Ashish Mishra, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

A Hybrid Approach For Secure Audio Communication Using Encryption And Watermarking

Ashish Mishra [1], Dheeraj Chillar [2]

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

Director, Sat Kabir Institute of Technology and Management, Ladrawan, Haryana[2]

Abstract: With the rapid growth of digital communication systems, ensuring the security and integrity of audio data has become increasingly important. This paper presents a hybrid approach for secure audio communication that integrates encryption and watermarking techniques to provide dual-layer protection. In the proposed system, the audio signal is first processed using transform-based watermarking methods, such as Discrete Wavelet Transform (DWT) and Singular Value Decomposition (SVD), to embed hidden information for authentication and ownership verification. The watermarked audio is then encrypted using the Advanced Encryption Standard (AES) to ensure confidentiality during transmission. At the receiver side, the encrypted signal is decrypted to recover the watermarked audio, followed by watermark extraction and verification to ensure data integrity. The performance of the proposed system is evaluated using quantitative metrics, including Signal-to-Noise Ratio (SNR), Mean Squared Error (MSE), and execution time. Experimental results demonstrate that the DWT-based approach provides better audio quality and robustness compared to the SVD-based method, while maintaining acceptable computational efficiency. The proposed hybrid framework effectively combines the strengths of encryption and watermarking, offering a secure, reliable, and practical solution for audio communication systems.

Keywords: Audio Security, Audio Watermarking, AES Secure Communication.

Introduction: In recent years, the widespread use of digital audio in communication systems has raised serious concerns regarding data security and authenticity. Audio signals, being easily accessible and transferable over networks, are highly susceptible to interception, duplication, and unauthorized modification. This has created a need for advanced techniques that not only protect the confidentiality of audio data but also ensure its integrity and ownership verification. Conventional encryption techniques have been extensively used to secure audio signals by converting them into unintelligible formats. Algorithms such as Blowfish and its variants have demonstrated efficiency in speech encryption due to their fast processing [1], [2]. Similarly, RSA-based methods provide strong security through public-key cryptography, although they are computationally intensive and less suitable for real-time applications [3]. Studies have also highlighted the importance of selecting appropriate cryptographic techniques based on application requirements, particularly in speech signal processing [4]. In addition, chaos-based encryption methods have gained attention for their ability to generate highly unpredictable sequences, offering enhanced security for multimedia data [5], [6]. Despite these advancements, encryption alone cannot guarantee complete protection, as it does not provide mechanisms for verifying data authenticity after decryption. To address this limitation, additional security strategies such as watermarking and authentication schemes, including fuzzy commitment techniques, have been explored to enhance data integrity and reliability [7]. Moreover, the unique characteristics of speech signals, as studied in linguistic frameworks [8], play a crucial role in the design of efficient and perceptually acceptable security systems. Therefore, there is a growing need for integrated approaches that combine multiple security mechanisms to achieve comprehensive protection. This work focuses on developing a hybrid framework that incorporates encryption and watermarking techniques to ensure both confidentiality and authenticity of audio signals. By leveraging the strengths of different methods, the proposed approach aims to provide a

secure and efficient solution for modern audio communication systems.

II. RESEARCH BACKGROUND

Over the years, numerous techniques have been developed to enhance the security of audio and speech signals, focusing on encryption, scrambling, and watermarking methods. Early research primarily focused on signal scrambling techniques, in which the speech signal is transformed in the time or frequency domain to prevent unauthorized interpretation. Chen et al. [9] proposed a frequency-domain speech scrambling method for end-to-end encryption, while Zeng et al. [10] introduced a compressed sensing-based approach that improves both security and efficiency. Traditional scrambling methods, such as the Fibonacci transformation [11] and Hadamard matrix-based techniques [12], have also been investigated for their simplicity and effectiveness.

