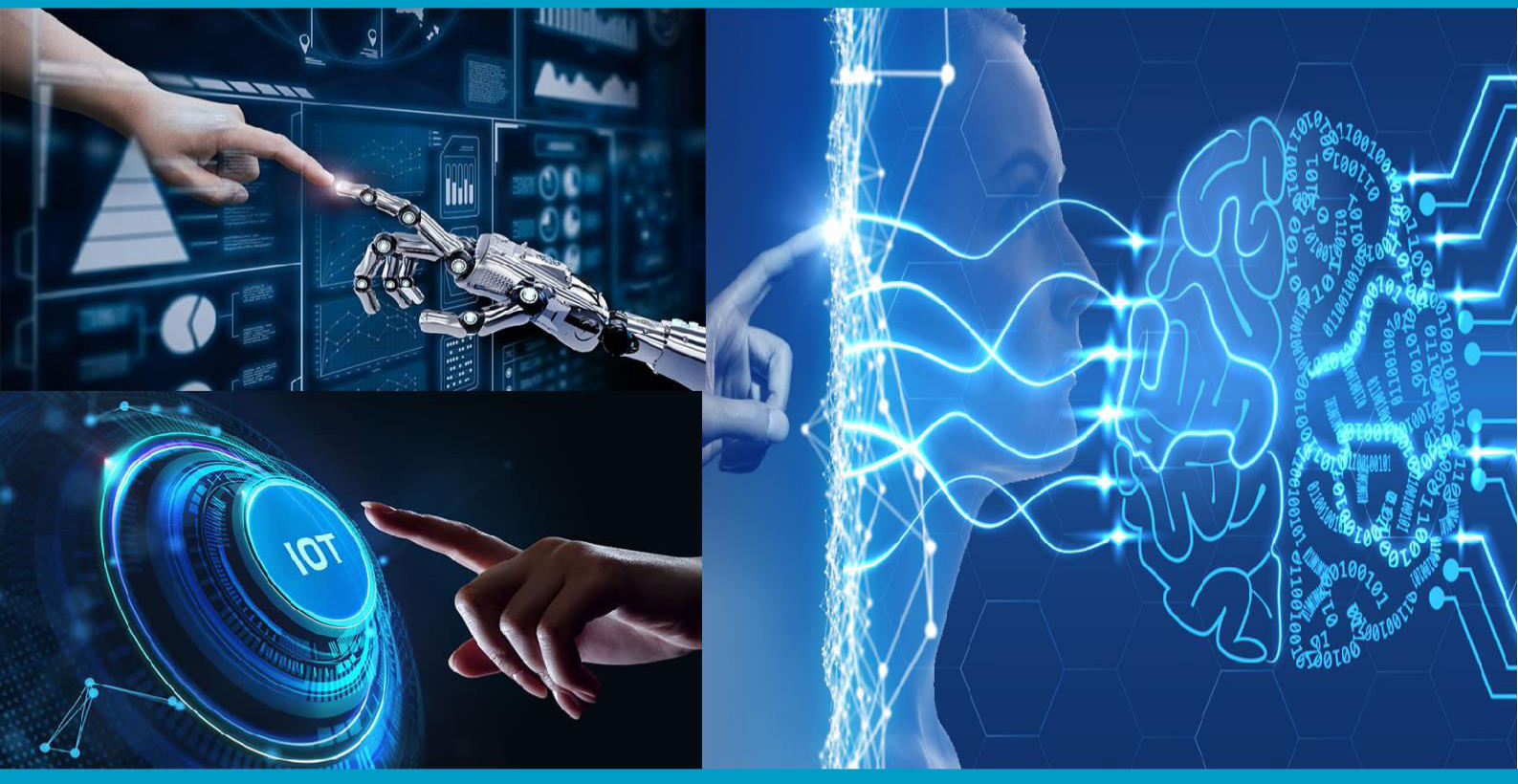
IJMRSETM

e-ISSN: 2395 - 7639



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING, TECHNOLOGY AND MANAGEMENT

Volume 12, Issue 6, June 2025

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.214

International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management (IJMRSETM)

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 8.214 | A Monthly Double-Blind Peer Reviewed Journal |

| Volume 12, Issue 6, June 2025 |

Voice Encryption Meets Watermarking: A Robust Framework for Secure Speech Communication

Jahanvi Pragati [1], Ritu Dagar [2]

P.G. Student, Department of CSE, Sat Kabir Institute of Technology and Management, Haryana, India [1]

Assistant Professor, of CSE, Sat Kabir Institute of Technology and Management, Haryana, India [2]

ABSTRACT: Secure and verified audio transmission has become essential in the age of ubiquitous digital communication. By combining Data Encryption Standard (DES) encryption with spatial domain watermarking (SDW), this research offers a thorough approach to secure audio communication. First, the input voice signal is digitized. Next, a strong digital watermark is embedded into the audio using SDW techniques The watermarked signal, or rather the embedded watermark, is encrypted using DES, a symmetric key encryption algorithm that is well-known for its ease of use and efficiency in resource-constrained settings, to further improve confidentiality. After encryption, the audio is safely transferred or stored. The embedded watermark is extracted for authentication and integrity verification from the original watermarked audio, which is recovered at the receiver's end using DES decryption. In addition to guaranteeing the speech signal's confidentiality and integrity, this two-pronged strategy offers ownership verification and watermark tamper detection. In addition to guaranteeing the speech signal's confidentiality and integrity, this two-pronged strategy offers ownership verification and tamper detection via watermark verification. For encrypted voice communication in wireless networks, Internet of Things devices, and teleconferencing applications, the suggested architecture provides a real-time, lightweight solution.

KEYWORDS: Speech Signal Encryption, Multilayer Security, Steganography in Audio

I. INTRODUCTION

In today's digitally interconnected world, the transmission of speech and audio data has become ubiquitous across a range of applications such as VoIP services, online conferencing, military communication, and Internet of Things (IoT) devices [1]. However, the open nature of communication networks renders audio data susceptible to threats such as eavesdropping, tampering, forgery, and unauthorized access. Consequently, ensuring the confidentiality, authenticity, and integrity of voice communication has become a pressing research challenge [2]. One popular method for preventing unwanted access to private audio data is speech signal encryption. Because of its basic simplicity and performance in real-time systems, the Data Encryption Standard (DES) is still one of the most widely used symmetric encryption algorithms [3]. DES ensures end-to-end data secrecy by converting digital audio into an incomprehensible format that can only be decoded by authorized persons with the right decryption key.

