A Polar Code-based Audio Encryption Scheme for Secure Wireless Communication Systems

Dinh Van Linh [1], Tran Duc Thiep [1], Pham Van Tiep [1], Pham Manh Tuan [2], and Pham Duy Phong [3,*]

[1] Faculty of Electronics and Microcircuits, Academy of Cryptography Techniques, Hanoi, Vietnam
[2] Department of Computer Engineering, School of Information and Communications Technology, Hanoi University of Science and Technology, Hanoi, Vietnam
[3] Faculty of Electronics and Telecommunications, Electric Power University, Hanoi, Vietnam
Email: vanlinh@actvn.edu.vn (D.V.L.); thiept598@gmail.com (T.D.T.); ptiep94@gmail.com (P.V.T.); tuanpm@soict.hust.edu.vn (P.M.T.); Phongphd@epu.edu.vn (P.D.P.)
*Corresponding author

***Abstract*—Secure transmission of multimedia files, particularly audio, is a challenging task in secure wireless communication systems. As a result, encryption methods are progressively becoming essential to meet the needs of audio content security. This work proposes an audio encryption method that provides an effective security level. Specifically, the secret key distilled from wireless channel properties is processed through the Advanced Encryption Standard-Counter (AES-CTR) mode to generate the corresponding sub-keys for the audio data blocks. Each sub-key is divided into two parts: the first part is Exclusive-OR (XOR) with the information bits that carry the audio data, while the remaining part is allocated to the frozen bits. The experiments demonstrate that the proposed technique achieves good results and maintains robust effectiveness across various attack scenarios.**

***Keywords*—audio encryption, polar codes, information bits, frozen bits, secret key, wireless channel characteristics**

I. Introduction

The security of audio signals plays an important role in Fifth-Generation (5G) networks, as audio signals not only facilitate convenient information transmission and communication, but also create potential opportunities for criminal misuse [1]. Encryption methods are commonly employed to protect audio signals. However, an audio signal is more difficult to encrypt than different types of data since there is an association between adjacent samples. Various audio encryption methods have been proposed, such as mathematical-based approaches [2–4], chaos-based schemes [5–14], or chaotic maps combined with Deoxyribonucleic Acid (DNA) [15–18]. Mathematical-based audio encryption methods typically exhibit weak security due to their deterministic structure, limited randomness, and vulnerability to statistical and cryptanalytic attacks. Chaos-based audio encryption schemes provide high sensitivity and randomness. Recent advancements have further diversified these approaches; for instance, the utilization of memristor-based chaotic systems [7] has demonstrated robust security metrics for audio encryption. Furthermore, cutting-edge physical-layer cryptographic techniques, such as those leveraging optical chaos and photonics [19], highlight a growing and highly effective trend of integrating physical hardware properties for robust data protection. However, they often suffer from parameter instability and limited cryptographic rigor. Many of these methods are vulnerable to known-plaintext and differential attacks if not carefully designed. Moreover, their computational complexity can be relatively high. Although DNA-based approaches introduce additional randomness, they often incur high complexity and lack strong theoretical security guarantees. In addition, their practical implementation is relatively challenging.

In recent years, crypto-coding has emerged as a physical layer security approach that has attracted considerable research attention thanks to its integration of data encryption and channel coding into a single unified processing stage. Crypto-coding techniques can be implemented on Turbo codes [20–23] and polar codes [24–27]. However, crypto-coding techniques based on polar codes offer more significant advantages, as polar codes are currently adopted in 5G networks. While recent crypto-coding schemes based on polar codes [24–27] have successfully integrated data encryption and channel coding, they typically rely on bit manipulation with a single, static shared key. The security effectiveness of these conventional methods is often validated only against passive eavesdropping attacks. Moreover, utilizing a fixed shared key across multiple data blocks can compromise systemic security, rendering it vulnerable to active attacks or key-reuse exploits. To address these limitations and advance beyond existing frameworks, this study proposes an advanced audio encryption method that deliberately bridges Physical-Layer Security (PLS) and formal cryptography. This method significantly extends our previous work reported in Ref. [25]. While prior works

Manuscript received March 16, 2026; revised April 14, 2026; accepted May 15, 2026; published October 30, 2026.

often treat key generation and coding as somewhat isolated or rely on static mapping, our proposed framework tightly couples them [25]. The conceptual contribution extends beyond simply substituting a static key with a dynamic one; it leverages the Advanced Encryption Standard-Counter (AES-CTR) mode is used to dynamically partition the sub-keys derived from wireless channel properties. By explicitly allocating one segment of the dynamic sub-key to mask the information bits and the other to randomize the frozen bits, the proposed architecture achieves a dual-layer security mechanism. It mathematically obfuscates the audio data while simultaneously inducing severe error propagation at an unauthorized decoder.

TABLE I. QUALITATIVE COMPARISON OF THE PREVIOUS METHODS AND OUR METHOD

Features	Methods				
	[24]	[25]	[26]	[27]	Proposed Scheme
Cryptographic Core	Chaotic sequences	Physical attributes	Random hopped scrambling	Channel entropy combinations	AES-CTR Stream Cipher
Key Dynamics	Static shared key	Single key extracted from channel	Pseudo-random hopping	Hybrid entropy mapping	Block-wise dynamic sub-keys
Frozen Bit Handling	Scrambled via chaos	Pseudo-randomly assigned	Random hopped	Fixed/Contextual	Dynamic randomization via AES keystream
Security Mechanism	Data obfuscation	PLS signal degradation	Positional uncertainty	Rate & entropy enhancement	SC decoding path disruption (Avalanche effect)

To explicitly highlight the novelty and technical differentiation of the proposed scheme against recent advancements in polar-based crypto-coding (specifically [24–27]), Table I provides a comprehensive, itemized technical comparison. While prior arts successfully merge encryption with channel coding, they predominantly rely on chaotic sequences or static shared keys. These conventional approaches often leave static anchor points in the codeword, making them vulnerable to structural cryptanalysis or key-reuse exploits. In contrast, the proposed scheme introduces the standardized AES-CTR mode as a dynamic sub-key expansion engine. By utilizing a block-specific counter mechanism, the AES-CTR engine continuously generates unique, non-repeating sub-keys for each audio frame. This enables the unpredictable and dynamic randomization of the frozen bits. Rather than merely obfuscating data, this architectural shift uniquely transforms the inherent error-propagation sensitivity of the physical-layer Successive Cancellation (SC) decoder into a robust structural security advantage. Specifically, the continuous disruption of frozen bits completely eliminates the static anchor points typically exploited by adversaries, deliberately triggering a catastrophic avalanche effect in any unauthorized decoding path. This integration ensures that the coding layer contributes meaningfully to structural security against active attacks, overcoming the incremental limitations of existing polar-code-based crypto-coding. The results show that the proposed approach encrypts/decrypts well and is highly effective in resisting attacks on encrypted audio files. Several contributions are summarized as follows:

- The shared key is derived from the properties of the wireless communication channel, making this proposed approach effective against passive eavesdropping attacks.
- The Advanced Encryption Standard (AES) key derivation function in Counter (CTR) mode is incorporated to reinforce the security of the secret key instead of relying on a single fixed key.
- The audio data contained in the information bits is protected using one part of the sub-key, while the remaining part is stored in the frozen bits to enhance diffusion and confusion. This approach enhances audio encryption performance, withstands passive eavesdropping attacks, and ensures dependable audio data delivery over wireless links.

To guide the reader through the study, the remainder of this article is organized as follows. Section II details the proposed hybrid crypto-coding framework, encompassing the physical-layer key derivation mechanism, the dynamic AES-CTR sub-key expansion, and the joint polar encoding/decoding architecture. Section III provides the simulation outcomes and discussion. Finally, Section IV summarizes the core contributions of this work and outlines potential trajectories for future research, including hardware implementation and formal information-theoretic security proofs.

II. PROPOSED AUDIO ENCRYPTION METHOD BASED ON POLAR CODES

We propose a crypto-coding system architecture based on polar codes for encrypting audio signals. The system uses a secret key generated from the wireless channel and expands it into sub-keys by applying AES in CTR mode, enabling simultaneous control of the information bits and frozen bits of the polar codes. The general model is shown in Fig. 1.

A. Key Generation Mechanism and Subkey Hierarchy

In this study, the primary focus is on the novel architectural integration of the derived key into the polar coding structure.

