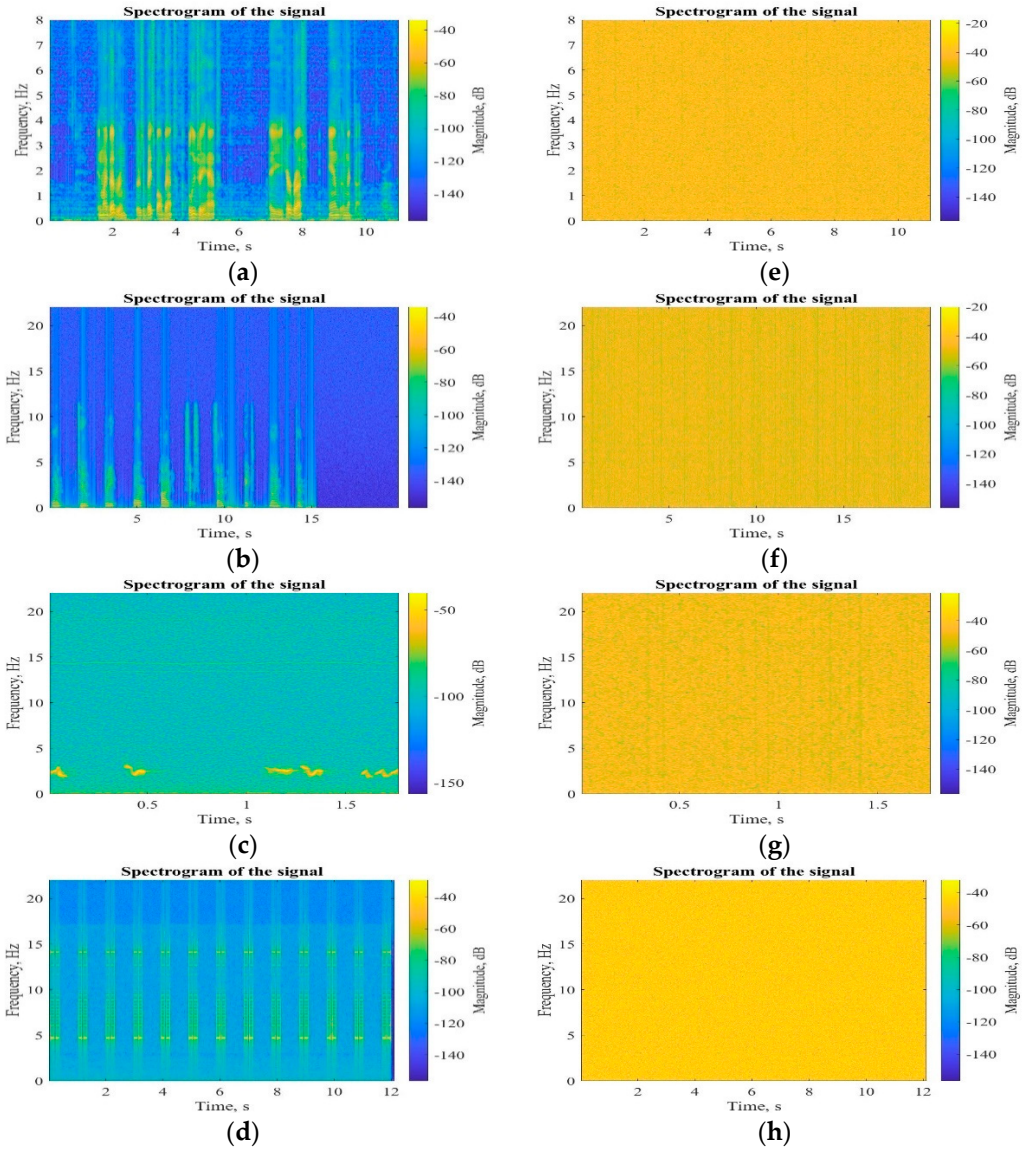

Figure 3. Spectrogram graph of the female, bird, male, and alarm (**a**–**d**) original audio (**e**–**h**) corresponding encrypted audio.

4.2.2. Histogram Analysis

Statistical attacks on cryptosystems can be tested using histogram analysis. It is conceivable that the cryptosystem will turn the original data into noise and generate unpredictability in the data. The most plausible scenario is that cryptosystems generate randomness in the data by turning the original information into noise. In a practical cryptosystem, the encrypted data probably does not provide any information that enables the encrypted data to be decoded without needing a secret key. Figure 4 shows the histogram statistics of our encryption algorithm. Figure 4a shows the histogram of the original audio, whereas Figure 4b shows the histogram of the cryptographed audio. One can see that the original audio signals have a chaotic and single-pointed histogram, while encrypted audio files have a uniform histogram.

