Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Speech Encryption Scheme for secure Voice Communication

Shraddha Patil, Priti Patil, Sujit Yadav, Tejas Bavache
UG students, Department of Electronics & Telecommunication Engineering, KIT'S College of Engineering, Kolhapur, Maharashtra, India

Prof. V. V. Patil
Professor, Department of Electronics & Telecommunication Engineering, KIT'S College of Engineering, Kolhapur, Maharashtra, India

Abstract - A secure voice communication system is developed using ESP32 microcontrollers to facilitate real-time, encrypted audio transmission. Digital voice signals are captured via an INMP441 microphone through the I2S protocol and protected using AES encryption to ensure data confidentiality. Transmission is managed by the ESP-NOW protocol, providing a low-latency, peer-to-peer connection that functions independently of external network infrastructure. Integration of a MAX98357A amplifier and a push-to-talk mechanism optimizes audio output quality while maximizing power efficiency during operation. Such a framework provides a robust and cost-effective solution for specialized communication needs in defense, industrial monitoring, and emergency response scenarios.

Keywords: AES Encryption, ESP-NOW Protocol, I2S Audio Interfacing, Speech Encryption

I. INTRODUCTION

Rapid advancements in wireless technology have increased the demand for secure and private communication channels, particularly in sensitive sectors such as defense and industrial operations. Conventional analog voice transmission is highly susceptible to eavesdropping and signal interference, making digital encryption a necessity for maintaining data integrity. Modern embedded systems now offer the processing power required to handle real-time audio digitization and complex cryptographic algorithms simultaneously.

The primary objective of this research is to develop a robust, peer-to-peer speech encryption framework using the ESP32 microcontroller. By integrating digital audio interfacing through the I2S protocol and applying AES encryption, the system ensures that voice data remains confidential during transit. Utilizing the ESP-NOW protocol allows for low-latency communication without relying on traditional network infrastructures like Wi-Fi routers. This approach results in a cost-effective, portable, and highly secure solution for short-range voice communication in environments where privacy is paramount.

The hardware integration focuses on high-fidelity digital signal processing to eliminate the noise associated with traditional analog circuits. By employing the I2S (Inter-IC Sound) protocol, the system maintains a direct digital path from the microphone to the encryption engine, preserving audio clarity and reducing the computational overhead of signal conversion. This seamless synchronization between the INMP441 sensor and the ESP32's processing cores allows for the execution of the Advanced Encryption Standard (AES) in real-time, ensuring that the latency remains below the threshold of human perception. Consequently, the project demonstrates that sophisticated cryptographic security can be achieved on low-power, cost-effective embedded hardware without compromising the user experience of a fluid, natural conversation.

LITERATURE SURVEY

The development of secure voice communication systems requires a multidisciplinary approach involving real-time digital signal processing, cryptographic efficiency, and low-latency wireless transmission. The following research works provide the foundational framework for the proposed speech encryption scheme.

1.1 Design and implementation of a real-time AES-256 encrypted audio transmission system on an ARM Cortex-M4 platform.

A significant benchmark in the field was established by Gupta and Chen (2023), who demonstrated the feasibility of implementing AES-256 encryption on an ARM Cortex-M4 platform. Their research proved that high-level encryption standards could be executed on low-power microcontrollers without compromising audio quality or introducing significant latency. This validates the use of the ESP32 in the current project, as it possesses the computational capacity to handle complex cryptographic

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

tasks in real-time.

1.2 A lightweight hybrid cryptography scheme for secure voice communication in IoT devices. Internet of Things and Cyber-Physical Systems. Another study titled "A Lightweight Hybrid Cryptography Scheme for Secure Voice Communication in IoT Devices" by K. Sharma, P. Patel, and R. Kumar (Internet of Things and Cyber-Physical Systems, 2023) explores hybrid encryption methods combining 14 symmetric and asymmetric algorithms for secure IoT-based voice applications. The paper highlights that hybrid methods can enhance key management and strengthen overall data confidentiality. While our project focuses on symmetric AES encryption for simplicity and speed, this study provides insight into future improvements such as key exchange mechanisms and hybrid encryption for multi-node communication systems.