Simultaneously, digital watermarking methods have become effective means of proving authenticity, claiming ownership, and identifying tampering. In particular, Spatial Domain Watermarking (SDW) algorithms provide a quick and easy way to encode metadata without requiring a large amount of computing overhead by directly embedding watermark information into the speech signal's amplitude or sample values [4]. Although spatial domain approaches are typically more brittle than transform domain approaches, they are ideal for real-time communication situations due to their simplicity of use.

In order to provide dual-layer security, this research suggests a secure audio communication system that combines DES encryption and spatial domain watermarking. In the spatial domain, the speech stream is first digitalized and watermarked. Next, DES is used to encrypt the entire audio signal or just the watermark. To confirm the validity and integrity of the signal, the watermark is recovered after the encrypted content has been decrypted at the receiver's end. The suggested approach is appropriate for crucial applications where secure and real-time voice transmission is crucial since it takes care of both data confidentiality and authentication.

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 8.214 | A Monthly Double-Blind Peer Reviewed Journal |

| Volume 12, Issue 6, June 2025 |

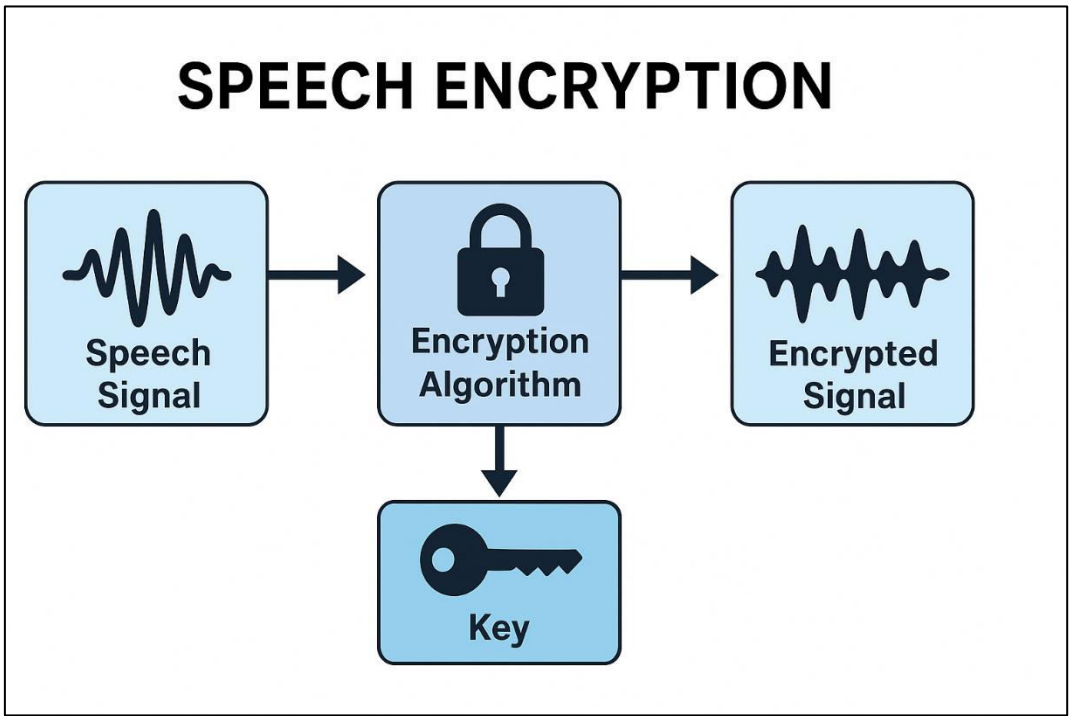


Figure 1: A general instance of Speech Encryption

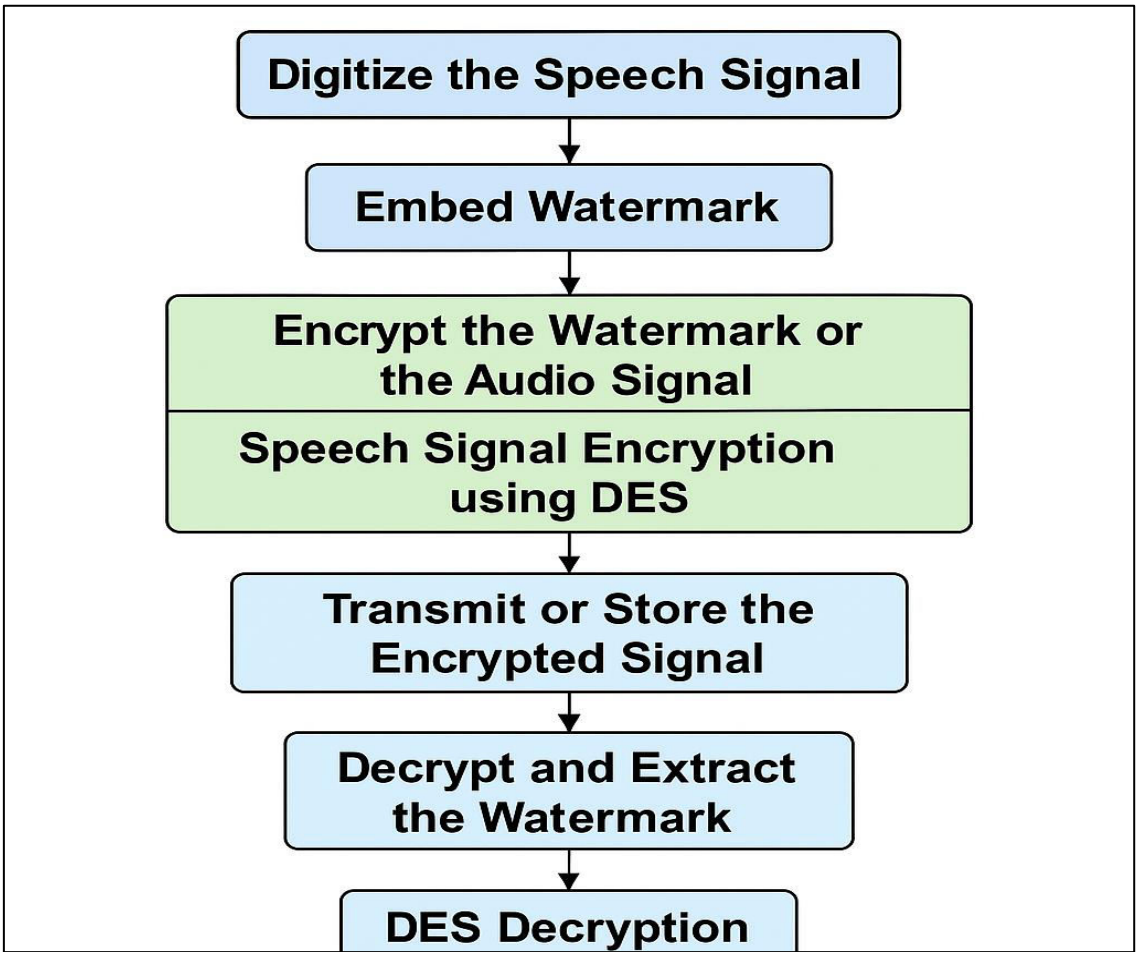


Figure 2: Work Flow

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 8.214 | A Monthly Double-Blind Peer Reviewed Journal |

| Volume 12, Issue 6, June 2025 |

II. RESEARCH BACKGROUND

In [5], a chaotic system with both the confusion and diffusion techniques is used to encrypt the dual-channel audio data with a one-time key. The method makes use of a large keyspace to prevent brute-force attacks. The cosine number transformation has already been applied to non-compressed 16-bit audio data to create a secret key block by block [6]. The storyline of the sequence blends economic and Henon maps. The confusion and diffusion technique is frequently employed to encrypt plain audio data when computing encrypted audio data [7]. Additionally, chaotic systems and DNA coding have been utilized to conflate and spread audio data.

Additionally, chaotic systems and DNA coding have been utilized to conflate and spread audio data. In a new encryption technique, the audio signal is converted into data using a lifting wavelet methodology, and then encrypted using a chaotic dataset and a hyperbolic function. The beginning value of the chaotic system is determined using the audio's hash value [8]. The concepts of a block cipher and chaotic maps are used in block-by-block encryption for audio recordings. A chaotic tent map is used during the permutation step. The retrieved block was then XORed with a key block. The resulting block is substituted using the multiplication inverse-based approach [9].