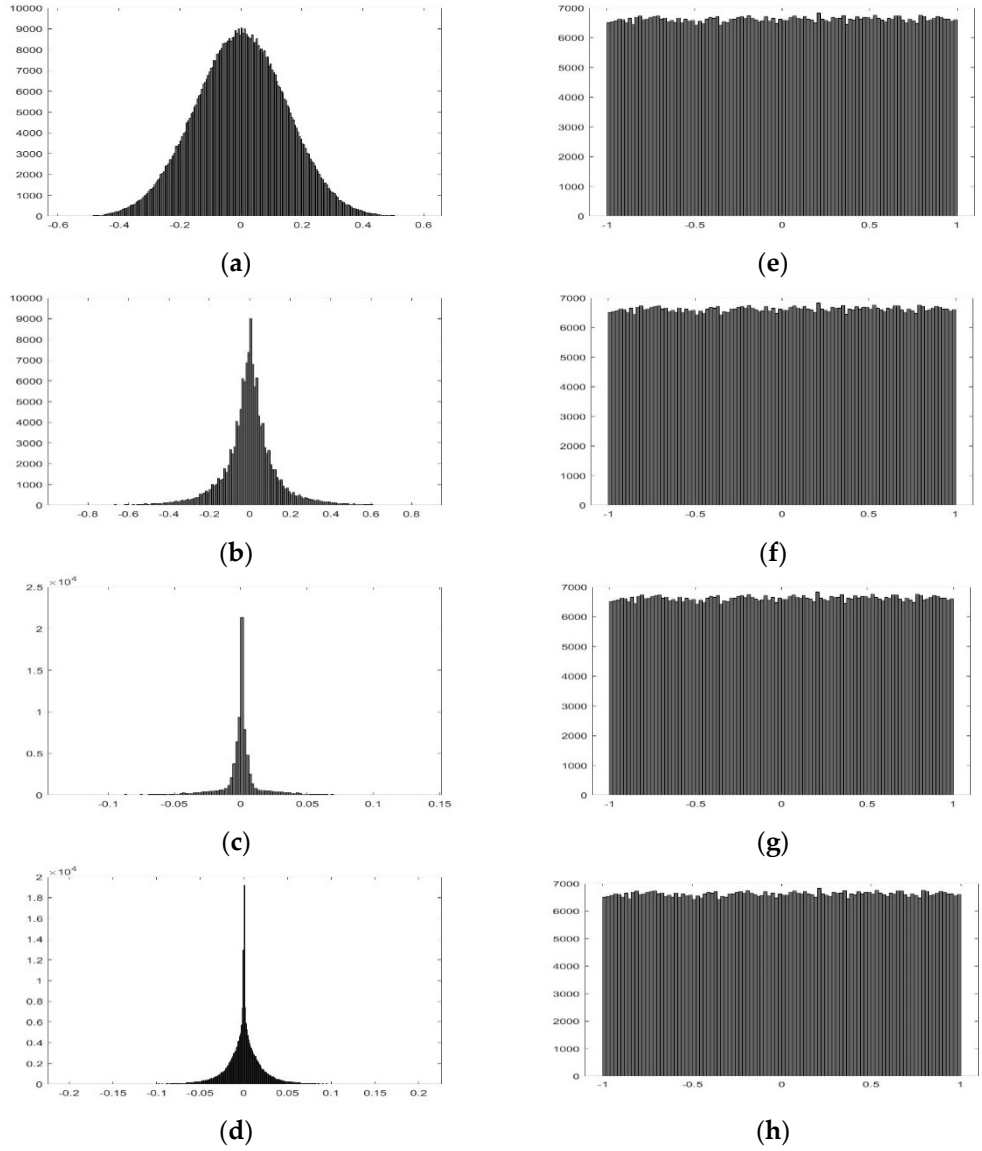

Figure 4. Histogram graphs of the female, alarm, male, and bird (**a**–**d**) original audio (**e**–**h**) corresponding encrypted audio.

4.2.3. Correlation Analysis

Any cryptosystem's resistance to statistical attacks can be assessed using several techniques, including the correlation coefficient because the segments of the data in multimedia data are highly connected. As a result, a reliable cryptosystem will prevent data segments from correlating. The correlation study analyses the correlation between the data segments that are most like each other. The correlation coefficient can be expressed mathematically as:

$$\alpha_{uv} = \frac{cov(p^*, q^*)}{\sqrt{D(p^*).D(q^*)}} \tag{10}$$

where

$$cov(p^*, q^*) = \frac{1}{w}\sum_{j=1}^{w}(p_j^* - \beta(p^*))(q_j^* - \beta(q^*)) \tag{11}$$