As cryptographic research advanced, more sophisticated encryption techniques were introduced. Quasigroup-based encryption methods have shown strong performance in securing speech signals due to their multilevel structure [13]. Public key cryptography, although computationally intensive, has been applied for secure communication in audio systems [14]. Symmetric encryption algorithms such as AES [15] and Blowfish [16] are widely used for their efficiency and robustness, with modified versions specifically designed for speech encryption applications [17]. Comparative studies have also highlighted the importance of selecting suitable cryptographic algorithms based on performance requirements [4].
In recent years, chaos-based encryption techniques have attracted significant attention due to their high sensitivity to initial conditions and strong randomness. Fridrich [18] demonstrated the use of chaotic maps for the design of secure cryptographic systems, while Kocarev [19] provided an overview of chaos-based cryptography. Advanced implementations, such as synchronized hybrid chaotic generators for real-time speech encryption [20] and chaotic shift keying methods, have further enhanced system security. Comparative analyses of chaotic systems for audio encryption have shown improved resistance to attacks compared to traditional methods. Hybrid approaches combining chaos with classical algorithms, such as Blowfish, have also been proposed to strengthen encryption performance.

Beyond encryption, audio watermarking techniques have been developed to ensure data integrity, authentication, and copyright protection. The theoretical foundation of information hiding was established by Moulin and O'Sullivan [21], highlighting key trade-offs between robustness and imperceptibility. Transform-domain watermarking techniques, particularly those based on wavelet transforms and singular value decomposition, have been widely adopted due to their robustness against signal processing attacks. For example, adaptive watermarking using SVD in the wavelet domain has shown improved performance [30], while neural network-based watermarking methods have enhanced robustness and perceptual quality [22]. Recent developments also include real-time audio watermarking solutions for practical applications [23].

Despite significant advancements in both encryption and watermarking, many existing approaches address these techniques independently, leading to limitations in achieving comprehensive security. Encryption ensures confidentiality but lacks mechanisms for post-decryption verification, while watermarking provides authentication but does not prevent unauthorized access. Therefore, there is a growing need for hybrid approaches that integrate encryption and watermarking to provide dual-layer protection. This work builds upon existing research by combining AES-based encryption with transform-domain watermarking techniques and evaluating their performance for secure audio communication.

iII. PROPOSED METHODOLOGY

Figure 1: Proposed Methodology

1. **Audio Recording (Real-Time Speech Signal):** The process begins with recording a real-time speech signal using a microphone. The analog speech signal is converted into a digital signal through sampling and quantization. Mathematically, the continuous-time signal x(t) is converted into a discrete-time signal:

$$x[n] = x(nT_s) \quad (1)$$

Where, T_s is the sampling interval, and n is the sampling index. This produces a digital audio sequence suitable for further processing.

2. **Input Audio Signal (Preprocessing):** The recorded audio signal is preprocessed to ensure consistency and improve performance. This includes normalization and framing. The signal is normalized as:

$$x_{norm}[n] = \frac{x[n]}{\max(|x[n]|)} \quad (2)$$

This ensures the amplitude lies within [-1, 1], preventing distortion and making the signal suitable for watermarking and encryption.

3. **A. DWT Watermarking (Embedding):** In the DWT-based approach, the audio signal is decomposed into approximation and detail coefficients:

$$x[n] \rightarrow DWT\{CA, CD\} \quad (3)$$

The watermark w[n] is embedded into the approximation coefficients:

$$CA' = CA + \alpha . w[n] \quad (4)$$

Where, α is the embedded strength. The watermarked signal is reconstructed using inverse DWT.

$$x_w[n] = IDWT(CA', CD) \quad (5)$$

B. SVD Watermarking (Embedding): In the SVD approach, the audio signal is reshaped into a matrix AAA, and Singular Value Decomposition is applied:

$$A = USV^T \quad (6)$$

The watermark is embedded by modifying singular values:

$$S' = S + \alpha W \quad (7)$$

The watermarked matrix is reconstructed as:

$$A' = US'V^T \quad (8)$$

and converted back to a 1D audio signal.