A unique audio transmission method is explored using DNA coding, chaotic maps, self-adaptive scrambling, and cipher feedback mechanisms. Five distinct chaotic maps and eight control parameters are used to generate a pseudo-random number [10]. A proposed encryption method for audio data generates the pseudo-random number using the chaotic circle map and updated rotation equations [11]. A suggested audio encryption technique is developed by permuting audio samples utilising a discrete improved Henon map and then performing a substitution operation. The keystream is obtained from the modified Lorenz–hyperchaotic system.

Numerous quality standards have been devised to evaluate the encryption method's quality [12]. One innovative method for encrypting voice signals is the use of several chaotic maps with cryptographic protocols. The scrambling process uses a cubic map to divide the incoming signal into four parts. All of the chaotic map parameters are protected by combining the private key with the blowfish algorithm. Between the sender and recipient endpoints, the hashing method and the system's blowfish key are implemented. Throughout secure communication, the chaotic map parameters are authenticated and verified using the message digest.

A number of statistical tests are carried out to show how effective the method is [13]. A new multiuser speech encryption method was developed using a chaos-based cryptosystem. Transmitters and receivers generate chaotic encryption and decryption keys using Chua chaotic systems. The speech stream is encrypted by combining the XOR operation with a chaotic matrix operation for randomization. The security analysis shows how vulnerable secret keys are and how a wide key space is necessary to fend off a brute-force attack. The transmitter's battery life has been extended via strong diffusion and confusion processes [14].

III. PROPOSED METHOD

1. **Digitize the Speech Signal**
   The first stage in encrypted audio communication is digitizing the speech signal, which involves converting analog sound into a digital representation that can be encrypted, transmitted, and stored. Here is a thorough breakdown of the procedure, step-by-step:

   (a) **Speech Signal Acquisition (Analog Signal Capture) and Preprocessing:** The process begins with capturing human speech using a microphone. This captured signal is analog continuous in time and amplitude, representing air pressure variations over time. It enhance quality and prepare the signal for digitization. This step performs noise filtration, normalization and amplification to boost low-power signals.
   (b) Sampling (Temporal Discretization): The analog signal is sampled at regular intervals i.e., the continuous-time signal is converted into discrete-time. To prevent aliasing, the sampling rate must be at least twice the maximum frequency in the signal.
   (c) **Quantization (Amplitude Discretization):** Each sampled value is mapped to the nearest level from a finite set of values. Introduces quantization noise, but makes it easier to represent values in binary.
   (d) Analog-to-Digital Conversion (ADC): It combines sampling and quantization. Each sample is assigned a binary value — e.g., a 16-bit number per sample.
   (e) **Output: Digital Speech Signal:** The final output is a stream of binary numbers representing the voice signal in digital form.

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 8.214 | A Monthly Double-Blind Peer Reviewed Journal |