$$D(p^*) = \frac{1}{w}\sum_{j=1}^{w} p_j^* \beta(p^*) \tag{12}$$

and

$$\beta(p^*) = \frac{1}{w}\sum_{j=1}^{w} p_j^* \tag{13}$$

Samples p_j^* and q_j^* are used in the above equation to represent adjacent samples at the j_{th} position. Correlation analyses of data are frequently conducted in three directions: horizontal, vertical, and diagonal. Because we are working with audio data, we can only use correlation analysis for the single-string data in the horizontal direction. Table 3 displays the findings of the correlation study. The original audio correlation is precisely one, indicating that the audio data's various segments are highly correlated. Correlation analysis for audio deciphered by the suggested method is close to zero, signifying that the proposed method analytically interferes with the audio segment's correlations. Figure 5 depicts the correlations between the original and encrypted audio. Because the intercorrelation of the audio file is gradually reduced, this shows that our method is effective. As a result, we have designed our approach to be impervious to malevolent statistical attacks.

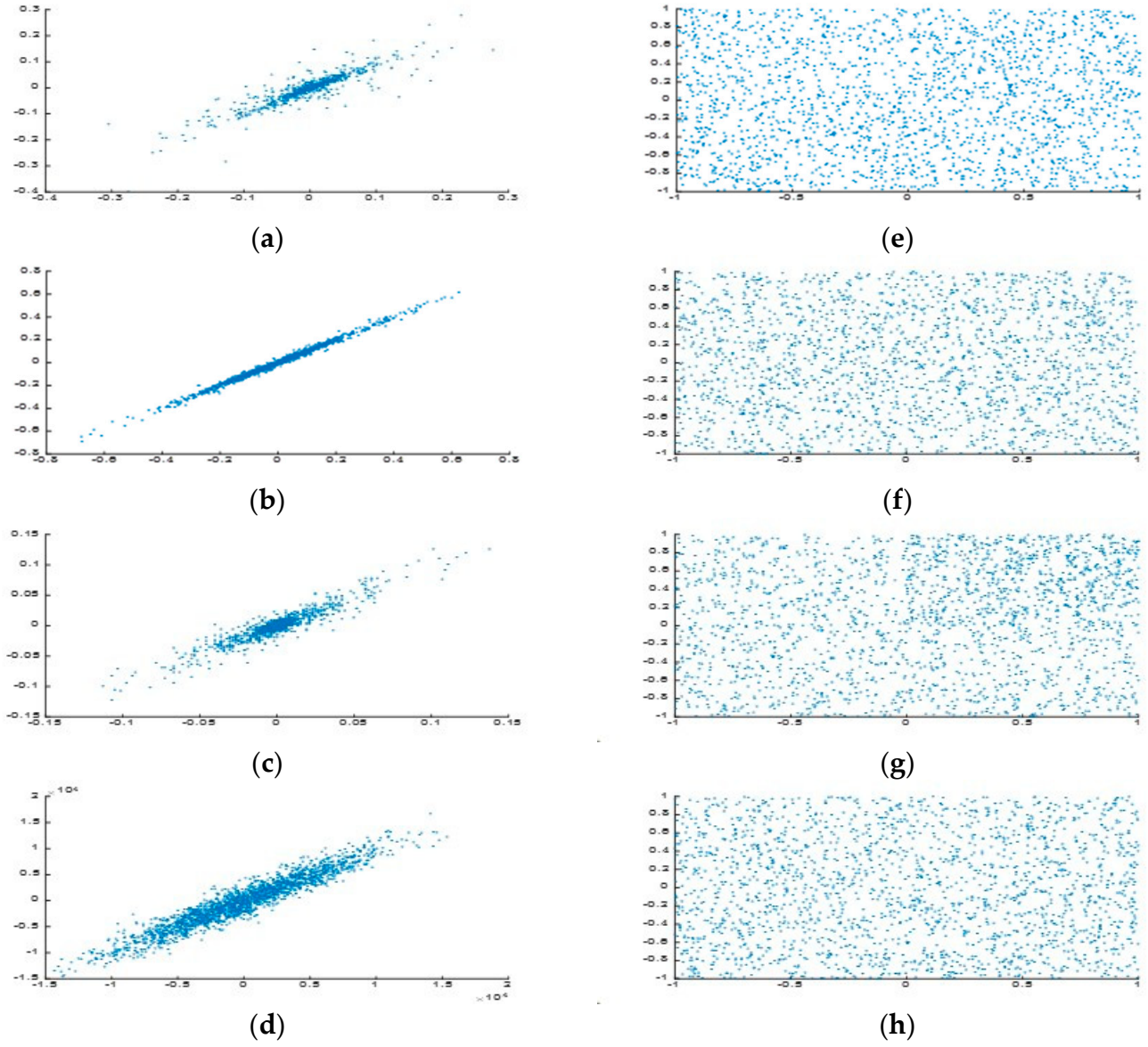

Figure 5. Correlation graph of the machine, animal, bird, and male (**a**–**d**) original audio (**e**–**h**) corresponding encrypted audio.