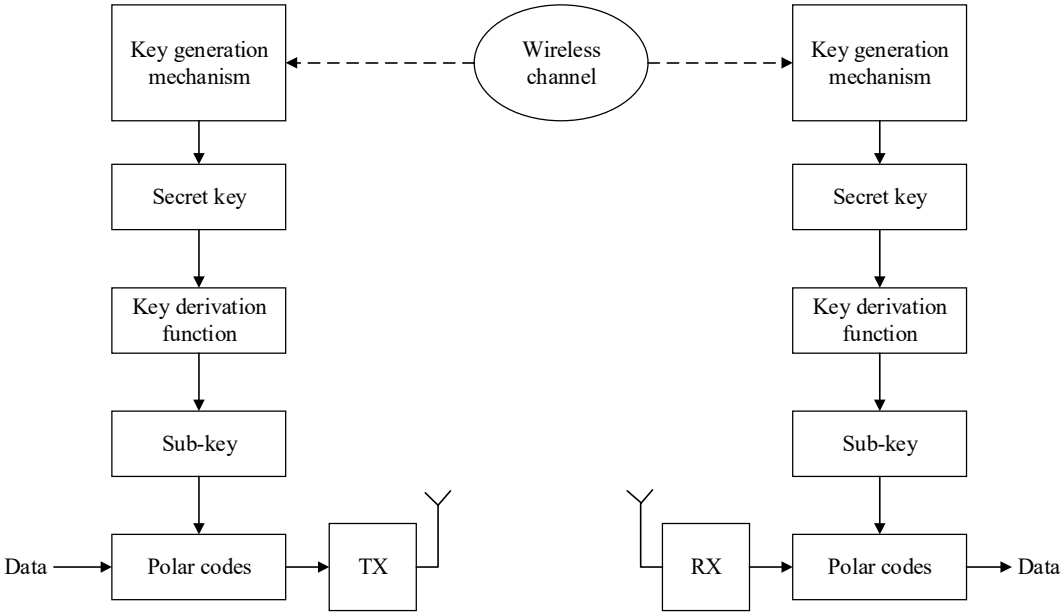


Fig. 1. The general model of the proposed method.

Let K_{secret} denote the secret key of the symmetric key encryption system. To ensure dynamic security, the AES/CTR algorithm is used as the key derivation function. For the i^{th} input audio data block (S_i), the sub-key is generated as follows Eq. (1):

$$K^i = \text{AES-CTR}\left(K_{\text{secret}}, IV, Ctr_i\right) \tag{1}$$

where IV is the initialization vector, and Ctr_i is the counter value of block i.

The sub-key sequence K^i has a length equal to the polar codes' length N, and is mapped into two index sets:

$$K^i = \left\{K^i_{info}, K^i_{frozen}\right\} \tag{2}$$

K^i_{info} (length K): the payload length mapped to the information bits.

K^i_{frozen} (length $N-K$): Used to insert into frozen bits.

It is crucial to explicitly acknowledge that practical physical-layer key agreement is highly non-trivial and acts as a strict operational bottleneck for the proposed scheme. The effectiveness of this framework against passive eavesdropping is inherently conditioned upon a sufficiently high-entropy channel and robust reciprocity. Practical impairments, such as low-Key Generation Rates (KGR) in static environments, or high Key Mismatch Rates (KMR) due to high mobility, imperfect reciprocity, and severe noise, directly compromise the systemic security. Specifically, because the AES-CTR expansion amplifies any key mismatch through the avalanche effect, a failure in the initial key agreement phase will not only weaken security but also critically disrupt the legitimate receiver's decoding process. While the detailed protocols to optimize KGR and minimize KMR are treated as preceding modular stages, we emphasize that the security claims presented herein are strictly valid only under the premise of a successfully negotiated, high-entropy initial secret key.

B. Encoding Process

The polar codes (N, K) classify N transmission channels into a set of reliable channels A and a set of noisy channels A_c. The input audio data D is divided into Q small blocks D_i of length K.

$$D = \bigcup_{i=1}^{Q} D_i \tag{3}$$

The proposed process for handling the audio block D_i (of length K) is carried out in three steps to achieve the objective of the joint encryption–coding process. The detailed process is shown in Fig. 2.

To ensure confidentiality, the data D_i is concealed by an XOR operation with each subkey extracted from the AES-CTR key stream. The resulting bit vector corresponding to the information set A is given by:

$$u_A = D_i \oplus K^i_{info} \tag{4}$$

Unlike conventional polarcoding, where frozen bits are fixed to zero, the proposed scheme fills the frozen set Ac using a random sub-key sequence K^i_{frozen} :

$$u_{A^c} = K^i_{frozen} \tag{5}$$

While AES-CTR independently guarantees formal computational security against cryptanalysis, its integration into the polar coding structure is not redundant; rather, it establishes a defense-in-depth mechanism. Standard cryptography assumes the eavesdropper

perfectly receives the ciphertext. By dynamically randomizing the frozen bits, our scheme denies the eavesdropper this prerequisite. Without the exact subkey to correctly assign the frozen bits, the Successive Cancellation (SC) decoder experiences catastrophic error propagation at the physical layer. Consequently, the eavesdropper recovers only pure noise ($c \approx 0.5$) instead of a structured ciphertext. Thus, the polar coding layer meaningfully contributes to structural security by physically degrading the intercepted signal, while AES-CTR ensures computational invulnerability.

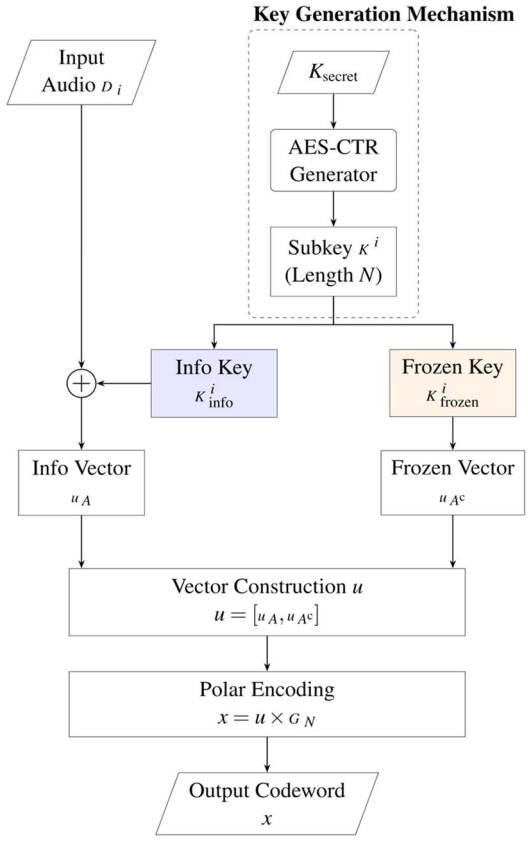


Fig. 2. Block diagram of the proposed joint encoding architecture.

The insertion of the derived subkey sequence enhances the diffusion property and disrupts the decoding process at the eavesdropper, resulting in severe error propagation under Successive Cancellation (SC) decoding.

The composite input vector $u = [u_A, u_{A^c}]$ is mapped to the corresponding positions and encoded using the Polar generator matrix *GN* to produce the transmitted codeword:

$$x = u \times G_N = u \times \left(B_N \times F^{\otimes n}\right) \tag{6}$$

where $G_N \in R^{N \times N}$, B_N is the bit-reversal permutation matrix, and $F^{\otimes n}$ represents the n^{th} Kronecker power of the kernel matrix *F*.

C. Decoding Process

At the receiver side, the legitimate user uses K_{secret} to generate the corresponding subkey sequence K_i by AES-CTR mode. The decoding process employs SC/Successive Cancellation List (SCL) decoding. Fig. 3 depicts the block diagram of the decoding procedure. Meanwhile, the decoder removes the frozen-bit subkey K^i_{frozen} to recover the frozen bits. After decoding the vector u, the original audio data is recovered as:

$$\hat{D}_i = \hat{u}_A \oplus K^i_{info} \tag{7}$$

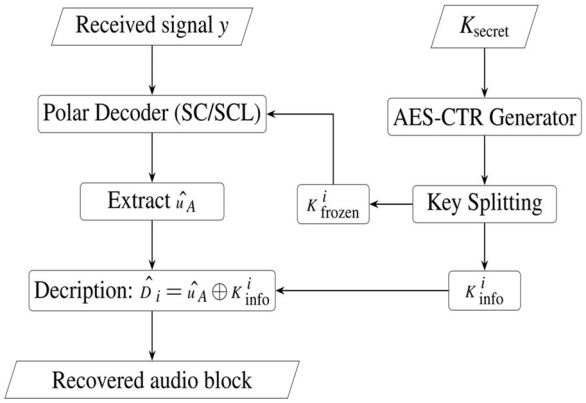


Fig. 3. Diagram of audio decoding processing process.

The use of AES-CTR and XOR operations with frozen bits does not significantly increase the inherent computational complexity of the system. The total computational cost of the polar coding layer remains bounded by $O(N \log N)$. In terms of system performance metrics, this low complexity translates directly to minimal processing latency, making the framework highly suitable for real-time audio streaming. Furthermore, the stream-cipher nature of AES-CTR requires minimal state retention, ensuring a low memory footprint compatible with resource-constrained wireless nodes. Regarding scalability, the proposed architecture can flexibly adapt to varying payload sizes and channel conditions by simply adjusting the polar codeword length *N* and the corresponding AES keystream generation, without requiring structural changes to the core crypto-coding mechanism.