4. **AES Encryption:** The watermarked audio is encrypted using AES. The audio signal is first converted into byte form:

$$x_b = typecast(x_w[n], unit8) \quad (9)$$

Encryption is performed as

$$C = AES_k(x_b) \quad (10)$$

where KKK is the secret key and CCC is the encrypted data.

5. **AES Decryption:** At the receiver, the encrypted signal is decrypted using the same key:

$$x_b = AES_K^{-1}(C) \quad (9)$$

The byte data is converted back to audio:

$$x_d[n] = \frac{int16(x_b)}{32767} \quad (10)$$

6. **Reconstructed Audio**

The decrypted signal represents the reconstructed audio signal:

$$x_r \approx x_n \quad (11)$$

This signal contains the embedded watermark and is used for evaluation.

The proposed system begins of acquiring a real-time speech signal with a microphone, which is converted to a digital audio signal via sampling and quantization. For example, a 5-second speech recording sampled at 8 kHz produces a discrete signal x[n]. This signal is then preprocessed by normalizing its amplitude to ensure uniformity and avoid distortion, using equation 2. The normalized audio is then processed through two parallel paths for comparison.

In the first path, the Discrete Wavelet Transform (DWT) is applied to decompose the audio signal into approximation and detail coefficients. A watermark signal w[n], such as a random binary sequence or logo representation, is embedded into the approximation coefficients using a scaling factor α using equation 4. The modified coefficients are then reconstructed using inverse DWT to obtain the watermarked audio signal xw[n]. In the second path, the audio signal is reshaped into a matrix and Singular Value Decomposition (SVD) is applied using equation 6. The watermark is embedded by modifying the singular values using equation 7, and the signal is reconstructed as A′, which is then converted back into a one-dimensional audio signal. After watermarking, both signals are encrypted using the AES. The audio is first converted into byte format and encrypted using a secret key K, resulting in encrypted data C (equation 10). At the receiver side, the encrypted signal is decrypted using the same key to recover the watermarked audio. The decrypted signal is then converted back into its original numerical format to obtain the reconstructed audio.

IV. PERFORMANCE ANALYSIS

The system performance is evaluated using three metrics: Signal-to-Noise Ratio (SNR), Mean Squared Error (MSE), and Comparison Output.

OUTPUTS

Figure 3 shows the time-domain waveform of the audio signal recorded by the microphone in the proposed system. The horizontal axis indicates the sample index (or time progression), while the vertical axis represents the amplitude of the audio signal, typically normalized within a range close to [−1,1]. The waveform clearly shows amplitude variations corresponding to different speech segments, with higher peaks representing louder speech (higher speech intensity) and flatter regions corresponding to silence or low-energy portions of the signal. The non-uniform pattern of the waveform reflects the natural characteristics of human speech, including pauses, syllables, and varying pitch levels. For example, regions with dense oscillations indicate active speech, while near-zero amplitude sections represent silence or background noise. This recorded signal serves as the system's input, which is later processed through normalization, watermarking (DWT/SVD), and encryption. Mathematically, the signal can be represented as a discrete-time sequence x[n], where each sample corresponds to the amplitude of the speech signal at a specific time instant. This raw waveform is essential for further processing, as it retains all the necessary information required for secure audio transformation and analysis.

Figure 2: Main Window

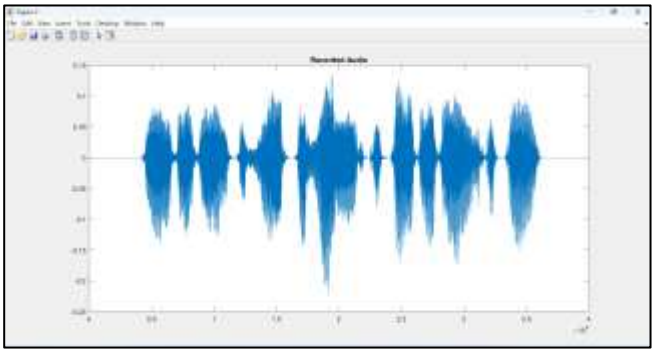

Figure 3: Recorded Audio

ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science,
Engineering and Technology
An Open Access Journal