| Volume 12, Issue 6, June 2025 |

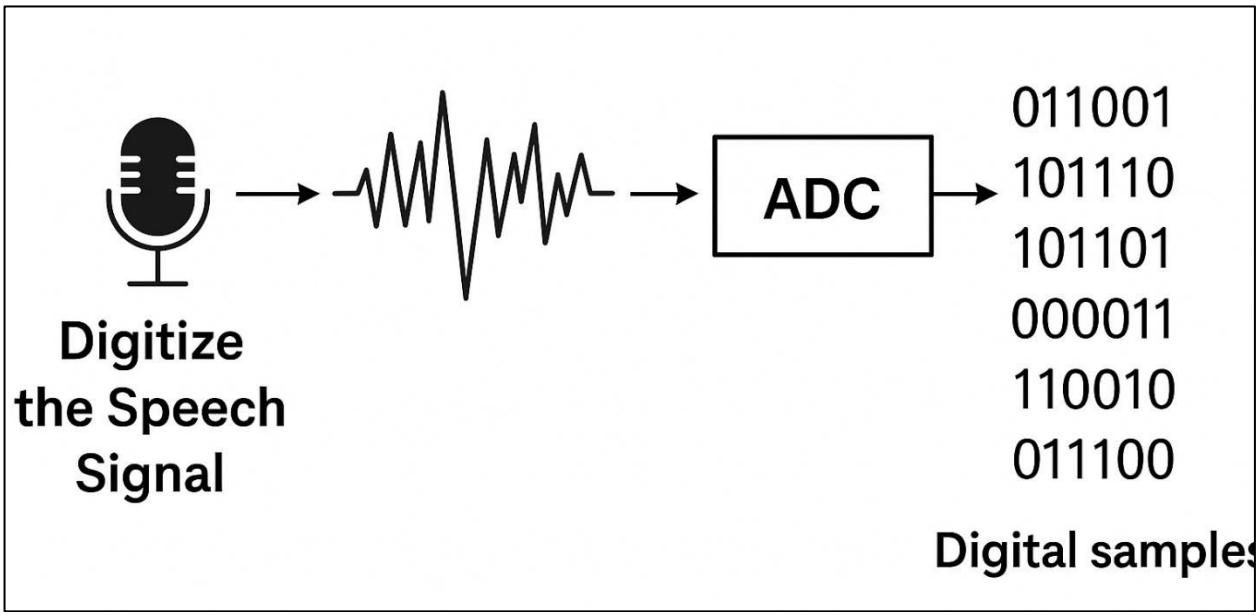


Figure 3: Digitize Speech Signal

2. **Embed Watermark:** One of the simplest and most straightforward methods for digital watermarking, including in audio, image, and video processing, is spatial domain watermarking (SDW). This method involves directly embedding the watermark into the spatial samples of the signal, usually by altering the audio or pixel values' least significant bits (LSBs). SDW involves inserting watermark data into a digital signal's actual values, such as a speech waveform's amplitude values. Although these alterations are typically invisible to the human eye, they can be identified and extracted with the right techniques. Key Steps in SDW for Audio Signals are as follows.
   (a) Preprocessing the Audio: Convert the analog speech signal into digital form. Normalize or filter the signal for consistency.
   (b) Watermark Generation: The watermark can be a text string, biometric pattern, ID, or logo converted into a binary or signal format.
   (c) Embedding the Watermark: Modify the LSBs of audio samples to insert watermark bits.
   (d) Transmission or Storage: The watermarked audio is now ready for secure storage or transmission. Since changes are minimal, audio quality remains largely unaffected.

3. **Encrypt the Watermarked Audio Signal:** The Data Encryption Standard (DES) is a symmetric-key block cipher used to encrypt digital data, including watermarked audio signals, for secure transmission or storage. DES works by encrypting data in 64-bit blocks using a 56-bit secret key. The main idea is to make the watermarked audio unintelligible to unauthorized users, protecting both the content and the embedded watermark. Here's a detailed breakdown of the process of encrypting a watermarked audio signal using DES:

   (a) **Preparation of the Watermarked Audio Signal**: The audio signal (in .wav format) contains the embedded watermark. Convert the audio sample values to binary format. Group the binary stream into 64-bit blocks, as DES operates on these fixed-size blocks.
   (b) **Key Generation**: Choose a 56-bit secret key. DES internally expands this to generate 16 round keys using permutation and shift operations.
   (c) **Initial Permutation (IP):** Each 64-bit block of the audio signal undergoes a fixed initial permutation. This rearranges the bits to create diffusion in the ciphertext.
   (d) **16 Rounds of Feistel Network**: Divide the block into Left (L) and Right (R) halves. Expand the right half (32 bits → 48 bits) using an expansion permutation. XOR the expanded R with the round key. Apply Substitution (S-boxes) to reduce 48 bits → 32 bits. Apply a permutation (P-box). XOR the output with L. Swap L and R (except in the last round).
   (e) Final Permutation (FP): After 16 rounds, a final permutation is applied to the concatenated block (R16 + L16). The output is a 64-bit ciphertext block.
   (f) **Output the Encrypted Audio:** Combine all encrypted 64-bit blocks to form the encrypted audio signal.

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 8.214 | A Monthly Double-Blind Peer Reviewed Journal |

| Volume 12, Issue 6, June 2025 |

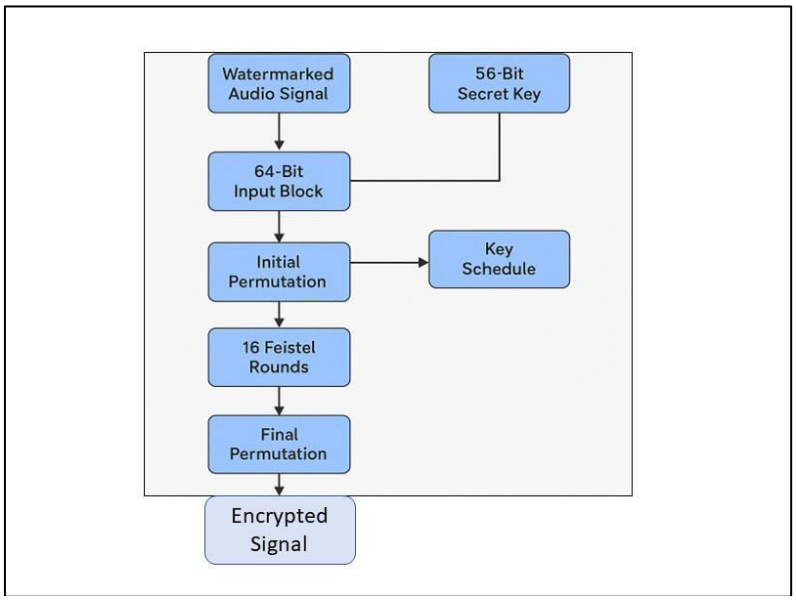


Figure 4: Watermarked Audio Signal

4. **Decrypt and Extract the Watermark:** It involves reversing the encryption process to retrieve both the original audio signal and the embedded watermark. Here's a step-by-step explanation of the decryption and extraction process:

   (a) Input: Encrypted Audio Signal: The receiver gets the encrypted audio signal that contains the watermark embedded in its samples.
   (b) **Divide into 64-bit Blocks:** DES operates on 64-bit blocks, so the encrypted signal is split into 64-bit chunks. Each block is processed individually for decryption.
   (c) **Initial Permutation (IP):** The encrypted 64-bit block undergoes an initial permutation, the same as in encryption. This step rearranges bits to prepare them for the Feistel rounds.
   (d) **16 Feistel Rounds (in Reverse):** Use the same 16 round keys generated during encryption but applied in reverse order.
   (e) **Final Permutation (FP):** After the 16 rounds, a final permutation (inverse of initial permutation) is applied to get the decrypted 64-bit block.
   (f) **Merge Decrypted Blocks:** All the decrypted 64-bit blocks are combined to reconstruct the decrypted audio signal.
   (g) **Watermark Extraction**: Once the audio signal is decrypted: We used the same spatial domain watermarking (SDW) technique that was used for embedding to extract the watermark. Check the LSB of the message. Detect patterns or amplitude modifications in specific segments. The extracted watermark is compared with the original to verify integrity and authenticity.

Simulation and outcomes: We have implemented the proposed method in MATLAB 2024. We have created voice samples using microphone.
Test 1:

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 8.214 | A Monthly Double-Blind Peer Reviewed Journal |

| Volume 12, Issue 6, June 2025 |

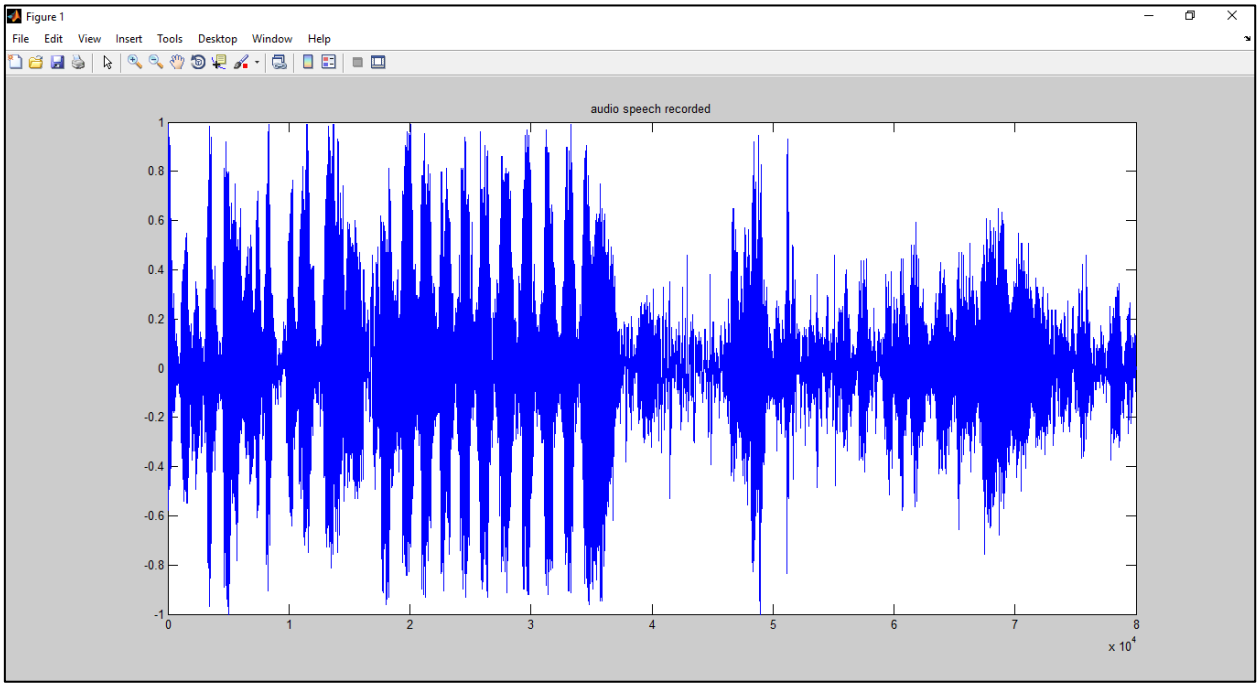


Figure 5: Original Audio

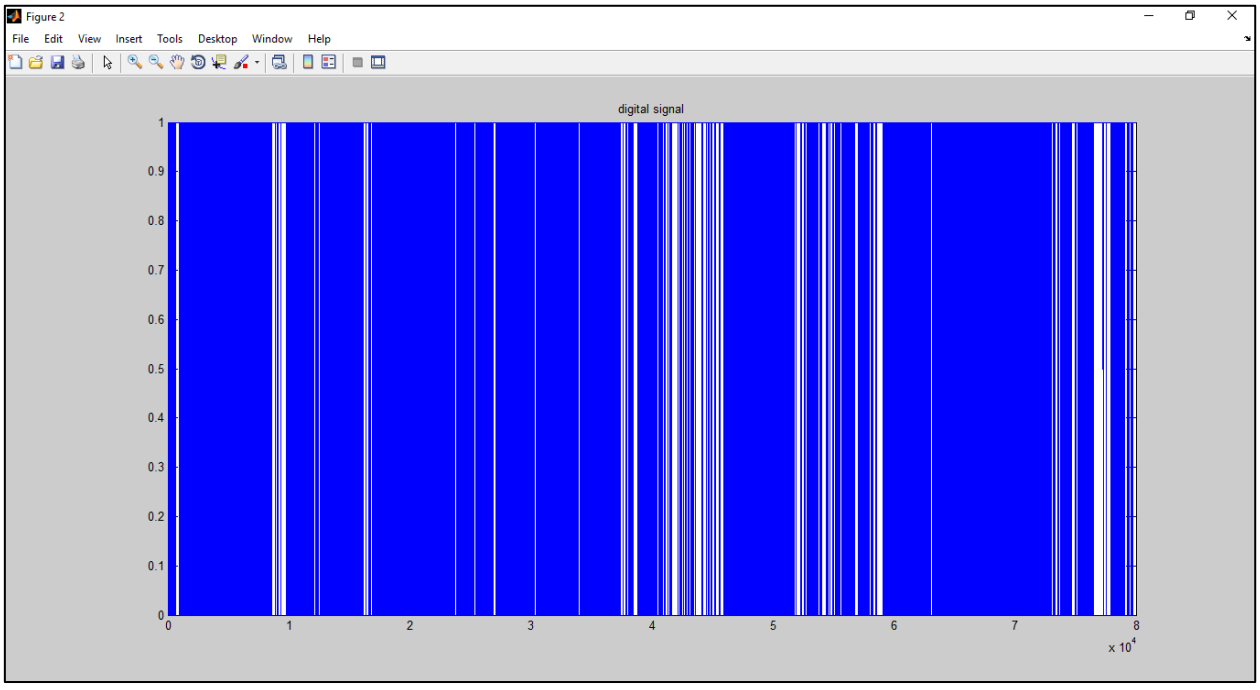


Figure 6: Digitized Speech

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 8.214 | A Monthly Double-Blind Peer Reviewed Journal |