To rigorously justify the security enhancement of this approach beyond empirical observations, we provide a formal information-theoretic analysis of the SC decoder's error propagation. In polar coding theory, the reliable recovery of the information bits u_A strictly depends on the perfect knowledge of the frozen bits u_{A^c}. Let Y be the received signal vector. Under SC decoding, the estimation of the i^{th} bit, $\hat{u}_i$, relies on the conditional likelihood ratio $L_i(y, \hat{u}_0^{i-1})$. In the proposed scheme, the frozen bits are dynamically randomized by the AES-CTR sub-key K^i_{frozen}. For an unauthorized eavesdropper lacking the secret key, K^i_{frozen} is computationally indistinguishable from a perfectly uniform random sequence. Consequently, the eavesdropper's SC decoder is forced to guess the frozen bits. Because SC decoding is sequential, an incorrect guess at index $j \in A^c$ completely distorts the

transition probabilities for all subsequent bits $i > j$. Formally, the mutual information between the transmitted information bits u_A and the eavesdropper's observations collapse $I(u_A; Y, \hat{u}_{A^c}) \to 0$. As the mutual information approaches zero, the equivocation (conditional entropy) of the eavesdropper approaches the unconditional entropy of the source $H(u_A|Y, \hat{u}_{A^c}) \to H(u_A)$. This mathematical bound demonstrates that the polar coding layer provides a formal PLS mechanism. It guarantees that the eavesdropper gains virtually no structural information about the plaintext from the physical channel, thereby acting as a robust theoretical shield that complements the computational Indistinguishability under Chosen Plaintext Attack (IND-CPA) security of the AES-CTR layer.

Furthermore, this structural security guarantee extends comprehensively beyond the standard SC decoder to advanced decoding algorithms, such as Successive Cancellation List (SCL) and Belief Propagation (BP). An adversary attempting to exploit an SCL decoder inherently relies on known frozen bit values to prune incorrect paths and manage list size. When the frozen bits are dynamically randomized by the AES-CTR engine and remain unknown to the eavesdropper, the SCL decoder is forced to branch at each frozen bit index. Because the number of possible frozen bit combinations (2^{N-K}) vastly exceeds any practical list size of L, the correct decoding path is inevitably purged from the list, leading to catastrophic failure. Similarly, a BP decoder requires deterministic Log-Likelihood Ratios (LLRs), which are typically initialized to $\pm\infty$ at the frozen bit indices to propagate reliability across the factor graph and achieve convergence. Without the secret key, an eavesdropper must initialize these frozen bit LLRs to zero, representing complete uncertainty. This deprives the factor graph of its essential structural constraints, strictly preventing the BP algorithm from converging to the correct codeword. Consequently, even when an adversary is fully aware of the scheme and deploys advanced SCL or BP decoders, the structural obfuscation of the frozen bits ensures the unauthorized receiver still yields a completely degraded signal with a BER of about 0.5.

III. Simulation Results and Discussion

This section evaluates the proposed scheme via MATLAB simulations. The dataset comprises five uncompressed 16-bit, 44.1 kHz '.wav' audio files (5–8 s, mono/stereo) sourced from royalty-free databases like Freesound. To validate robustness across diverse acoustic structures, we select one representative sample per category: continuous harmonics (Piano), percussion (Drum), vocal formants (Human), and ambient noise (Animal, Car). While this compact dataset sufficiently validates the core structural security, future exhaustive statistical evaluations will utilize large-scale benchmarks such as ESC-50.

A. Visual Analysis

To assess the effectiveness of structure destruction in the original signal, we conduct a visual evaluation from three perspectives: the time domain (Waveform), the frequency domain (Spectrogram), and the amplitude distribution (Histogram) for both mono and stereo audio signals.

Fig. 4 depicts how the original audio waveform changes as a result of the encryption process. It can be observed that the original signal exhibits clear amplitude variations depending on the information content. In contrast, the encrypted signal shows a noise-like distribution within the range $[-1, 1]$, completely concealing the amplitude characteristics of the original signal.

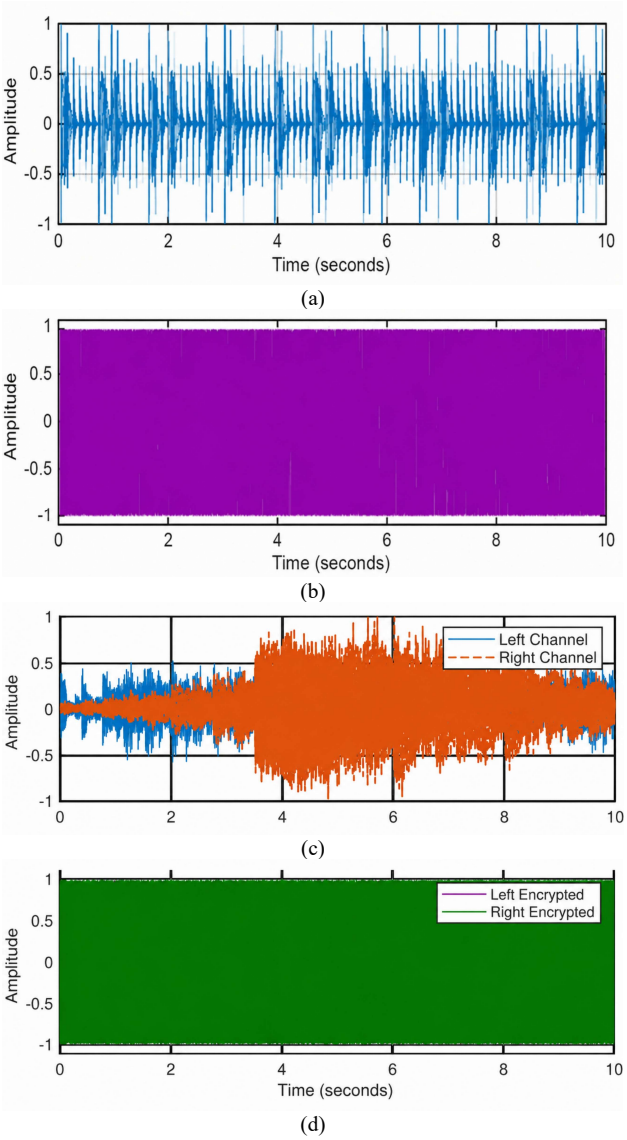


Fig. 4. Waveform analysis of both original and encrypted signals: (a) Original mono signal, (b) Encrypted mono signal, (c) Original dual signal, (d) Encrypted dual signal.

Fig. 5 compares the energy spectra of mono and dual-channel audio signals before and after encoding. The original audio signals exhibit energy concentration mainly in the low- and mid-frequency bands, with distinct spectral structures that reflect their underlying information content. After encoding, the spectral energy is uniformly distributed across the entire bandwidth from 0 to 22 kHz. This spectral flattening effect transforms the encoded

signals into noise-like patterns and effectively suppresses frequency domain analysis attacks.

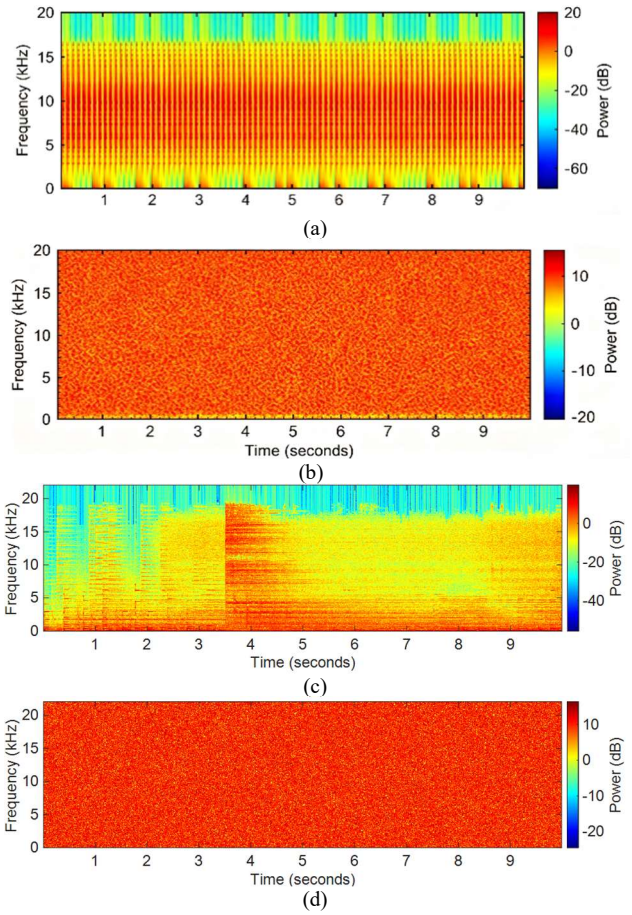


Fig. 5. Spectrogram analysis of both original and encrypted signals: (a) Original mono spectrogram, (b) Encrypted mono spectrogram, (c) Original dual spectrogram, (d) Encrypted dual spectrogram.

Fig. 6 illustrates the spectral representations of different audio samples (animal, car, drum, human voice, and piano) before and after encryption, where the horizontal axis denotes time and the vertical axis indicates frequency, and the signal energy is represented by color intensity. The original audio signals exhibit distinct energy concentrations at specific frequency bands, reflecting their inherent acoustic characteristics, as indicated by the non-uniform color distributions. By contrast, the encrypted audio spectra exhibit an approximately even energy spread across the full frequency band from 0 to 22 kHz. This spectral flattening effect demonstrates that the proposed method effectively disperses the signal energy, obscures the original spectral features, and thereby enhances resistance against frequency-domain analysis attacks.