Path 1: AES+DWT

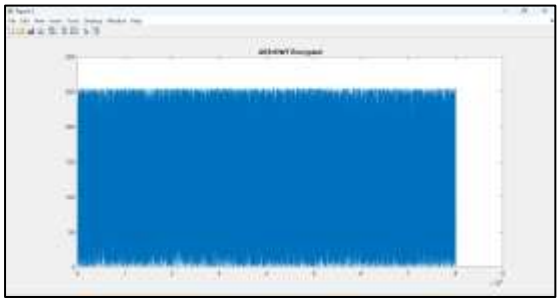

Figure 4: AES + DWT hybrid encrypted audio signal

Figure 4 shows the encrypted audio signal obtained after applying DWT-based watermarking and then AES encryption. Unlike the original audio waveform, which shows structured patterns corresponding to speech, this encrypted signal appears highly random and noise-like. This randomness is a key characteristic of a secure encryption process. The horizontal axis represents the sample index, while the vertical axis shows the amplitudes of the encrypted data**,** now in byte form (typically ranging from 0 to 255). The absence of recognizable patterns or smooth variations indicates that the original speech information has been completely transformed into an unintelligible form. This transformation occurs because AES encryption performs multiple rounds of substitution and permutation operations, effectively scrambling the statistical structure of the signal. The dense and uniform distribution of values across the plot confirms that the encryption process has achieved a high level of randomness, making it extremely difficult for an attacker to extract meaningful information without the correct key.

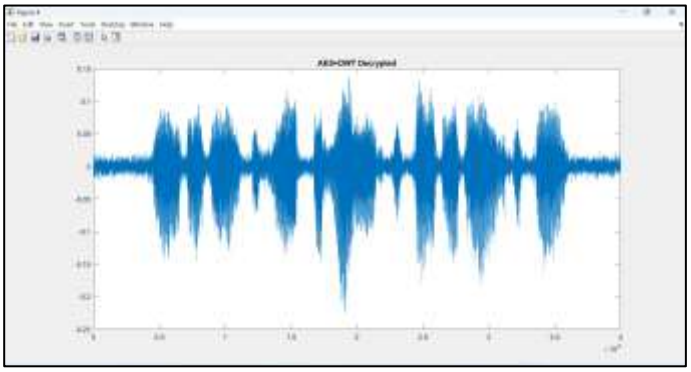

Figure 5: AES_DWT Decrypted

Figure 5 shows the decrypted audio signal obtained after applying AES decryption to the encrypted watermarked audio (DWT-based). Unlike the encrypted signal, which appears completely random, this waveform closely resembles the original recorded audio signal, indicating successful recovery of the speech information. The horizontal axis represents the sample index (time progression), while the vertical axis shows the amplitude of the audio signal. The waveform exhibits clear speech patterns, including peaks and valleys corresponding to different phonetic components of the spoken signal. Regions of higher amplitude correspond to louder speech segments, while lower-amplitude regions indicate pauses or silence. The decryption process reverses the encryption operation using the same secret key K. Although the waveform closely matches the original signal, slight variations may be observed due to the watermark embedding process and numerical transformations. These minor differences contribute to small reconstruction errors, which are later quantified using performance metrics such as SNR and MSE. The figure confirms that the AES decryption process is effective, successfully restoring the audio signal while preserving the embedded watermark, thereby maintaining both security and signal integrity.