Table 3. Correlation Analysis and Comparison with existing Schemes.

No.	Audio	Plain Audio	Cipher Audio
1	Machine	0.9538	0.0013
2	Male	0.8910	0.0010
3	Audio	0.9935	0.0040
4	Animal	0.9953	0.0067
5	Sound	0.9847	0.0023
6	Alarm	0.7317	0.0012
7	Applause	0.8368	−0.0013
8	Female	0.9933	−0.0024
9	Bird	0.9321	−0.0018
10	Bells	0.9962	0.0017
11	Ref. [22]	-	−0.0020
12	Ref. [23]	-	0.0040
13	Ref. [28]	-	0.0016
14	Ref. [29]	-	0.0119

4.2.4. Information Entropy

Information entropy analysis can confine the degree of uncertainty in coded data. As the rate of uncertainty rises, the entropy of encrypted audio data also rises. Entropy can be expressed as:

$$H = -\sum_{k=0}^{l} P(k) \log_2(k) \tag{14}$$

Here, l is the grayscale of the audio file, and $P(k)$ is the chance of the gray-value k appearing. The audio size, usually quantified in kilobytes (KB), denotes the volume of digital storage capacity essential to store an audio file. The magnitude of an audio file is contingent upon several elements, including the audio format, sample rate, bit depth, and the duration of the recording. A 16-bit audio file of various sizes in (Kb) corresponds to a theoretical value of H. A secure cryptosystem is deemed in place when the ciphered file's information entropy is precisely 16. Table 4 shows the results of our information entropy analysis from our new proposed system. The table shows that our proposed technique's information value for every ciphered audio is equal to nearly 16, resulting in optimal confusion in the audio file. So, our system can withstand attacks from entropy.

Table 4. Entropy Analysis of Original and Encrypted Audios.

No.	Audio	Original	Cipher	Size
1	Machine	14.1688	15.898	26,000/Kb
2	Male	10.6914	15.945	345/Kb
3	Audio	14.8475	15.446	900/Kb
4	Animal	8.0065	15.554	530/Kb
5	Sound	14.8549	15.434	11,000/Kb
6	Alarm	9.8183	15.452	24,000/Kb
7	Applause	13.4401	15.923	783/Kb
8	Female	8.5125	15.986	32/Kb
9	Bird	4.5625	15.670	307/Kb
10	Bells	13.4216	15.345	32,000/Kb

4.2.5. Differential Analysis

The two most used techniques for analyzing differential attacks are the number of pixel change rates (NPCR) and the unified average changing intensity (UACI). The sensitivity of the cryptosystem is calculated by these two. Cryptographic algorithms must be sensitive so that even the slightest change in original data results in an incredibly diverse ciphertext. Mathematically, NPCR and UACI are defined as:

$$NPCR = \frac{\sum_{a^*,b^*} f(a^*,\ b^*)}{K^*} \times 100\% \tag{15}$$

In Equation (15), K^* represents the cardinality of audio data set and $f(a^*, b^*)$ is defined as:

$$f(a^*, b^*) = \begin{cases} 1 & if \quad N_1(a^*, b^*) = N_2(a^*, b^*) \\ 0 & if \quad N_1(a^*, b^*) \neq N_2(a^*, b^*) \end{cases} \tag{16}$$

and UACI is defined as:

$$UACI = \frac{1}{K^*}\left[\sum_{a^*,b^*} \frac{abs\left(N_1^*(a^*, b^*) - N_2^*(a^*, b^*)\right)}{2^{k^*} - 1}\right] \times 100\% \tag{17}$$

where the value 2^{k^*} denotes the bit order in the audio data collection. Tables 5 and 6 show the results of our testing of the suggested audio encryption method utilizing the NPCR and UACI analyses. We take an audio file in this study and change its 1st and 100th byte, respectively. As the alarm audio file has 85,529 bytes, firstly, we change its first byte and analyze, and then at the 100th place, the same procedure is applied, and the results are displayed in Table 5. Similarly, we analyzed all these audio files and evaluated the value of NPCR. Furthermore, we considered all the UACI values of these audio files by making small changes, as given in Table 6. This strategy is expected to nullify differential attacks.

Table 5. NPCR Analysis and their Comparison.

No.	Audio	NPCR-1	NPCR-2
1	Machine	99.9967	99.9977
2	Male	99.9983	99.9979
3	Audio	99.9865	99.9856
4	Animal	99.9976	99.9923
5	Sound	99.9845	99.9847
6	Alarm	99.9983	99.9924
7	Applause	99.9973	99.9987
8	Female	99.9967	99.9923
9	Bird	99.9989	99.9912
10	Bells	99.9992	99.9990
11	Ref. [29]	99.9972	-
12	Ref. [30]	99.9989	-
13	Ref. [31]	99.6521	-

Table 6. UACI Analysis and their Comparison.