1.3 Efficient hardware–software co-design for voice data encryption on edge devices.

Wang et al. in their paper "Efficient Hardware–Software Co-Design for Voice Data Encryption on Edge Devices" (Microprocessors and Microsystems, 2023) proposed the use of DMA (Direct Memory Access) and buffer management techniques to optimize real time performance in embedded encryption systems. Their results demonstrate that offloading repetitive data transfers to DMA significantly reduces latency and processor load. This concept has been applied in our project to enable simultaneous data sampling, encryption, and transmission, ensuring smooth voice streaming without delays. with phrases like "Welcome"—thereby enhancing the mascot's interactivity and making it more engaging and lifelike.

1.4 Real-Time Secure Voice Transmission for Embedded Systems

Sharma and colleagues demonstrated in their research on "Real-Time Secure Voice Transmission for Embedded Systems" that implementing AES in CTR mode (Counter mode) provides a balanced trade-off between computational efficiency and security. Unlike CBC mode, which requires sequential block processing, CTR mode allows parallel encryption of blocks, enabling faster processing suitable for streaming audio applications. This insight influenced the selection of AES-CTR mode in this project to maintain real time encryption with minimal delay.

1.5 ESP32 for Real-Time Processing

Kumar et al. presented their work on "Real-Time Audio Processing using Embedded Microcontrollers," where they highlighted the capabilities of the ESP32-WROOM-32 in handling high-speed data processing and wireless communication simultaneously. Their research demonstrated that ESP32 can efficiently manage real-time audio streams with minimal delay. This finding supported the selection of ESP32 as the core processing unit in this project.

II.PROPOSED SYSTEM

Developing an end-to-end encrypted link, this system utilizes the ESP32 microcontroller to facilitate secure, real-time wireless voice communication. High-fidelity digital audio is captured via the I2S protocol using an INMP441 microphone and protected with robust AES encryption to ensure data confidentiality. The framework leverages the connectionless ESP-NOW protocol for low-latency, peer-to-peer transmission, effectively bypassing the need for traditional network infrastructure. Final audio reconstruction is managed by a MAX98357A amplifier, while a push-to-talk mechanism optimizes operational power efficiency. This integrated hardware-software approach provides a compact and cost-effective solution for secure communication in defense, industrial, and emergency sectors.

III. METHODOLOGY

Establishing a secure channel for voice data begins with the seamless integration of digital hardware and cryptographic firmware. The process initiates at the transmitter, where acoustic signals are captured by an INMP441 digital microphone and converted into 24-bit audio data via the I2S protocol. This digitized stream is fed into the ESP32 microcontroller, where it is partitioned into discrete packets and subjected to Advanced Encryption Standard (AES) processing. By utilizing the dual-core capabilities of the ESP32, the system ensures that high-speed data acquisition and intensive encryption tasks occur simultaneously, preventing any latency that could disrupt the natural flow of human speech.

Once the audio packets are encrypted, they are dispatched across a 2.4 GHz wireless link using the connectionless ESP-NOW protocol. This method of transmission is specifically chosen to bypass the time-consuming handshakes and network overhead associated with traditional Wi-Fi or Bluetooth protocols. By communicating directly through MAC addresses, the system achieves a peer-to-peer connection that is both high-speed and infrastructure-independent. This localized approach not only ensures low-latency performance essential for real-time conversation but also adds a layer of physical security by keeping the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

data off public networks or centralized routers.