Path 2: AES+SVD

Figure 6 illustrates the encrypted audio signal obtained after applying SVD-based watermarking followed by AES encryption. Similar to the AES + DWT encrypted signal, this waveform appears completely random and noise-like, with no visible structure or recognizable speech patterns. This randomness indicates that the original audio information has been successfully transformed into an unintelligible form. The horizontal axis represents the sample index, while the vertical axis shows the amplitude values of the encrypted data, typically ranging between 0 and 255 due to byte-level representation. The dense, uniform distribution of values across the entire signal confirms that the encryption process has effectively removed any statistical or temporal correlation in the original speech signal. The absence of identifiable features in the waveform demonstrates the strong security provided by AES encryption, ensuring confidentiality even if the signal is intercepted. This confirms that the combination of SVD watermarking and AES encryption successfully protects the audio data from unauthorized access.

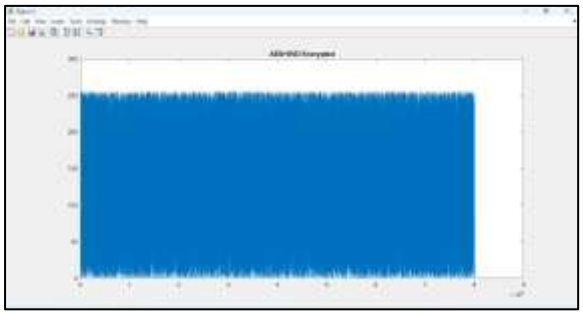

Figure 6: AES+SVD Encryption

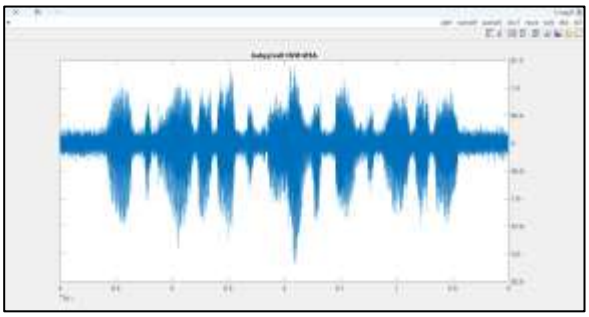

Figure 7: AES+SVD Decryption

Figure 7 shows the decrypted audio signal obtained after applying AES decryption to the SVD-based watermarked audio. Unlike the encrypted signal, which appears completely random, this waveform clearly resembles the structure of the original recorded speech signal, indicating successful recovery of the audio data. The horizontal axis represents the sample index (time progression), while the vertical axis shows the amplitude of the audio signal. The waveform displays distinct speech patterns, including peaks and troughs corresponding to variations in speech intensity. Regions of higher amplitude indicate active speech segments, while lower amplitude regions represent pauses or silence. The decryption process reverses the AES encryption using the same secret key. Although the overall structure of the waveform is similar to the original signal, slight distortions or variations can be observed. These differences arise from the SVD watermark-embedding process, particularly the reshaping of the audio signal into a matrix and the modification of its singular values. This can introduce minor reconstruction errors, which are reflected in slightly lower SNR and higher MSE values compared to the DWT-based approach.

Quantitative performance comparison of the two proposed methods, AES + DWT and AES + SVD, evaluated using Signal-to-Noise Ratio (SNR), Mean

Squared Error (MSE), and execution time. The AES + DWT method achieves an SNR of 11.81 dB, which is significantly higher than the 8.68 dB obtained by AES + SVD. A higher SNR indicates that the reconstructed audio signal is closer to the original signal, with less noise and distortion. This confirms that the DWT-based watermarking approach preserves audio quality more effectively than the SVD-based method.