No.	Audio	UACI-1	UACI-2
1	Machine	33.1988	33.1982
2	Male	33.9501	33.9512
3	Audio	33.7656	33.7662
4	Animal	33.8734	33.8765
5	Sound	33.9002	33.9009
6	Alarm	33.8377	33.8323
7	Applause	33.8969	33.8945
8	Female	33.9082	33.9024
9	Bird	33.8003	33.8014
10	Bells	33.9016	33.9040
11	Ref. [29]	-	-
12	Ref. [30]	33.3421	-
13	Ref. [31]	33.2122	-

4.2.6. Peak Signal-to-Noise Ratio (PSNR)

The following equations can be used to obtain the mean squared error for two vectors, U^* and V^*:

$$MSE = \frac{1}{n^*}\sum_{j=1}^{n^*}(U^*[j] - V^*[j]) \tag{18}$$

If U^* is the host audio file and V^* is its encoded version, then the PSNR of an audio version can be calculated as follows:

$$PSNR = 10 * \log_{10}\left(\frac{MAX^2}{MSE}\right) \tag{19}$$

where MAX represents the stream's highest possible value. Using this method, we were able to calculate the PSNR for each of the audio recordings and present them in Table 7. It should be emphasized that the numbers are quite small. It is preferable to have lower PSNR values for encrypted audio files because of the increased noise in these files and the increased resistance against attacks that these files provide.

Table 7. PSNR and MSE Analysis.

No.	Audio	PSNR	MSE
1	Animal	10.5881	3.2456×10^4
2	Male	10.6779	3.2655×10^4
3	Alarm	10.8705	3.2664×10^4
4	Audio	10.4673	3.2664×10^4
5	Female	10.8820	3.2634×10^4
6	Machine	10.6799	3.2614×10^4
7	Bird	10.8908	3.2564×10^4
8	Applause	10.2306	3.2344×10^4
9	Sound	9.7340	3.2564×10^4
10	Audio	10.1106	3.2534×10^4

4.2.7. Root Mean Square (RMS) and Crest Factor (CF) Value

The RMS method is used to determine the average amplitude of an audio signal. If the input signal has a mean value of zero, the RMS and standard deviation are the same and calculated as:

$$RMS = \sqrt{\frac{1}{N^*}\sum_{j=1}^{N^*}\left|A_j^*\right|^2} \tag{20}$$

The crest factor (CF) is a waveform parameter that measures the ratio of peak values to the effective value. Its primary purpose is to identify the minimum possible value for the peaks in a waveform. A CF ratio of 0 dB indicates no peaks, resembling a direct current (DC) signal. A higher CF value corresponds to the presence of peaks. The CF is calculated as follows:

$$CF = 20\log_{10}\left(\frac{V_{peak}^*}{V_{RMS}^*}\right) \tag{21}$$

Table 8 displays the results of RMS and CF calculations for the given values in the proposed algorithm. The coded audio files exhibit approximately 0.56 RMS and 4.7 CF values, as shown in Table 8. The absence of a statistical correlation between the host audio files and the coded audio files provides evidence that there is no such correlation.

Table 8. RMS and CF values.

No.	Audio	RMS	CF
1	Machine	0.5772	4.7732
2	Male	0.5764	4.7853
3	Audio	0.5871	4.7933
4	Animal	0.57731	4.7715
5	Sound	0.58637	4.9057
6	Alarm	0.59943	4.8809
7	Applause	0.58943	4.9909
8	Female	0.57637	4.7857
9	Bird	0.56943	4.8909
10	Bells	0.56983	4.7123

4.2.8. Complexity Analysis

Complex analysis measures the resources, such as time and memory, to execute. There are various ways to determine the complexity of the algorithm. However, the most common method is a big 'O' notation. In this section, we analyzed the proposed scheme using big O. Since the proposed scheme is a substitution permutation network, the scheme initially generates S-boxes and utilized them for substitution. Afterwards, the method uses the generated numbers in the encryption process. For the permutation module, the generated sequence is used and shuffles the audio data in linear time. Therefore, the permutation module permutes the data in $O(M \times N)$. Similarly, constant time is required to substitute the fixed pixel. So, the permutation module permutes the data in $O(M \times N)$. So overall, the time complexity of the proposed algorithm is $O(M \times N)$.

4.2.9. Key Sensitivity

To assess the overall behavior of the encryption method, we conducted a test where we employed nearly identical keys for encrypting and decrypting the audio files. The decryption key was obtained by modifying a single digit in one of the parameters used to generate the key space. Figures 6 and 7 show that decryption is unsuccessful despite using a similar secret key. We just changed the value from five to seven and generated different sequences by using the technique presented in Section 3.1, which is utilized for encryption and decryption. A single-digit change in the key causes decryption to fail. The experiment provides evidence of the suggested audio encryption algorithm's high key sensitivity.