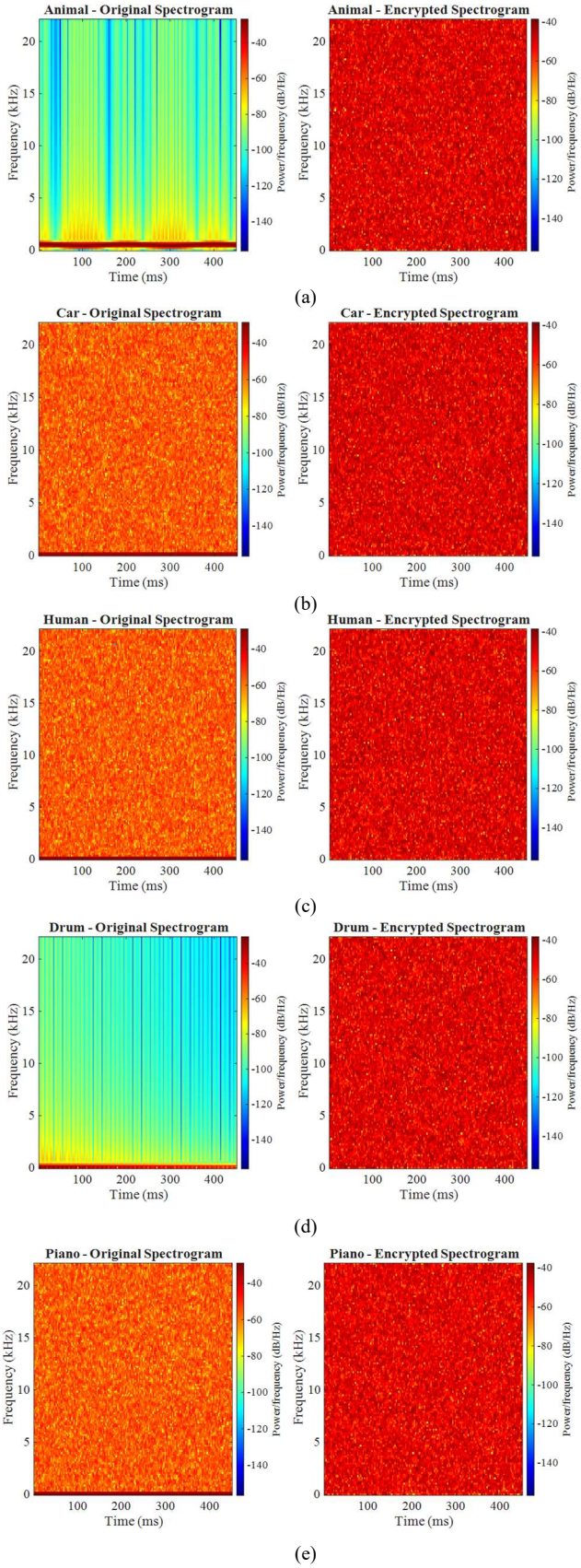


Fig. 6. Spectrogram representations of the original and encrypted audio signals: (a) Audio Animal, (b) Audio Car, (c) Audio Human, (d) Audio Drum, (e) Audio Piano.

Fig. 7 shows the probability distribution of the signal amplitude. The histogram of the original signal follows a Gaussian distribution (centered around 0). After applying the proposed algorithm, the probability distribution becomes a uniform distribution. This outcome confirms the algorithm's strong resilience against statistical attacks.

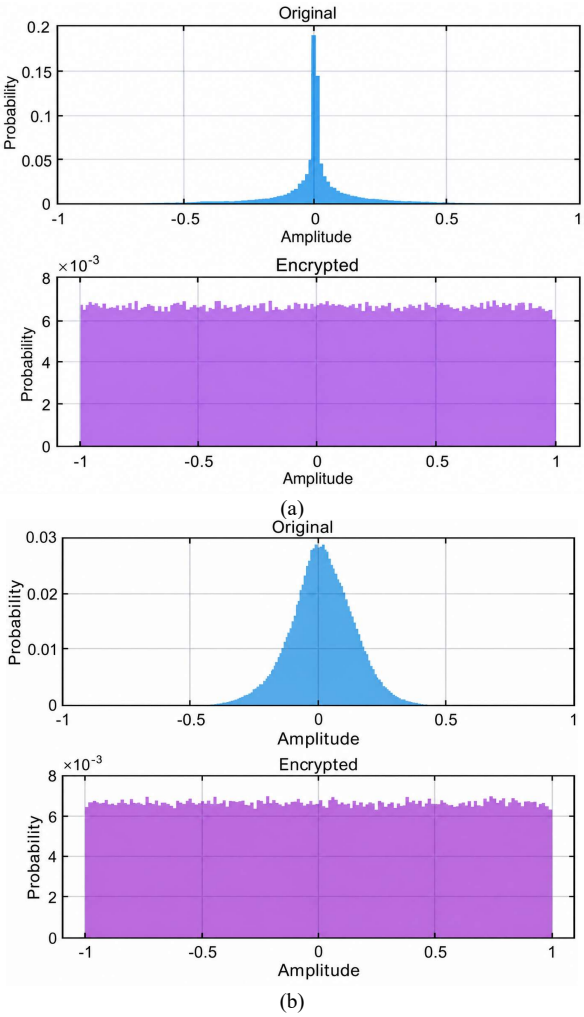
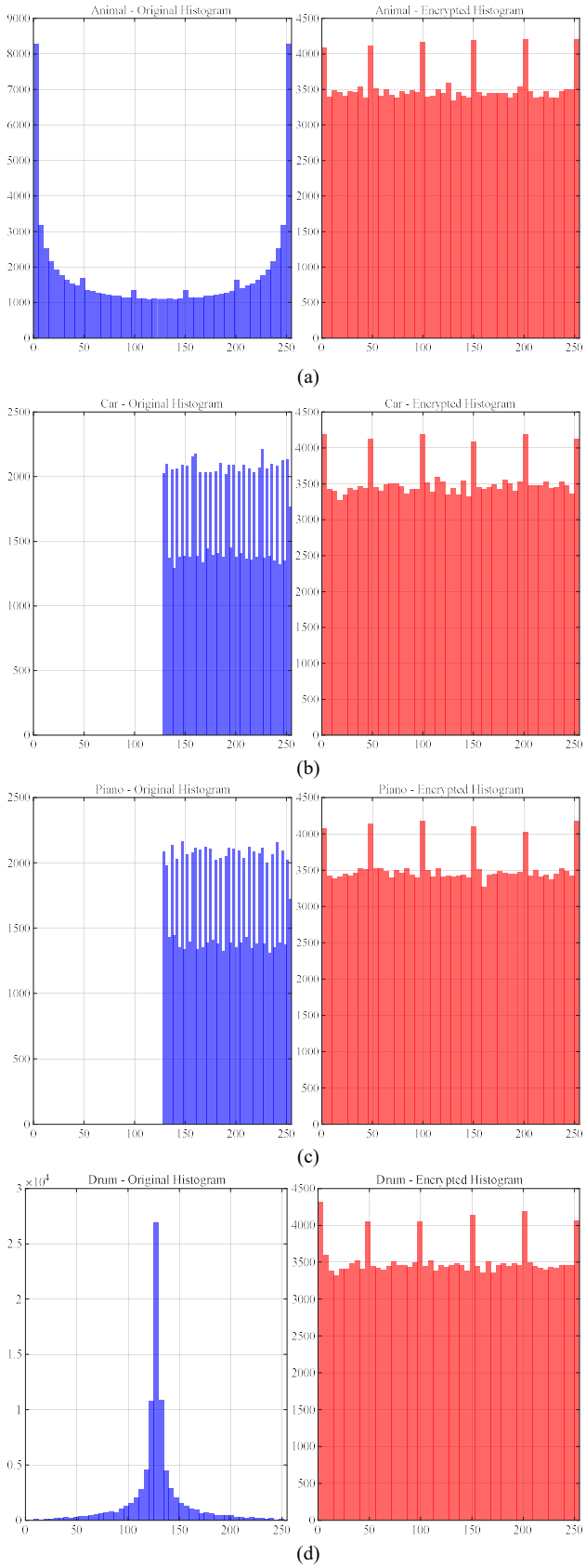


Fig. 7. The probability distribution of the signal amplitude: Mono: (a) Original and encrypted histograms, (b) Dual: Original and encrypted histograms.

Fig. 8 presents the frequency histograms, which are used to analyze and evaluate changes in the amplitude probability distribution of the given audio signals with respect to statistical attack resistance. As shown in the figure, the original audio samples (Animal, Car, Piano, Drum, and Human) exhibit distinct non-uniform histogram distributions, typically showing a strong concentration around zero or following a Gaussian (bell-shaped) distribution. Such original data can be easily exploited by attackers.

After applying the proposed encryption method, the frequency histogram of each audio sample becomes flat and closely approaches a uniform distribution. The amplitude values are evenly distributed across the entire dynamic range, with no noticeable concentration or regular pattern. These results demonstrate that the proposed encryption algorithm effectively destroys the statistical information of the plaintext audio, providing strong resistance against statistical analysis attacks, as no exploitable amplitude patterns can be identified in the encrypted signals.