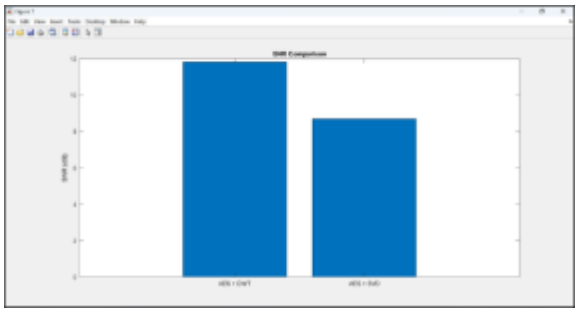

Figure 8: SNR Comparison

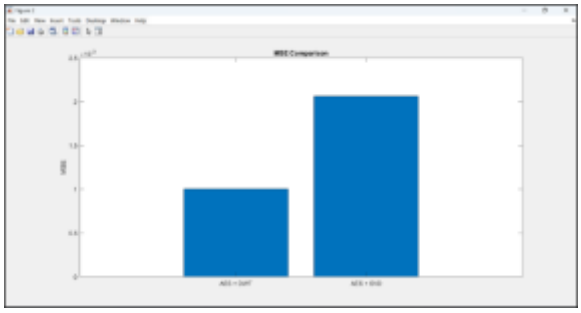

Figure 9: MSE Comparison

Similarly, the MSE value for AES + DWT is 0.001003, which is lower than the 0.002063 observed for AES + SVD. Since MSE measures the average error between the original and reconstructed signals, a lower value indicates better reconstruction accuracy. This further validates that the DWT method introduces less distortion compared to SVD. However, in terms of computational efficiency, the AES + SVD method performs faster, with an execution time of 0.0023 seconds, compared to 0.0042 seconds for AES + DWT. This is because DWT involves additional decomposition and reconstruction steps, increasing computational complexity. Thus, DWT is more suitable for applications requiring high-quality audio reconstruction, while SVD may be preferred in time-sensitive scenarios.

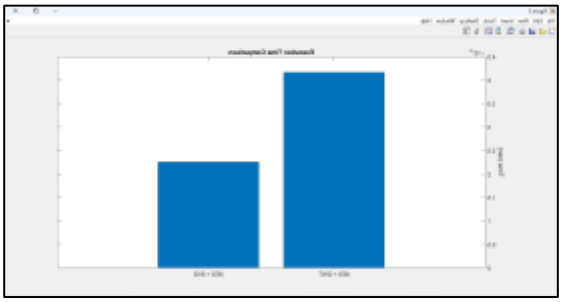

Figure 10: Execution Time Comparison

V. CONCLUSION

The experimental results demonstrate the effectiveness of the proposed hybrid audio security system, which combines AES encryption with transform-based watermarking techniques. A comparative analysis between the AES + DWT and AES + SVD approaches shows that both methods successfully secure and reconstruct the audio signal after encryption and decryption. However, clear differences are observed in terms of performance metrics. The AES + DWT method achieves superior audio quality, as indicated by its higher SNR (11.81 dB) and lower MSE (0.001003), demonstrating better preservation of the original signal and reduced distortion. In contrast, the AES + SVD method exhibits lower computational time (0.0023 seconds), making it more efficient in terms of processing speed. These results highlight a trade-off between signal quality and computational efficiency. The study concludes that DWT-based watermarking is more suitable for audio applications where signal fidelity is critical, while SVD-based methods are advantageous in time-sensitive scenarios. The integration of AES encryption ensures strong security in both cases, making the proposed system a reliable solution for secure audio communication.