The communication cycle concludes at the receiver unit, where the incoming ciphertext is decrypted using a pre-shared secret key to restore the original digital audio values. These restored signals are then routed through a MAX98357A I2S Class-D amplifier, which converts the digital data back into a high-fidelity analog output to drive a speaker. Throughout this process, a push-to-talk mechanism manages the activation of the radio frequency components, ensuring that the system remains power-efficient and only occupies the wireless spectrum during active speech. This comprehensive hardware-software co-design results in a robust and portable solution for secure communication in sensitive industrial or tactical environments.

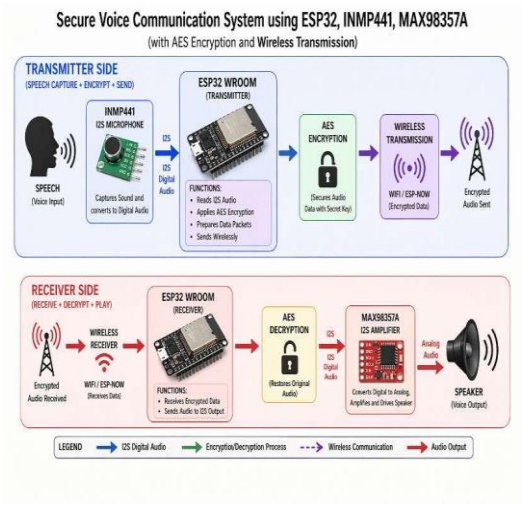


Figure 1: Block Diagram

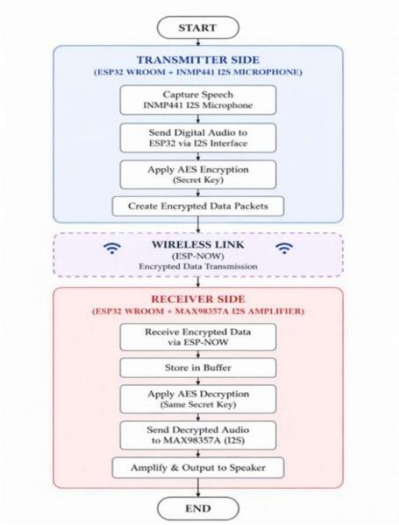


Figure 2: Flow Chart

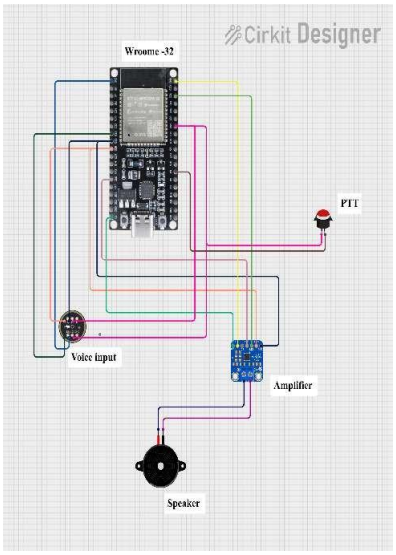


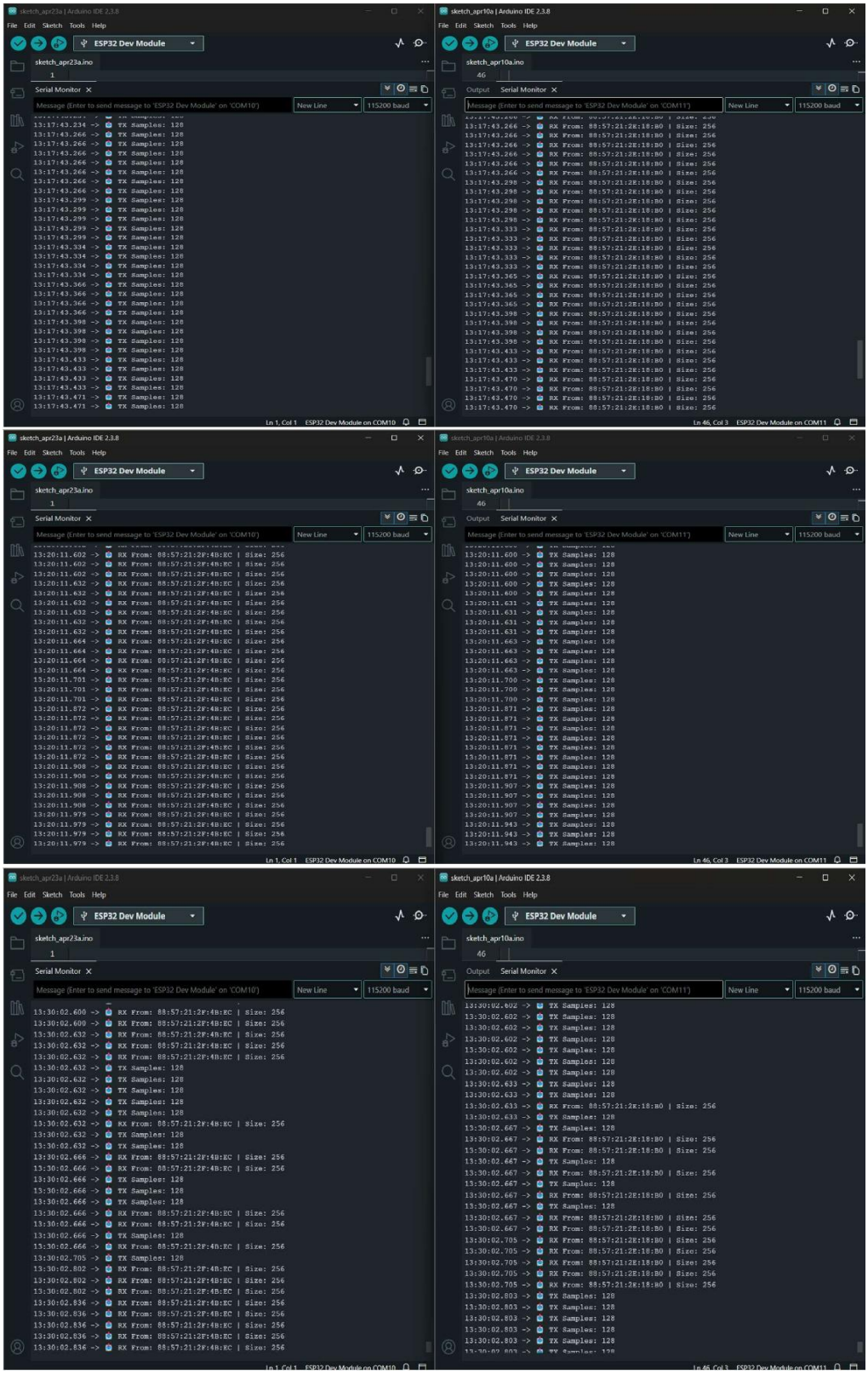
Figure 3: Circuit Diagram

IV. Result And Discussion

Result:

The system employs ESP-NOW, a connectionless communication protocol operating at the Data Link Layer (OSI Layer 2), which eliminates the need for the TCP/IP stack and traditional Wi-Fi association, enabling direct device-to-device communication with latency below 10ms. Reliable and secure point-to-point transmission is achieved through MAC address-based authentication, where each ESP32's unique 48-bit hardware address is predefined to establish a dedicated communication link while filtering out unrelated traffic at the hardware level. To ensure data confidentiality, the system integrates AES-256 encryption in Cipher Block Chaining (CBC) mode, utilizing a 16-byte Initialization Vector (IV) generated from the ESP32's internal True Random Number Generator (TRNG) for each packet, thereby preventing pattern leakage and enhancing security even during repetitive or silent transmissions. The receiver reconstructs the original 16-bit PCM audio using the shared key and transmitted IV. By combining low-latency ESP-NOW communication, hardware-level filtering, and strong cryptographic protection, the system establishes a secure, efficient, and zero-trust wireless communication framework where intercepted data remains computationally infeasible to decode in real time.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026