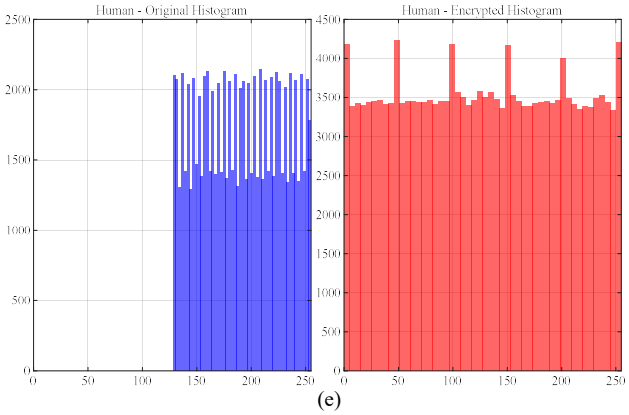


Fig. 8. Histogram before and after audio encryption: (a) Audio animal, (b) Audio car, (c) Audio piano, (d) Audio drum, (e) Audio human.

B. Quantitative Analysis

To confirm safety, we use the following indicators: Correlation Coefficient, Entropy, Number of Samples Change Rate (NSCR)/Unified Averaged Changed Intensity (UACI), and Peak Signal-to-Noise Ratio (PSNR)

Correlation Coefficient quantifies the strength of linear dependence between neighboring samples. The calculation is performed as described in Ref. [8]:

$$r_{xy} = \frac{cov(x,y)}{\sqrt{D(x)}\sqrt{D(y)}} \tag{8}$$

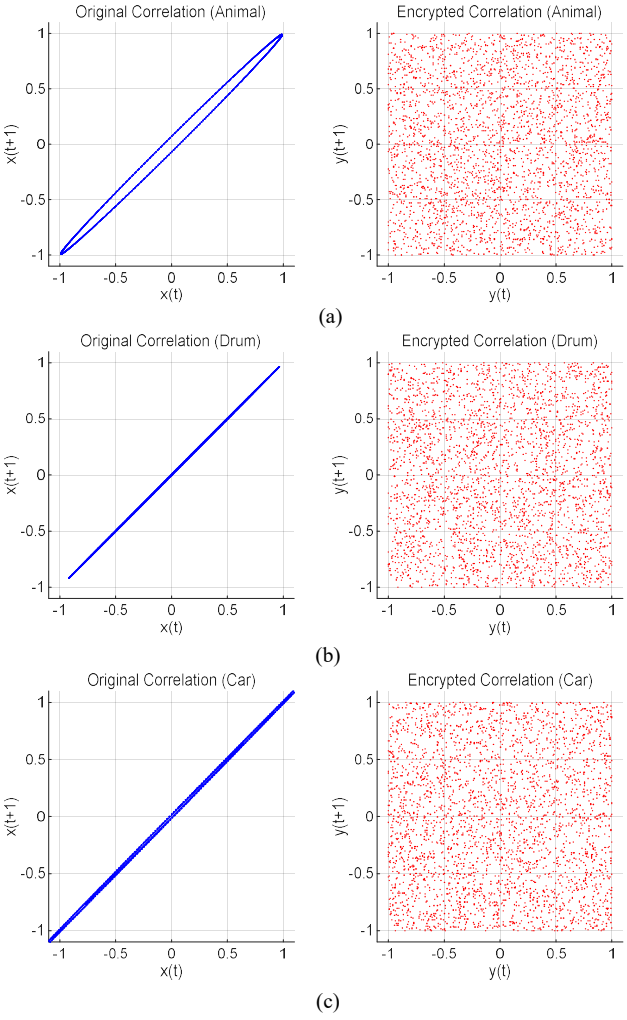
where $D(x)$ and $cov(x,y)$ denote the variance and the covariance, respectively.

The correlation coefficients of the original samples are 0.9921 (Animal), 0.9966 (Drum), 0.9695 (Piano), 0.9176 (Car), and 0.9815 (Human). These results prove that the original audio signals exhibit extremely high correlation and confirm strong statistical dependence. The results reported in Table II show the correlation coefficients of the encrypted samples. It can be seen that the encrypted signals yield correlation values close to zero, indicating that the statistical relationship between the signal samples has been effectively eliminated. As detailed in Fig. 9 presents the distribution of correlation between adjacent samples across different audio signals, comparing the cases before encoding and after encoding. For the original audio signals, a strong linear correlation can be observed, as the data points are densely concentrated along the diagonal, indicating a high dependence between neighboring samples and revealing the inherent signal structure. In contrast, the encoded audio signals exhibit high dispersion and randomness, with no clear correlation pattern. The uniform distribution of points across the entire plane suggests that the proposed encoding method effectively eliminates statistical correlation between adjacent samples. This significant reduction in correlation demonstrates the effectiveness of the proposed method to resist correlation-based and statistical assaults.

As detailed in Table II, the correlation coefficients of the encrypted signals are −0.00157 (Animal), 0.00154 (Drum), −0.00056 (Piano), 0.001171 (Car), and 0.00009 (Human), respectively. These results confirm that the encrypted samples are statistically independent, ensuring effective decorrelation. The effectiveness of the proposed scheme is assessed by benchmarking its results against those of existing encryption methods [11, 17, 28]. As shown in the table, our method achieves correlation values comparable to or lower than these benchmarks. For instance, the result for the 'Animal' sample (−0.00157) in our scheme shows a competitive decorrelation performance compared to the value reported in Method [17] (0.0014). Furthermore, our results align with the low correlation ranges reported in Method [11] (0.0009–0.0013) and Method [28] (0.0011), validating the proposed method's robustness against statistical attacks.

TABLE II. CORRELATION COEFFICIENTS OF ENCRYPTED SIGNALS

Audio Samples	Ours	Ref. [11]	Ref. [17]	Ref. [28]
Animal	−0.00157	–	0.0014	–
Drum	0.00154	–	–	–
Piano	−0.00056	–	0.0004	–
Car	0.001171	–	–	–
Human	0.00009	–	−0.0005	–
Other samples	–	0.0009 to 0.0013	–	0.0011



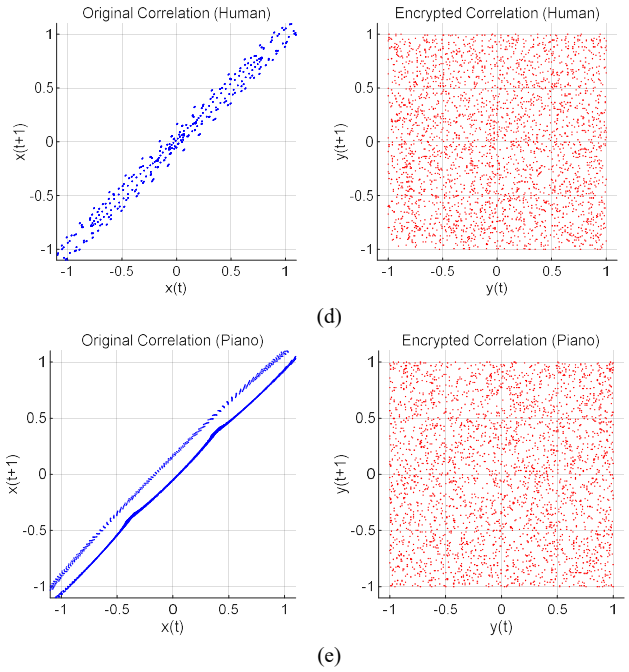


Fig. 9. Correlation before and after audio encryption: (a) Animal, (b) Drum, (c) Car, (d) Human, (e) Piano.

Fig. 9 presents the distribution of correlation between adjacent samples across different audio signals, comparing the cases before encoding and after encoding. For the original audio signals, a strong linear correlation can be observed, as the data points are densely concentrated along the diagonal, indicating a high dependence between neighboring samples and revealing the inherent signal structure. In contrast, the encoded audio signals exhibit high dispersion and randomness, with no clear correlation pattern. The uniform distribution of points across the entire plane suggests that the proposed encoding method effectively eliminates statistical correlation between adjacent samples. This significant reduction in correlation demonstrates the effectiveness of the proposed method to resist correlation-based and statistical assaults.

Information Entropy indicates the level of uncertainty and randomness within a signal. The entropy H is defined as:

$$H(s) = -\sum_{i=0}^{2^N-1} P(s_i) \log_2 P(s_i) \tag{9}$$

The ideal entropy for an 8-bit encoded signal is 8 bits/sample [2], representing a truly random source where all 256 possible amplitude values occur with equal probability. Table III displays that the entropy of the samples after encoding all reached approximately 7.9989; the proposed scheme achieved entropy values of 7.9989 (Animal), 7.9993 (Drum), 7.9992 (Piano), 7.9983 (Car), and 7.9993 (Human), respectively. These results (7.9983–7.9993) indicate a high degree of randomness in the ciphertexts very close to the theoretical benchmark, confirming the high degree of randomness achieved by the proposed scheme. Therefore, the proposed method is extremely effective in resisting statistical analysis attacks.