REFERENCES

1. K. Nagaraj, "Understanding Blowfish Encryption Algorithm." https://cyberw1ng.medium.com/understanding-blowfish-encryption-algorithm-2023-24eb8f69f85b.
2. A. A. Abd El-Sadek, A. Talaat, and M. M. Fouad, "Speech encryption applying a modified Blowfish algorithm," in 2014 International Conference on Engineering and Technology (ICET), 2014, pp. 1–6.
3. S. F. Yousif, "Encryption and decryption of audio signal based on Rsa algorithn," Int. J. Eng. Technol. Manag. Res., vol. 5, no. 7, pp. 57–64, 2018.L. Nan, S. Yanhong, and Z. Jiancheng, "An audio scrambling method based on Fibonacci transformation," J. North China Univ. Technol, vol. 16, no. 3, pp. 8–11, 2004.
4. R. Aparna and P. I. Chithra, "A review on cryptographic algorithms for speech signal security," Int. J. Emerg. Trends Technol. Comput. Sci., vol. 5, no. 5, pp. 84–88, 2016..
5. H. Oğraş and M. Türk, "A secure chaos-based image cryptosystem with an improved sine key generator," Am. J. Signal Process., vol. 6, no. 3, pp. 67–76, 2016.
6. A. Belazi and A. A. Abd El-Latif, "A simple yet efficient S-box method based on chaotic sine map," Optik (Stuttg)., vol. 130, pp. 1438–1444, 2017.
7. A. Juels and M. Wattenberg, "A fuzzy commitment scheme," in Proceedings of the 6th ACM conference on Computer and communications security, 1999, pp. 28–36.
8. A. Akmajian, A. K. Farmer, L. Bickmore, R. A. Demers, and R. M. Harnish, Linguistics: An introduction to language and communication. MIT press, 2017.
9. Lee, L.S., Chou, G.C. and Chang, C.S., 1984. A new frequency domain speech scrambling system which does not require frame synchronization. IEEE transactions on communications, 32(4), pp.444-456.
10. Gao, Z., Dai, L., Han, S., Wang, Z. and Hanzo, L., 2018. Compressive sensing techniques for next-generation wireless communications. IEEE Wireless Communications, 25(3), pp.144-153.
11. Zou, J., Ward, R.K. and Qi, D., 2004, May. The generalized Fibonacci transformations and application to image scrambling. In 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing (Vol. 3, pp. iii-385). IEEE.
12. Palanisamy, S. and Alshalali, T.A.N., 2025. A novel Hadamard matrix based hybrid compressive sensing technique for enhancing energy efficiency and network longevity. Scientific Reports, 15(1), pp.1-20.
13. M. Satti and S. Kak, "Multilevel indexed quasigroup encryption for data and speech," IEEE Trans. Broadcast., vol. 55, no. 2, pp. 270–281, 2009.
14. M. S. Anoop, "Public Key Cryptography," Retrieved Oct., vol. 6, p. 2010, 2007.
15. National Institute of Standards and Technology (NIST), Advanced Encryption Standard (AES), FIPS PUB 197, 2001.
16. K. Nagaraj, "Understanding Blowfish Encryption Algorithm." https://cyberw1ng.medium.com/understanding-blowfish-encryption-algorithm-2023-24eb8f69f85b.
17. A. A. Abd El-Sadek, A. Talaat, and M. M. Fouad, "Speech encryption applying a modified Blowfish algorithm," in 2014 International Conference on Engineering and Technology (ICET), 2014, pp. 1–6.
18. J. Fridrich, "Symmetric ciphers based on two-dimensional chaotic maps," Int. J. Bifurc. chaos, vol. 8, no. 06, pp. 1259–1284, 1998.
19. L. Kocarev, "Chaos-based cryptography: a brief overview," IEEE Circuits Syst. Mag., vol. 1, no. 3, pp. 6–21, 2001.
20. M. S. Azzaz, C. Tanougast, S. Sadoudi, and A. Bouridane, "Synchronized hybrid chaotic generators: Application to real-time wireless speech encryption," Commun. Nonlinear Sci. Numer. Simul., vol. 18, no. 8, pp. 2035–2047, 2013.
21. P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," IEEE Trans. Inf. theory, vol. 49, no. 3, pp. 563–593, 2003.
22. C. Maha, E. Maher, and B. A. Chokri, "A blind audio watermarking scheme based on neural network and psychoacoustic model with error correcting code in wavelet domain," in 2008 3rd International Symposium on Communications, Control and Signal Processing, 2008, pp. 1138–1143.

23. Y.-Y. Tai, “Audio watermarking algorithm is first to solve ‘second-screen problem’ in real time.” https://www.amazon.science/blog/audio-watermarking-algorithm-is-first-to-solve-second-screen-problem-in-real-time