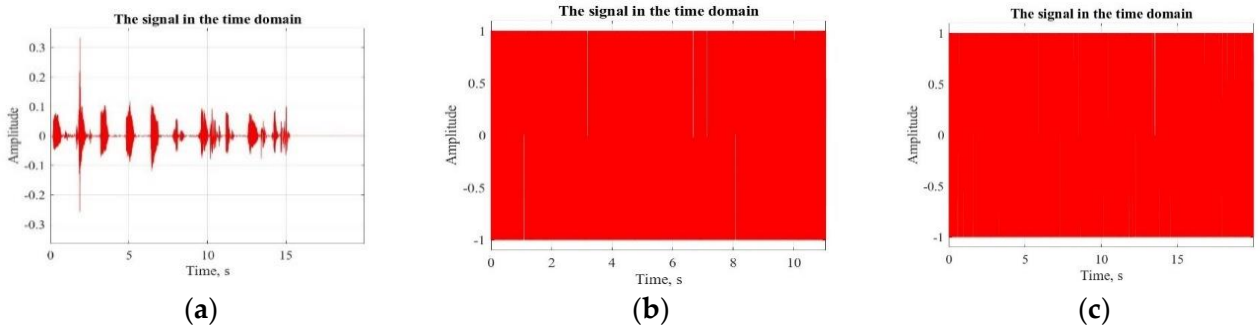

Figure 6. Waveforms plotting key sensitivity, (**a**) original audio (**b**) encrypted file using a secret key (**c**) decrypted file using the secret key with change.

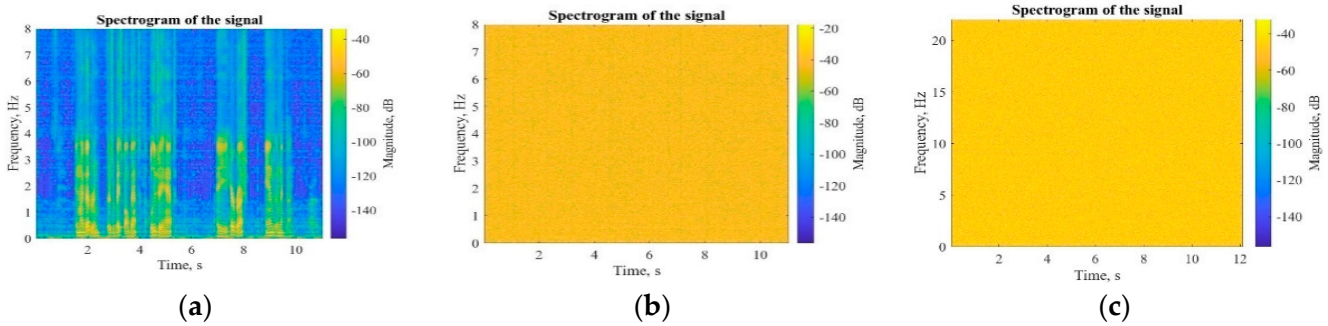

Figure 7. Spectrogram plotting Key Sensitivity, (**a**) original audio (**b**) encrypted audio using a secret key (**c**) decrypted file using the secret key with change.

4.2.10. NIST Statistical Attack

To evaluate the random number generator for cryptographic applications, we examined the sequence generated by the suggested random number generator. We first convert the created sequence to binary so that the NIST test can be used to determine its randomness [32]. Table 9 shows the sixteen statistical tests that make up the NIST statistical

test. Because it passed every randomness test, our method generates high-quality random numbers that may be used in various audio encryption schemes.

Table 9. NIST Statistical Analysis.

No.	Test Type	*p*-Value
1	Frequency test (Monobit)	0.925896567576757
2	Frequency Test	0.347657845646547
3	Discrete Fourier	0.197554645648668
4	Longest Run T	0.476574564654666
5	Run T	0.465265624677676
6	Non overlapping	0.686596575675757
7	Overlapping	0.189566757875478
8	Maurer's Universal	0.986525262525789
9	Binary Rank T	0.765635634532988
10	Linear Complexity	0.465343898603582
11	Approximate Entropy	0.057539774374433
12	Cumulative Sums (Forward)	0.964980450801528
13	Cumulative Sums (Reverse)	0.992180553046622
14	Serial	0.106776358367435
15	Random Excursions Test:	0.013685365347784
	State	*p*-value
	−4	0.616753379242737
	−3	0.736524242873434
	−2	0.624245262362632
	−1	0.538534724232327
	1	0.528535982599735
	2	0.438539422492473
	3	0.523853734738434
	4	0.438535635935979
16	Randomexcursions variant test:	*p*-value
	State	0.535635693593953
	−9	0.335683593496424
	−8	0.324738368267326
	−7	0.426534737483848
	−6	0.519544771283753
	−5	0.382974453828231
	−4	0.619382974453828
	−3	0.763273652388728
	−2	0.235394349393939
	−1	0.135939539393737
	1	0.253593943737939
	2	0.135395359343937
	3	0.534774348383838
	4	0.535353955935935
	5	0.738343473973939
	6	0.838535384829293
	7	0.735839534398268
	8	0.593635354378378
	9	0.538534272843343