By comparison, Ref. [13] reports entropy values ranging from 7.9371 to 7.9473, while Ref. [12] reports 7.9966 to 7.9995. These results depict that the proposed framework outperforms Ref. [13] and matches Ref. [12] near the ideal value of 8.00. Therefore, the proposed method outperforms previous approaches in terms of resistance to entropy attacks.

To assess the sensitivity of the algorithm (avalanche effect), we use two metrics, including NSCR and UACI. The calculation formulas are as follows Eqs. (10)–(11):

$$NSCR = \frac{\sum_{i}^{L} D(i)}{L} \times 100\% \tag{10}$$

$$UACI = \frac{1}{L} \sum_{i}^{L} \frac{|c_1(i) - c_2(i)|}{2^N - 1} \times 100\% \tag{11}$$

where c_1 and c_2 denote two ciphertexts generated from two slightly different plaintexts, L represents the data length. As reported in Table IV and Table V, the obtained NSCR (≈ 99.61%) and UACI (≈ 33.4%) values are in close agreement with the corresponding theoretical results, which confirms the suggested approach's strong defense against differential attacks.

In the case of 8-bit audio signals, the theoretically ideal NSCR value is 99.61%. Simulation results demonstrate that the proposed method exhibits near-perfect convergence to this ideal value across all test samples. Looking at Table IV, the NSCR indices for the proposed method are 99.62% for animal sounds, 99.62% for drum sounds, 99.62% for piano sounds, 99.62% for car sounds, and 99.63% for human sounds, respectively.

TABLE III. INFORMATION ENTROPY ANALYSIS RESULTS

Audio Samples	Ours	Ref. [12]	Ref. [13]
Animal	7.9989	–	–
Drum	7.9993	–	–
Piano	7.9992	–	–
Car	7.9983	–	–
Human	7.9993	–	–
All samples (range) Theory	–	7.9966 to 7.9995	7.9371 to 7.9473

Note: The theoretical maximum entropy for 8-bit audio is 8.00 bits/sample.

TABLE IV. NSCR EVALUATION RESULTS (%)

Audio Samples	Ours	Ref. [6]	Ref. [13]	Ref. [15]
Animal	99.62	–	–	–
Drum	99.62	–	–	–
Piano	99.62	–	–	–
Car	99.62	–	–	–
Human	99.63	–	–	–
All samples (range) Theory		99.60354 to 99.61265	99.5316 to 99.6201	98.43 ≈ 99.61

To further validate the security performance of our scheme, we compare its NSCR results with those reported in recent relevant studies [6, 13, 15]. As indicated in the literature, the NSCR values in Ref. [13] range from 99.5316% to 99.6201%, and Ref. [6] yields NSCR results between 99.60354% and 99.61265%. By contrast, the approach in Ref. [15] only achieves an NSCR of 98.43%. This shows that the proposed method consistently maintains a more stable NSCR value, closer to the theoretical optimal level of 99.61%. This confirms the clear advantages of the proposed algorithm over previous methods. For all audio samples in this study, the NSCR results are around the optimal value, showing that the suggested approach successfully reduces the statistical dependence between the raw and encrypted signals while keeping the encoding performance insensitive to the characteristics of the input audio. As a result, the proposed method provides comprehensive protection against statistical analysis attacks, surpassing conventional methods presented in Refs. [6, 15], which do not achieve comparable durability.

Table V reports the UACI results. The theoretical UACI value for random 8-bit data is approximately 33.46%. As shown in Table V, our scheme generates ciphertexts with differential intensities extremely close to this ideal value. The UACI values obtained for different audio categories are 33.41% for animal sounds, 33.55% for drum sounds, 33.52% for piano sounds, 33.62% for car sounds, and 33.47% for human sounds, respectively. These results are noticeably closer to the theoretical ideal than those of Ref. [7] (32.78%–32.86%), Ref. [16] (33.1%–33.4%), and Ref. [17] (33.3439%–33.3782%). This confirms that the proposed encryption scheme yields a more uniform distribution of differences, thereby offering superior resistance against differential analysis.

TABLE V. UACI EVALUATION RESULTS (IN %)

Audio Samples	Ours	Ref. [7]	Ref. [16]	Ref. [17]
Animal	33.41	–	–	33.3439
Drum	33.55	–	–	33.3547
Piano	33.52	–	–	33.3551
Car	33.62	–	–	33.3782
Human	33.47	–	v	33.3630
All samples (range)	–	32.7895 to 32.8661	33.1 to 33.4	–
Theory		≈ 33.46		

TABLE VI. PSNR RESULTS IN DB FOR ENCRYPTED AUDIO SIGNALS

Audio Sample	Ours	Ref. [7]	Ref. [16]	Ref. [17]
Animal	−29.2413	–	–	−10.1874
Drum	−29.4593	–	–	−12.1700
Piano	−26.2086	–	–	−14.4715
Car	−25.7494	–	–	−8.5007
Human	−25.8253	–	–	−12.1941
All samples (range)	–	4.4186 to 4.8659	−10.8877	–

PSNR is used to evaluate the level of distortion between the original and the corresponding encrypted audio sample. A lower PSNR value indicates a higher degree of dissimilarity and, consequently, better encryption performance [10]. The PSNR is calculated as follows in Eq. (12):

$$PSNR = 10 \times \log_{10}\left(\frac{MAX^2}{MSE}\right) \tag{12}$$

The PSNR values reported in Table VI are consistently low across all tested samples. The PSNR results obtained for different audio categories are −29.4593 dB for drum sounds, −26.2086 dB for piano sounds, −25.7494 dB for car sounds, and −25.8253 dB for human sounds. These negative values demonstrate that there is a substantial difference between the encrypted and raw audio samples and exhibit strong noise-like characteristics. In comparison with previous studies, the proposed method demonstrates superior encryption performance. While one prior method reports PSNR values from −8.50 dB to −14.47 dB, while another study reports positive values between 4.41 dB and 4.86 dB. Furthermore, the proposed result is also considerably lower than the −10.8877 dB reported in the literature. This substantial reduction in PSNR confirms that the proposed scheme introduces a much higher degree of distortion to the plain audio,

thereby ensuring better security against statistical attacks [7, 16, 17].

The blank cells in Tables II–VI indicate that the referenced studies used different, often proprietary, audio samples, making direct one-to-one comparison impractical. Because reproducing the exact physical-layer hardware and diverse coding architectures of all compared methods to test a single common audio file is impractical, a strict one-to-one numerical comparison on identical acoustic data cannot be established. Consequently, these tables are not intended to serve as an absolute benchmark. Rather, they are provided to demonstrate that the proposed scheme consistently achieves competitive or superior encryption metrics (such as NSCR, UACI, and Information Entropy) within the general operational ranges reported in current literature. Establishing a standardized, open-access audio benchmark dataset specifically tailored for evaluating physical-layer security and crypto-coding schemes remains a critical area for future research.

C. Chosen/Known Plaintext Analysis

Known Plaintext Attacks (KPA) and Chosen-Plaintext Attacks (CPA) typically rely on discovering linear relationships or deterministic patterns between the plaintext and the ciphertext. While modern cryptographic primitives require formal provable security models (e.g., IND-CPA), the theoretical cryptographic security of our scheme is fundamentally inherited from the standardized AES algorithm in CTR mode. Owing to the nonlinear nature of AES and the use of a varying initialization vector, deriving the original secret key from plaintext–ciphertext pairs is computationally intractable.

However, to evaluate the structural robustness of the integrated physical-layer scheme, we augment this theoretical baseline with an empirical evaluation. In our hybrid system, the polar coding layer restricts the adversary's ability to accurately intercept the ciphertext. We carry out simulations of two hypothetical attack scenarios to validate the empirical resistance of the proposed approach. It should be explicitly noted that while these simulation-based Bit Error Rate (BER) metrics and statistical tests (Entropy, NSCR, UACI) confirm strong signal randomization and physical-layer obfuscation, they serve as empirical indicators rather than formal cryptographic proofs. However, from an information-theoretic perspective, the physical-layer security can be bounded by the equivocation (conditional entropy). Because the dynamic randomization of frozen bits physically degrades the intercepted signal, the mutual information between the original audio source S and the eavesdropper's intercepted signal Z is minimized. Consequently, the equivocation lower bound approaches the source entropy, $H(S|Z) \approx H(S)$, mathematically demonstrating that the eavesdropper gains virtually no structural information about the plaintext from the physical layer. Establishing a rigorous formal framework for the joint crypto-coding mutual information leakage remains a subject for future theoretical modeling.

KPA scenario: It is assumed that the attacker knows the plaintext of the Animal audio file and extracts the corresponding keystream (Key A). The attacker then reuses this keystream to decrypt another file, Drum (encrypted with Key B). The results show that the two keystreams are completely different.

The BER metric is used to evaluate the difference between the extract keystream and the actual keystream, calculated as:

$$BER = \frac{1}{L}\sum_{j=1}^{L}\left|K_{extracted}^{j} - K_{real}^{j}\right| \times 100\% \qquad (13)$$

where L is the keystream length.