| Volume 12, Issue 6, June 2025 |

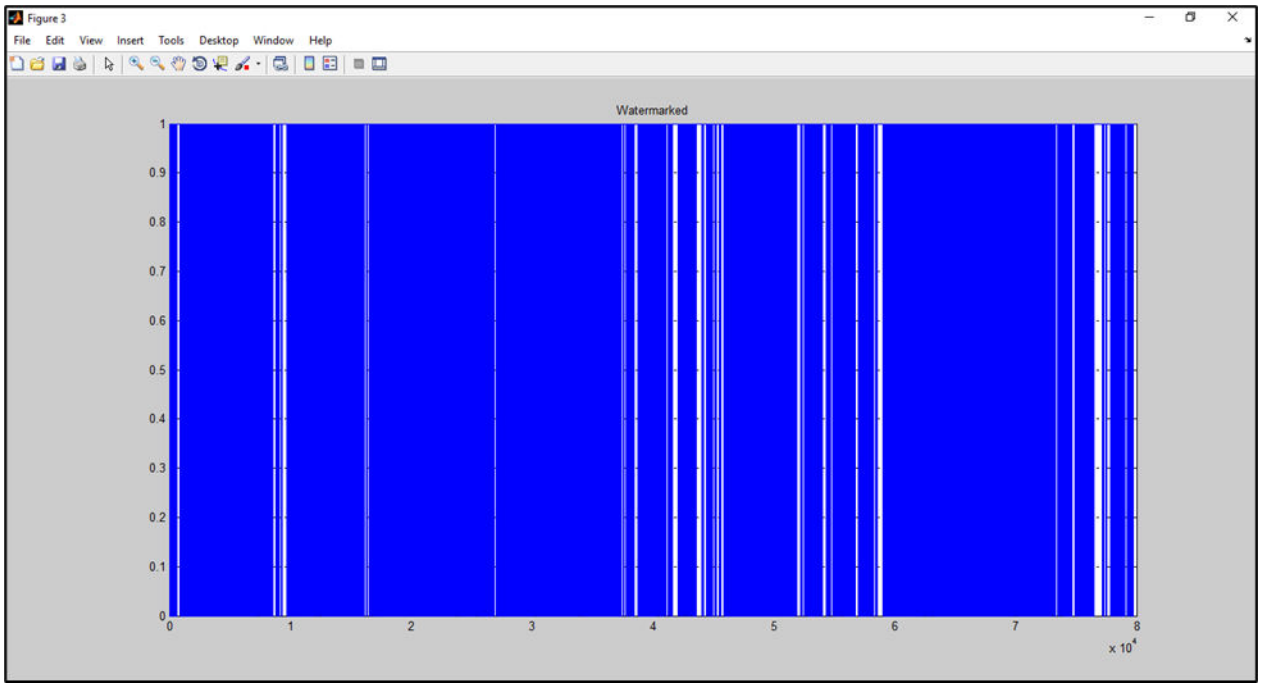


Figure 7: Speech after Watermarking

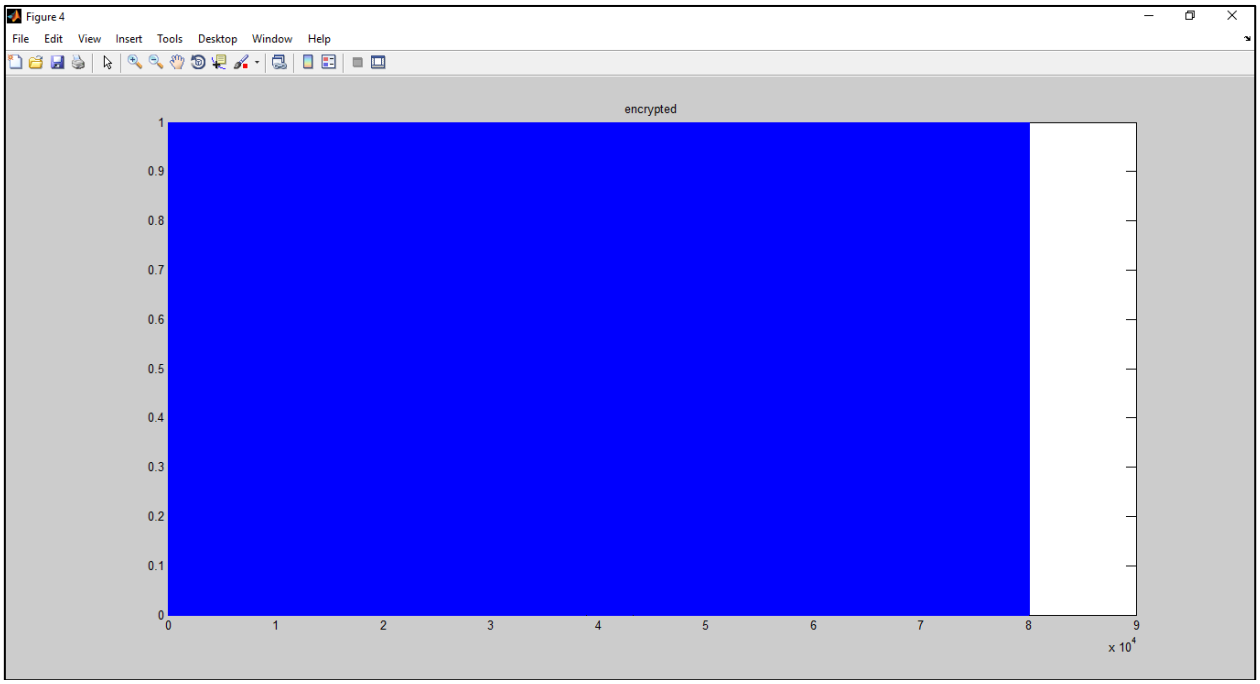


Figure 8: DES encryption based speech

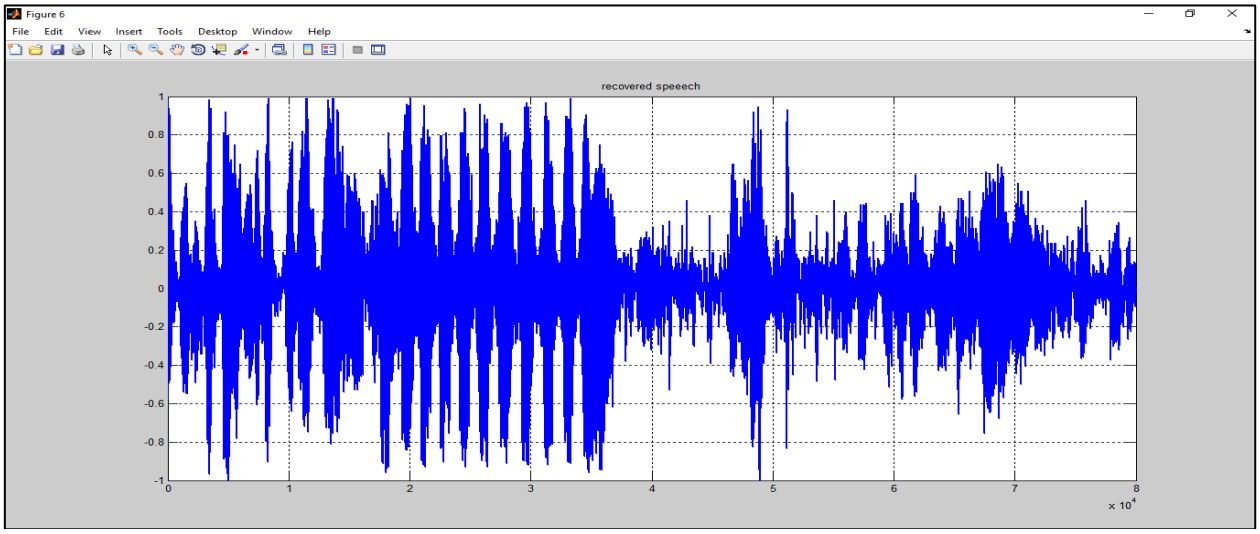


Figure 9: Speech after decryption and watermark removal

Table 1: Comparative analysis of the proposed method with other conventional methods

Criteria	Proposed Method (SDW + DES)	FFT + Chaos-Based Encryption [Younis & Bukhari, 2014]	DWT + RSA Encryption [Poonia & Jain, 2015]	Deep Learning-Based Autoencoder [Zhang & Luo, 2020]
Watermarking Domain	Spatial Domain (e.g., LSB modification)	Frequency Domain (FFT components)	Wavelet Domain (DWT coefficients)	Learned latent representations (autoencoders)
Encryption Algorithm	DES (Symmetric, Block Cipher)	Chaos + FFT Key Generation (Pseudo-random)	RSA (Asymmetric, Key Exchange)	Implicit encoding through trained models
Security Level	Moderate (DES is known but relatively outdated)	High (chaos is unpredictable)	Very High (RSA is asymmetric and strong)	High (security depends on model secrecy and adversarial robustness)
Imperceptibility	Good if watermark strength is optimized	Good but may introduce spectral noise	Excellent, as DWT affects perceptually insignificant components	Excellent (trained to minimize distortion)
Robustness to Attacks	Low–Moderate (Spatial watermarking vulnerable to compression or filtering)	Moderate (depends on FFT resolution and chaos strength)	High (DWT more robust to compression, RSA secure)	High (can generalize and resist distortions)
Complexity	Low (simple to implement and fast)	Medium (FFT and chaos generation require additional computation)	High (DWT + RSA needs high processing and key management)	Very High (requires model training and inference infrastructure)
Real-Time Applicability	Suitable for low-resource systems	Medium (slower due to FFT and key sync)	Less suitable (RSA is computationally expensive)	Depends on model size and inference speed
Key Management	Requires secure symmetric key distribution	Requires initial chaos parameters	Involves key pairs and public key infrastructure (PKI)	Requires model and parameter synchronization
Use Case Suitability	Lightweight, basic watermark-protected communication	Secure voice chats, moderate security needs	Secure and legal-grade transmission	Smart IoT audio, biometric protection, privacy-preserving systems

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 8.214 | A Monthly Double-Blind Peer Reviewed Journal |