5. Conclusions

In this paper, a novel cryptographic algorithm for digital audio encryption is presented. The proposed scheme is designed to follow the substitution permutation network architecture. The substitution and permutation module that is used for confusion and diffusion is based on the arithmetic operations of an EC over the extension field of the binary field $\mathbb{Z}_2$. Initially, the scheme used a particular type of EC and disordered the pixel positions of the plain audio. The result of this step demolishes the horizontal intra-correlation among the pixels of the audio. Since the audio data consist of sixteen-bit elements for the confusion module, the scheme generates multiple 4×4 good quality S-boxes. The generated S-boxes

are then used in parallel and substitute the pixels of the disordered audio. In the final step of the encryption, the procedure produced quality random numbers using the EC operation and added the sequences with substituted data using the addition operation of the extension field. The simulation illustrates that the proposed encryption method effectively secures the audio content, transforming it into a uniform, unidentifiable sound pattern. Furthermore, the system undergoes rigorous testing against various attack scenarios, with the performance analysis indicating that the proposed encryption approach demonstrates exceptional resilience to statistical and differential attacks.

Author Contributions: Conceptualization, H.U.R., T.S. and D.S.; Software, H.U.R. and D.S.; Writing—original draft, H.U.R.; Writing—review and editing, M.M.H., T.S. and Z.B. All authors have read and agreed to the published version of the manuscript.

Funding: The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through large group Research Project under grant number RGP2/5/44.

Data Availability Statement: No associated data related to that article.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

EC	elliptic curve
S-box	substitution box
PRNs	pseudo-random numbers
PIP	primitive irreducible polynomial

References

1. Ghazvini, M.; Mirzadi, M.; Parvar, N. A modified method for image encryption based on chaotic map and genetic algorithm. *Multimed. Tools Appl.* **2020**, *79*, 26927–26950. [CrossRef]
2. Iqbal, N.; Hanif, M.; Rehman, Z.U.; Zohaib, M. On the novel image encryption based on chaotic system and DNA computing. *Multimed. Tools Appl.* **2022**, *81*, 8107–8137. [CrossRef]
3. Alghafis, A.; Waseem, H.M.; Khan, M.; Jamal, S.S. A hybrid cryptosystem for digital contents confidentiality based on rotation of quantum spin states. *Stat. Mech. Its Appl.* **2020**, *554*, 123908. [CrossRef]
4. Pourasad, Y.; Ranjbarzadeh, R.; Mardani, A. A new algorithm for digital image encryption based on chaos theory. *Entropy* **2021**, *23*, 341. [CrossRef]
5. Liu, H.; Liu, Y. Cryptanalyzing an image encryption scheme based on hybrid chaotic system and cyclic EC. *Opt. Laser Technol.* **2014**, *56*, 15–19. [CrossRef]
6. Ullah, I.; Azam, N.A.; Hayat, U. Efficient and secure substitution box and random number generators over Mordell ECs. *J. Inf. Secur. Appl.* **2021**, *56*, 102619.
7. Rehman, H.U.; Shah, T.; Aljaedi, A.; Hazzazi, M.M.; Alharbi, A.R. Design of nonlinear components over a mordell elliptic curve on Galois fields. *Comput. Mater. Contin.* **2022**, *71*, 1313–1329.
8. Ramzan, M.; Shah, T.; Hazzazi, M.M.; Aljaedi, A.; Alharbi, A.R. Construction of s-boxes using different maps over ECs for image encryption. *IEEE Access* **2021**, *9*, 157106–157123. [CrossRef]
9. Liu, Z.; Guo, Q.; Xu, L.; Ahmad, M.A.; Liu, S. Double image encryption by using iterative random binary encoding in gyrator domains. *Opt. Express* **2010**, *18*, 12033–12043. [CrossRef]
10. Andono, P.N. Improved pixel and bit confusion-diffusion based on mixed chaos and hash operation for image encryption. *IEEE Access* **2022**, *10*, 115143–115156. [CrossRef]
11. Ibrahim, S.; Alharbi, A. Efficient image encryption scheme using Henon map, dynamic S-boxes and EC cryptography. *IEEE Access* **2020**, *8*, 194289–194302. [CrossRef]
12. Abbas, A.M.; Alharbi, A.A.; Ibrahim, S. A novel parallelizable chaotic image encryption scheme based on ECs. *IEEE Access* **2021**, *9*, 54978–54991. [CrossRef]
13. Servetti, A.; Martin, J.C. Perception-based partial encryption of compressed speech. *IEEE Trans. Speech Audio Process.* **2002**, *10*, 637–643. [CrossRef]
14. Thorwirth, N.J.; Horvatic, P.; Weis, R.; Zhao, J. Security methods for MP3 music delivery. In Proceedings of the Record of the Thirty-Fourth Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 29 October–1 November 2000; Volume 2, pp. 1831–1835.