The decryption error map in Fig. 10 indicates a BER of 49.56%, meaning the decrypted output is essentially random noise.

The attacker selects a special input, namely an all-zero sequence (''Silence''), feeds it into the system, and obtains the corresponding keystream (Key C). This keystream is then used to decrypt the ''Drum'' file. The results in Fig. 11 show a BER of 50.08%. BER values close to 50% in both scenarios confirm that the system achieves strong confusion and diffusion properties. Even a minor change in the plaintext or the key-generation parameters (IV/Counter) leads to a completely different keystream, effectively preventing any reuse of information from KPA/CPA attacks.

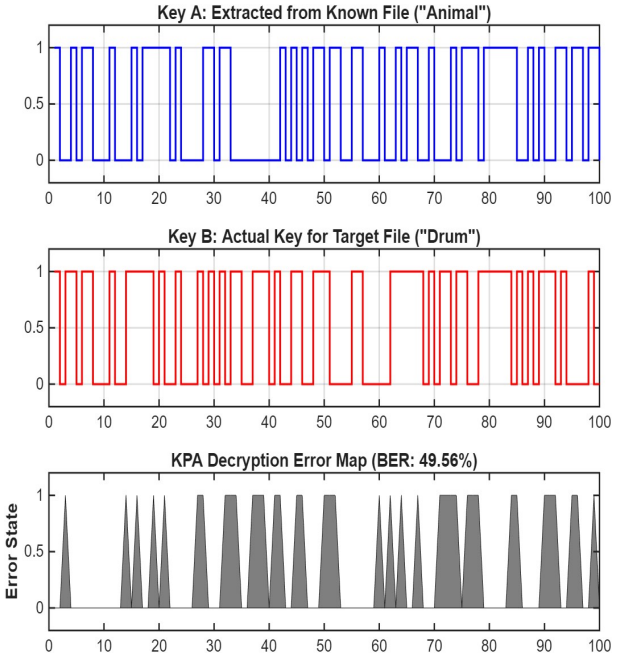


Fig. 10. Resistance to Known-Plaintext Attack (KPA) scenario.

The cryptanalysis in Ref. [29] demonstrates that the secret key can be fully recovered, which is equivalent to the BER dropping to 0% (a total break) and exposing the original audio. In contrast, our proposed scheme maintains a BER of 50.08% under similar CPA conditions (Fig. 11), proving that the attacker recovers nothing but random noise. These results confirm that the suggested approach successfully addresses the structural flaws of chaotic

key-chaining, offering enhanced cryptographic robustness to thwart chosen-plaintext analysis.

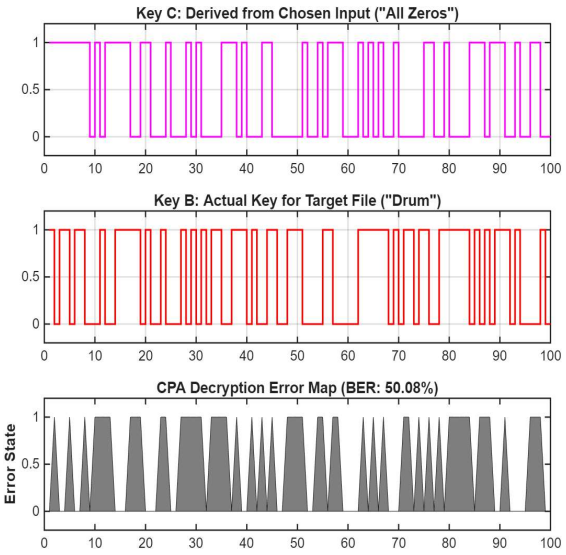


Fig. 11. Resistance to Chosen-Plaintext Attack (CPA) scenario.

D. Reliability and Channel Performance Analysis

While the previous subsections validate the cryptographic strength of the proposed scheme, it is fundamental for 5G physical-layer applications to evaluate communication reliability over a noisy channel. We simulate the BER versus E_b / N_0 over an Additive White Gaussian Noise (AWGN) channel using BPSK modulation with a code length of 1024 and a code rate of 0.5.

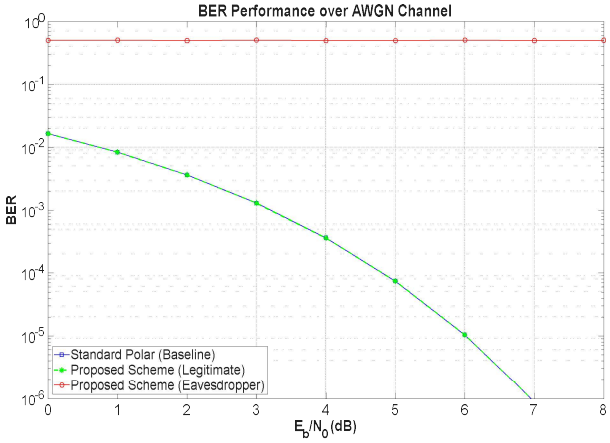


Fig. 12. BER performance comparison of the proposed crypto-coding scheme over an AWGN channel.

Fig. 12 illustrates the BER performance of the proposed crypto-coding scheme compared to a baseline standard polar code. The results clearly demonstrate that for the legitimate receiver possessing the correct secret key, the proposed scheme exhibits a BER curve that perfectly matches the standard baseline polar code. Because the legitimate receiver accurately generates the dynamic AES-CTR keystream to identify the frozen bits, the SC decoder operates with full error-correction capability. This proves that the proposed integration achieves robust security without sacrificing communication reliability, incurring zero performance penalty compared to standard transmission. Conversely, for an unauthorized eavesdropper lacking the exact dynamic sub-key, the SC decoder structurally fails, keeping the BER uniformly bounded at approximately 0.5 across all E_b/N_0 regimes. This confirms that eavesdroppers recover nothing but noise, regardless of the physical channel's clarity.

E. Impact of SCL List Size on Performance and Security

While the baseline decoding assumes a standard SC approach, practical 5G implementations often utilize the SCL decoder. For the legitimate receiver possessing the explicit AES-CTR sub-keys, increasing the list size L (e.g., $L \in \{2, 4, 8\}$) yields standard coding gain improvements, directly enhancing the BER versus E_b/N_0 performance exactly as it does in standard polar codes. Conversely, from a cryptanalysis perspective, an adversary might attempt to exploit an SCL decoder with a large L to empirically track and guess the randomized frozen bits. However, this approach is mathematically futile. Lacking the keystream, the eavesdropper is forced to treat the dynamic frozen bits as unknown variables, causing the SCL algorithm to branch at each frozen index. Because the total number of permutations (2^{N-K}) vastly exceeds any physically realizable memory constraint for L, the correct decoding path is rapidly and inevitably purged from the surviving list. Consequently, employing advanced SCL decoding provides zero computational advantage to the eavesdropper, guaranteeing that the signal remains completely degraded (BER ≈ 0.5) regardless of the chosen list size.

F. Computational Complexity and Execution Time

The theoretical computational complexity of the proposed scheme is dominated by the polar encoding and decoding operations, tightly bounded by $O(N \log N)$, as the AES-CTR keystream generation adds strictly linear $O(N)$ overhead. To empirically substantiate its suitability for real-time streaming, execution time measurements are conducted in the software simulation environment. The average total processing time (including sub-key extraction, AES-CTR generation, and polar encoding) per $N = 1024$ block was measured at approximately 0.1489 ms. This yields a baseline software throughput of approximately 3438.30 kbps (or ~3.44 Mbps). Given that standard uncompressed audio streaming typically requires bitrates around 1.4 Mbps, the unoptimized software implementation already comfortably meets and exceeds practical real-time constraints. Future implementations on dedicated hardware testbeds (e.g., using embedded platforms like Raspberry Pi or FPGAs) will fully exploit the parallelization potential of the AES engine and polar encoder, achieving substantially higher hardware throughput for real-time 5G physical-layer applications.

G. Practical Synchronization Considerations and Recovery Mechanism

It is necessary to note that this integrated process requires tight synchronization between the polar channel decoder and the AES-CTR decryptor. Because the AES

keystream is highly sensitive to counter offset, a loss of frame sync could potentially cause catastrophic error propagation (BER ≈ 0.5) for the legitimate receiver. To ensure the scheme is robust and practical over realistic 5G fading or Doppler channels, a synchronization recovery mechanism must be integrated. Similar to standard cellular protocols, this is achieved by appending a lightweight, unencrypted Sequence Number (SN) to the header of each transmitted polar frame. This explicit SN allows the receiver to derive the exact AES-CTR counter required for that specific block. Consequently, if a frame is entirely dropped due to severe channel impairments, the system does not permanently desynchronize. The receiver simply reads the SN of the next successfully detected frame and resynchronizes the counter instantly. Under this recovery mechanism, the cryptographic frame loss is strictly bounded by the physical layer's Block Error Rate (BLER), preventing cascading failures and guaranteeing practical resilience in real-world deployments.