Discussion:

The secure voice communication system achieves high-quality audio transmission with minimal delay. The use of the I2S protocol improves signal clarity by reducing noise compared to analog communication. The ESP32 efficiently handles AES encryption and decryption, maintaining real-time performance. ESP-NOW ensures stable and reliable wireless communication over moderate distances. The system also demonstrates good power efficiency, especially with push-to-talk functionality. Security is enhanced as encrypted data prevents unauthorized access. Overall, the system provides a low-cost and effective solution for secure voice communication.

V. CONCLUSION

Successful Implementation of Secure Audio: The project successfully demonstrates a functional end-to-end encrypted voice link using the ESP32, proving that

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

high-security communication can be achieved with low-cost embedded hardware.

Real-Time Performance and Low Latency: By utilizing the dual-core processing of the ESP32 and the connectionless ESP-NOW protocol, the system maintains real-time audio transmission with latency levels below the threshold of human perception.

Infrastructure Independence: The system operates as a reliable peer-to-peer network without the need for external Wi-Fi routers or internet connectivity, making it ideal for deployment in remote or tactically sensitive environments.

Optimized Resource Management: Integration of a push-to-talk mechanism and I2S digital interfacing ensures high audio clarity while significantly reducing power consumption for portable, battery-operated use.

VI. FUTURE SCOPE

- Multi-Node Mesh Networking: Transitioning from a point-to-point link to a mesh network architecture would allow multiple users to communicate simultaneously across a wider area, with each device acting as a signal repeater.

- Dynamic Key Exchange Protocols: Implementing asymmetric encryption (like Elliptic Curve Diffie-Hellman) for initial handshaking would allow the system to generate unique session keys for every conversation, enhancing security against long-term key compromise.

- Voice Activity Detection (VAD): Integrating AI-based VAD algorithms could replace the manual push-to-talk button, automatically triggering transmission only when human speech is detected to further optimize bandwidth and power.

- Biometric Authentication: Adding a fingerprint sensor or voice-print recognition to the ESP32 hardware would ensure that only authorized personnel can access the encrypted channel, adding a layer of physical security.

- Advanced Audio Compression: Utilizing lightweight codecs (like Opus or Speex) could compress the digital audio data before encryption, allowing for longer transmission ranges and better performance in environments with high signal interference.

- GPS and Metadata Integration: Future iterations could include the transmission of encrypted metadata alongside voice, such as GPS coordinates or device ID, providing real-time location tracking for search and rescue or tactical teams.

- Solar-Powered Integration: Developing a specialized power management circuit for solar charging would make the device entirely self-sustaining for long-term deployment in remote areas where traditional power sources are unavailable.

VII. REFERENCES

[1] Sharma, K., Patel, P., & Kumar, R. (2023). A lightweight hybrid cryptography scheme for secure voice communication in IoT devices. Internet of Things and Cyber-Physical Systems, 3, 180–191.
[2] Gupta, A., & Chen, S. Y. (2023). Design and implementation of a real-time AES-256 encrypted audio transmission system on an ARM Cortex-M4 platform. 2023 IEEE International Conference on Consumer Electronics (ICCE), 1–5.
[3] Wang, C., Li, H., & Zhang, Z. (2023). Efficient hardware–software co-design for voice data encryption on edge devices. Microprocessors and Microsystems, 96, 104752.
[4] Al-Asadi, M., Ahmed, N. A., & Saad, T. J. (2023). Enhancing security in wireless voice communication using a chaos-based encryption algorithm. Journal of Information Security and Applications, 76, 103525.
[5] Kumar, S., Verma, R., & Singh, A. (2022). Real-time audio processing using embedded microcontrollers. International Journal of Embedded Systems and Applications, 12(3), 45–52.