15. Servetti, A.; Testa, C.; De Martin, J.C. Frequency-selective partial encryption of compressed audio. In Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, Hong Kong, China, 6–10 April 2003; pp. 660–668.
16. Bhargava, B.; Shi, C.; Wang, S.Y. MPEG video encryption algorithms. *Multimed. Tools Appl.* **2004**, *24*, 57–79. [CrossRef]
17. Grangetto, M.; Magli, E.; Olmo, G. Multimedia selective encryption by means of randomized arithmetic coding. *IEEE Trans. Multimed.* **2006**, *8*, 905–917. [CrossRef]
18. Yan, W.Q.; Fu, W.G.; Kankanhalli, M.S. Progressive audio scrambling in compressed domain. *IEEE Trans. Multimed.* **2008**, *10*, 960–968. [CrossRef]
19. Lima, J.B.; da Silva Neto, E.F. Audio encryption based on the cosine number transform. *Multimed. Tools Appl.* **2016**, *75*, 8403–8418. [CrossRef]
20. Naskar, P.K.; Paul, S.; Nandy, D.; Chaudhuri, A. DNA encoding and channel shuffling for secured encryption of audio data. *Multimed. Tools Appl.* **2019**, *78*, 25019–25042. [CrossRef]
21. Babu, N.R.; Kalpana, M. Balasubramaniam. A novel audio encryption approach via finite-time synchronization of fractional order hyperchaotic system. *Multimed. Tools Appl.* **2021**, *80*, 18043–18067. [CrossRef]
22. Shah, D.; Shah, T.; Ahamad, I.; Haider, M.I.; Khalid, I. A three-dimensional chaotic map and their applications to digital audio security. *Multimed. Tools Appl.* **2021**, *80*, 22251–22273. [CrossRef]
23. Shah, D.; Shah, T.; Hazzazi, M.M.; Haider, M.I.; Aljaedi, A. An Efficient audio encryption scheme based on finite fields. *IEEE Access* **2021**, *9*, 144385–144394. [CrossRef]
24. Qureshi, A.; Shah, T. S-box on subgroup of Galois field based on linear fractional transformation. *Electron. Lett.* **2017**, *53*, 604–606. [CrossRef]
25. Farwa, S.; Sohail, A.; Muhammad, N. A novel application of ECs in the dynamical components of block ciphers. *Wirel. Pers. Commun.* **2020**, *5*, 1309–1316. [CrossRef]
26. Adams, C.; Tavares, S. The structured design of cryptographically good S-boxes. *J. Cryptol.* **1990**, *3*, 27–41. [CrossRef]
27. Biham, E.; Shamir, A. Differential cryptanalysis of DES-like cryptosystems. *J. Cryptol.* **1991**, *4*, 3–72. [CrossRef]
28. Kordov, K. A novel audio encryption algorithm with permutation-substitution architecture. *Electronics* **2019**, *8*, 530. [CrossRef]
29. Sathiyamurthi, P.; Ramakrishnan, S. Speech encryption using chaotic shift keying for secured speech communication. *EURASIP J. Audio Speech Music. Process.* **2017**, *1*, 20. [CrossRef]
30. Farsana, F.J.; Devi, V.R.; Gopakumar, K. An audio encryption scheme based on fast walsh hadamard transform and mixed chaotic keystreams. *Appl. Comput. Inform.* **2020**, *19*, 239–264. [CrossRef]
31. Habib, Z.; Khan, J.S.; Ahmad, J.; Khan, M.A.; Khan, F.A. Secure speech communication algorithm via DCT and TD-ERCS chaotic map. In Proceedings of the 4th International Conference on Electrical and Electronic Engineering (ICEEE), Ankara, Turkey, 8–10 April 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 246–250.
32. Bassham, L.; Rukhin, A.; Soto, J.; Nechvatal, J.; Smid, M.; Barker, E.; Leigh, M.; Levenson; Vangel, M.; Banks, D.; et al. *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*; NIST Special Publication, Tech. Rep. 800-22 Rev 1a; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2010.