IV. Conclusion

This research presents an audio encryption technique related to the crypto-coding method based on polar codes to ensure the reliability and security of data transmission in secure wireless communication systems. The proposed system employs the AES-CTR key generation function to generate dynamic sub-keys from the wireless channel for each data block, thereby overcoming the drawbacks of fixed-key usage and enhancing security at the physical layer. MATLAB simulation results prove the superior performance of the suggested approach, with information entropy close to the ideal value, a correlation coefficient close to zero, sensitivity metrics meeting security standards (NSCR $\approx$ 99.61%, UACI $\approx$ 33.48%), and a low PSNR value. These findings demonstrate that the suggested method is highly effective for structural audio encryption and offers robust defense against passive eavesdropping.

While these foundational simulation results establish a strong theoretical proof-of-concept for future 5G and real-time multimedia security applications, immediate practical readiness requires further investigation. Consequently, future work will explicitly address the synchronization resilience of the AES-CTR keystream over fading channels and transition from software-only evaluations to physical hardware prototypes (e.g., embedded transceivers) to validate real-time feasibility. Additionally, to bridge the gap between empirical validation and information-theoretic security, formal proofs regarding secrecy capacity, mutual information leakage, and equivocation will be rigorously developed, alongside evaluations against advanced decoding techniques, side-channel attacks, and active adversarial models.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Tran Duc Thiep and Pham Van Tiep did the simulation; Pham Duy Phong performed the data analysis and structured the manuscript; Pham Manh Tuan wrote and revised the paper; Dinh Van Linh verified the simulation outcomes and theoretical findings; all authors had approved the final version.

References

[1] J. He, P. Zhu, Z. Liu, and Y. Cao, "A novel digital audio Encryption and forensics watermarking scheme," *IEEE Access*, vol. 12, pp. 103565–103582, 2024. doi: 10.1109/ACCESS.2024.3434576

[2] A. Maity and B. C. Dhara, "An audio encryption scheme based on empirical mode decomposition and 2D cosine logistic map," *IEEE Lat. Am. Trans.*, vol. 22, no. 4, pp. 267–275, 2024. doi: 10.1109/TLA.2024.10472959

[3] G. Alzeer and H. Alhumyani, "Efficient audio encryption using a discrete cosine transform and baker map," in *Proc. 2022 Fifth Natl. Conf. Saudi Comput. Coll.*, 2022, pp. 201–208. doi: 10.1109/NCCC57165.2022.10067440

[4] J. B. Lima and E. F. D. S. Neto, "Audio encryption based on the cosine number transform," *Multimed. Tools Appl.*, vol. 75, no. 14, pp. 8403–8418, 2016. doi: 10.1007/s11042-015-2755-6

[5] H. Djamel, A. Herbadji, I. Haddad, H. Kahia, A. Belmeguenai, and N. Derouiche, "An enhanced logistic chaotic map based tweakable speech encryption algorithm," *Integration*, vol. 97, 102192, 2024. doi: 10.1016/j.vlsi.2024.102192

[6] M. Y. M. Parvees, J. A. Samath, and B. P. Bose, "Audio encryption −A chaos-based data byte scrambling technique," *Int. J. Appl. Syst. Stud.*, vol. 8, no. 1, pp. 51–75, 2018. doi: 10.1504/IJASS.2018.091847

[7] W. Dai, X. Xu, X. Song, and G. Li, "Audio encryption algorithm based on chen memristor chaotic system," *Symmetry (Basel)*, vol. 14, no. 1, pp. 1–19, 2022. doi: 10.3390/sym14010017

[8] I. A. Taqi and S. M. Hameed, "A secure audio encryption method using tent-controlled permutation and logistic map-based key generation," *Comput. Mater. Contin.*, vol. 85, no. 1, pp. 1653–1674, 2025. doi: 10.32604/cmc.2025.067524

[9] T. Haridas, S. D. Upasana, G. Vyshnavi, M. S. Krishnan, and S. S. Muni, "Chaos-based audio encryption: Efficacy of 2D and 3D hyperchaotic systems," *Franklin Open*, vol. 8, 100158, 2024. doi: 10.1016/j.fraope.2024.100158

[10] S. Adhikari and S. Karforma, "A novel audio encryption method using Henon–Tent chaotic pseudo random number sequence," *Int. J. Inf. Technol.*, vol. 13, no. 4, pp. 1463–1471, 2021. doi: 10.1007/s41870-021-00714-x

[11] F. J. Farsana, V. R. Devi, and K. Gopakumar, "An audio encryption scheme based on fast walsh hadamard transform and mixed chaotic keystreams," *Appl. Comput. Informatics*, vol. 19, no. 3–4, pp. 239–264, 2023. doi: 10.1016/j.aci.2019.10.001

[12] D. Shah, T. Shah, and S. S. Jamal, "Digital audio signals encryption by Mobius transformation and Hénon map," *Multimed. Syst.*, vol. 26, no. 2, pp. 235–245, 2020. doi: 10.1007/s00530-019-00640-w

[13] 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 s8symmetric group and chaotic systems," *Math. Probl. Eng.*, vol. 1, 5554707, 2021. doi: 10.1155/2021/5554707

[14] E. A. Albahrani, T. K. Alshekly, and S. H. Lafta, "A review on audio encryption algorithms using chaos maps-based techniques," *J. Cyber Secur. Mobil.*, vol. 11, no. 1, pp. 53–82, 2022. doi: 10.13052/jcsm2245-1439.1113

[15] M. Mujtaba and M. Singh, "Optimizing audio encryption efficiency: A novel framework using double DNA operations and chaotic map-based techniques," *Comput. Electr. Eng.*, vol. 123, 110088, 2025. doi: 10.1016/j.compeleceng.2025.110088

[16] A. Kumar and M. Dua, "Audio encryption using two chaotic map based dynamic diffusion and double DNA encoding," *Appl. Acoust.*, vol. 203, 109196, 2023. doi: 10.1016/j.apacoust.2022.109196

[17] X. Wang and Y. Su, "An audio encryption algorithm based on DNA coding and chaotic system," *IEEE Access*, vol. 8, pp. 9260–9270, 2020. doi: 10.1109/ACCESS.2019.2963329

[18] M. Nassef, "an Intelligent audio encryption and compression algorithm inspired by the encoding of various biological sequences," *IEEE* Access, vol. 13, pp. 126334–126354, 2025. doi: 10.1109/ACCESS.2025.3588764

[19] A. Shikder, S. K. Rajput, N. K. Nishchal, O. Matoba, and Y. Awatsuji, "Audio encryption using polarized light beam," *APL Photonics*, vol. 10, no. 3, 036110, 2025. doi: 10.1063/5.0234393

[20] V. Kovtun and Y. Ivanov, "Crypto coding system based on the turbo codes with secret keys," *ICT Express*, vol. 10, no. 2, pp. 330–335, 2024. doi: 10.1016/j.icte.2023.08.007

[21] D. V. Linh and V. V. Yem, "A novel scheme for joint error control and dynamic security coding using puncturing mechanism," *J. Commun.*, vol. 17, no. 11, pp. 948–955, 2022. doi: 10.12720/jcm.17.11.948-955

[22] D. V. Linh, Y. V. Van, and T. P. T. Hoang, "A hybrid of encryption and error correction method for multiple-input multiple-output wireless communication systems," *Int. J. Commun. Networks Distrib. Syst.*, vol. 30, no. 1, pp. 657–678, 2024. doi: 10.1504/ijcnds.2024.10061300

[23] D. V. Linh and V. V. Yem, "A turbo-based encryption and coding scheme for multiple-input multiple-output orthogonal," *IET Commun.*, no. 17, pp. 632–640, 2023. doi: 10.1049/cmu2.12568

[24] X. Lu, J. Lei, W. Li, K. Lai, and Z. Pan, "Physical layer encryption algorithm based on polar codes and chaotic sequences," *IEEE Access*, vol. 7, pp. 4380–4390, 2019. doi: 10.1109/ACCESS.2018.2888883

[25] D. V. Linh, V. V. Yem, and H. T. P. Thao, "Crypto-coding technique based on polar code and secret key generated from wireless channel characteristics for wireless communication systems," *PLoS One*, vol. 20, pp. 1–20, 2025. doi: 10.1371/journal.pone.0318110

[26] K. H. Moussa, S. Shaaban, and A. H. E. Sakka, "Secured polar code derived from random hopped frozen-bits," *Wirel. Networks*, vol. 29, no. 1, pp. 423–435, 2023. doi: 10.1007/s11276-022-03127-1

[27] C. An, M. Huang, X. Lu, L. Bi, and W. Li, "Polar code-based secure transmission with higher message rate combining channel entropy and computational entropy," *Cybersecurity*, vol. 7, no. 1, 36, 2024. doi: 10.1186/s42400-024-00229-5

[28] A. Belmeguenai, Z. Ahmida, S. Ouchtati, and R. Djemii, "A novel approach based on stream cipher for selective speech encryption," *Int. J. Speech Technol.*, vol. 20, no. 3, pp. 685–698, 2017. doi: 10.1007/s10772-017-9439-8

[29] I. E. Hanouti and H. E. Fadili, "Security analysis of an audio data encryption scheme based on key chaining and DNA encoding," *Multimed. Tools Appl.*, vol. 80, no. 8, pp. 12077–12099, 2021. doi: 10.1007/s11042-020-10153-8

Copyright © 2026 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).