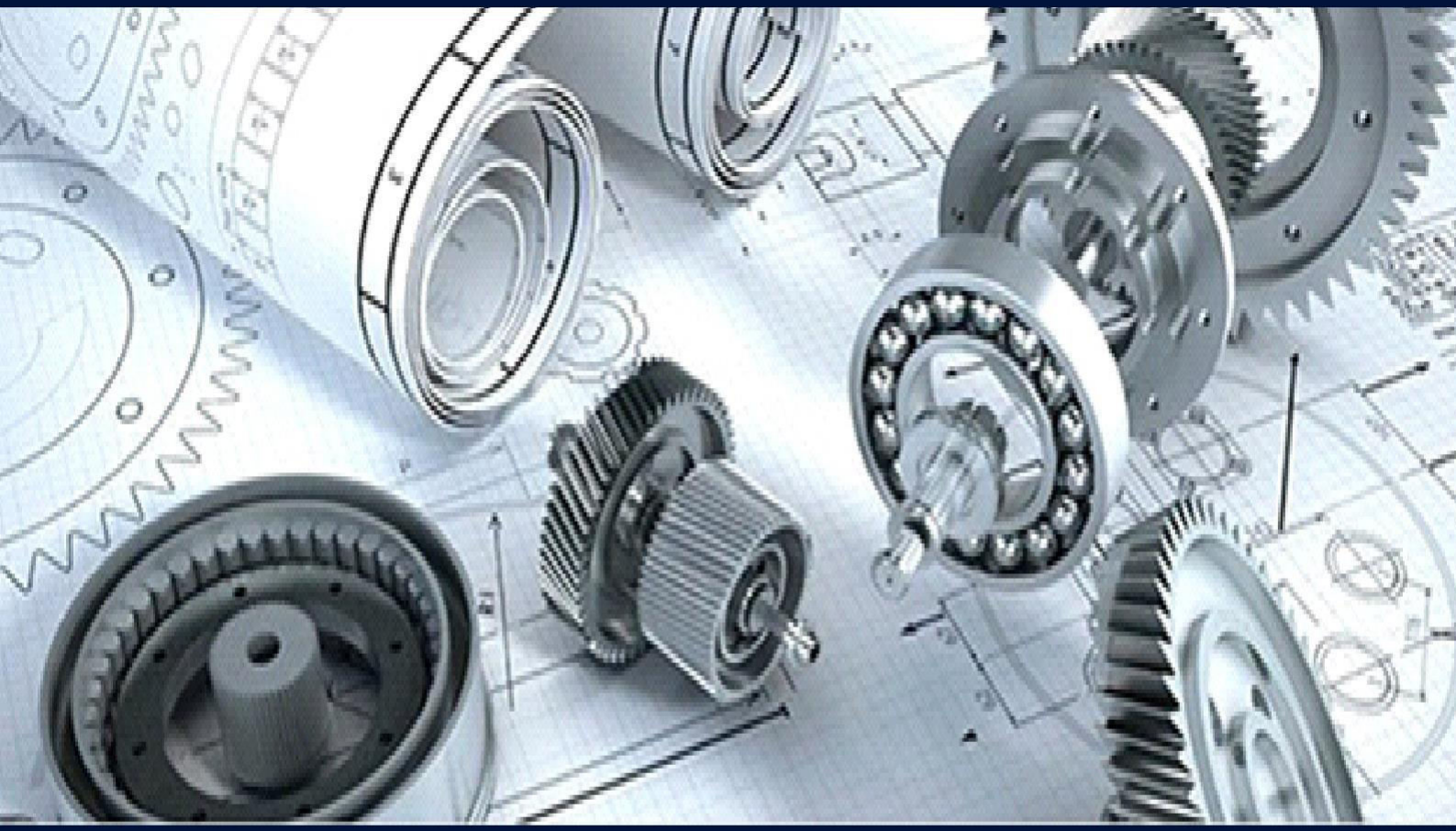
| Volume 12, Issue 6, June 2025 |

IV. CONCLUSION

The suggested technique for secure audio transmission combines DES encryption with Spatial Domain Watermarking (SDW) to provide dual-layer protection: embedded watermarking for authenticity and speech signal secrecy. To ensure imperceptibility and traceability, the method starts with digitizing the speech signal and then embeds a watermark in the spatial domain. The DES algorithm is then used to encrypt the watermarked audio, adding an extra degree of protection against unwanted access while it is being transmitted or stored. The system extracts the encoded watermark to confirm integrity and authenticity after performing DES decryption to retrieve the watermarked audio upon reception. Applications such as secure voice transmission, copyright protection, and tamper-proof communication can benefit from this method's practical harmony among security, computational efficiency, and simplified implementation. DES is perfect for systems with limited resources because of its speed and ease of use, even though it might not provide the strongest cryptography available today. When combined with SDW, the technique guarantees that the watermark will still give forensic attribution even in the event that the encryption is breached. By substituting more sophisticated encryption algorithms like AES for DES and investigating strong watermarking in transform domains for increased resistance to compression and signal deterioration, further research can raise the security level.

REFERENCES

1. Ravinder, Sumit Dalal, Sumiran and Rohini Sharma, “A comprehensive review of voice encryption techniques,” Synergy: Cross-Disciplinary Journal of Digital Investigation, volume 02, issue 6, 2024.
2. Stallings, W. (2017). Cryptography and Network Security: Principles and Practice. Pearson Education.
3. National Bureau of Standards (1977). Data Encryption Standard (DES). FIPS PUB 46.
4. Cox, I. J., Miller, M. L., Bloom, J. A., Fridrich, J., & Kalker, T. (2007). Digital Watermarking and Steganography. Morgan Kaufmann.
5. Liu, H.; Kadir, A.; Li, Y. Audio encryption scheme by confusion and diffusion based on multi-scroll chaotic system and one-time keys. Optik 2016, 127, 7431–7438.
6. Farsana, F.; Devi, V.; Gopakumar, K. An audio encryption scheme based on Fast Walsh Hadamard Transform and mixed chaotic keystreams. Appl. Comput. Inform. 2019. Online ahead of print.
7. Adhikari, S.; Karforma, S. A novel audio encryption method using Henon–Tent chaotic pseudo random number sequence. Int. J. Inf. Technol. 2021, 13, 1463–1471.
8. Wang, X.; Su, Y. An Audio Encryption Algorithm Based on DNA Coding and Chaotic System. IEEE Access 2019, 8, 9260–9270.
9. Albahrani, E.A. A new audio encryption algorithm based on chaotic block cipher. In Proceedings of the 2017 Annual Conference on New Trends in Information & Communications Technology Applications (NTICT), Baghdad, Iraq, 7–9 March 2017; pp. 22–27.
10. Abdelfatah, R.I. Audio Encryption Scheme Using Self-Adaptive Bit Scrambling and Two Multi Chaotic-Based Dynamic DNA Computations. IEEE Access 2020, 8, 69894–69907.
11. Kordov, K.; Bonchev, L. Using Circle Map for Audio Encryption Algorithm. Math. Softw. Eng. 2017, 3, 183–189.
12. Farsana, F.; Devi, V.; Gopakumar, K. An audio encryption scheme based on Fast Walsh Hadamard Transform and mixed chaotic keystreams. Appl. Comput. Inform. 2019. Online ahead of print.
13. Yasser, I.; Mohamed, M.; Samra, A.; Khalifa, F. A Chaotic-Based Encryption/Decryption Framework for Secure Multimedia Communications. Entropy 2020, 22, 1253.
14. Hashemi, S.; Pourmina, M.A.; Mobayen, S.; Alagheband, M.R. Multiuser wireless speech encryption using synchronized chaotic systems. Int. J. Speech Technol. 2021, 24, 651–663.



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING, TECHNOLOGY AND MANAGEMENT

+91 99405 72462 +91 63819 07438 ijmrsetm@gmail.com