IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 20 March 2024, accepted 19 May 2024, date of publication 24 May 2024, date of current version 3 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3405317

SURVEY

Secure End-to-End Voice Communication: A Comprehensive Review of Steganography, Modem-Based Cryptography, and Chaotic Cryptography Techniques

ALBERTUS ANUGERAH PEKERTI, ARIF SASONGKO, AND ADI INDRAYANTO
School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Bandung 40132, Indonesia
Corresponding author: Albertus Anugerah Pekerti (33222017@std.stei.itb.ac.id)

ABSTRACT Secure end-to-end voice communication is crucial, but it has several issues. Researchers in this field can benefit from a review paper with broader evaluation parameters than existing literature and from a coverage of security algorithms for voice communication. This paper provides a comprehensive overview of end-to-end secure voice communication and evaluation of various methods for securing voice transmissions over communication networks. The evaluation parameters include System Throughput Capacity, Transmission Error Rate, Recovered Speech Quality, and Security of the System. The analysis results show that Steganography category is the most popular technique in recent years and the most promising category to implement high system throughput capacity. Modem-based Cryptography category is the most promising category to implement standard security algorithms with low system throughput capacity. Meanwhile, chaotic Cryptography provide suitable cryptography for voice security. Furthermore, the results reveal gaps in proposed solutions, including the evaluation and improvement of speech quality, the need for further development of compatible method to implement standard security algorithms, and the extensive development space for Artificial Intelligent based method.

INDEX TERMS Chaotic cryptography, cryptography, end-to-end security, steganography, voice communication.

I. INTRODUCTION

Voice communication plays a significant role in communication due to its audible medium and the ability to convey the speaker's identity and contain the intrinsic information such as emotion, enthusiasm, and health of the speaker. Moreover, voice communication has a real-time feedback characteristic. Moreover, voice communication does not require high data rate and wide bandwidth, thus the information contained is not easily disturbed by noise. Unlike information in data communication which is easily disturbed by noise, because it requires high data rates and wide bandwidth.

Many devices and networks provide voice communication services, such as Landline Telephone, Two-Way Radio, Satellite Phone, Mobile Phone, and LTE Broadband Radios. Despite the rapid rate of expansion of telecommunications networks in recent years, digital services are more widely available in developed countries than in developing countries, especially in large urban centers [1]. Approximately 46.4% of the current world population does not have regular access to the Internet, which represents a total of 3.61 billion people [2]. Global System for Mobile Communications (GSM) channel is the most widely used technology for voice

The associate editor coordinating the review of this manuscript and approving it for publication was Ahmed Almradi.

communication and the only communication facility accessible in most rural areas and underdeveloped countries [3]. Voice communication over limited bandwidth such as, GSM, Radio, Satellite, even Landline Telephone is more reliable, especially in the rural and remote area.

Voice communication is highly susceptible to attacks such as eavesdropping, man-in-the-middle (MitM), and spying malware attacks. An eavesdropping attack is where an attacker can intercept and listen to the conversation without the knowledge or consent of the communicating parties. An MitM attack is where the attacker intercepts communication between two parties and can manipulate or record the conversation. A spying malware (spyware) attack is where a malicious software installed directly on communication devices and collect information from the device, including voice information, without the user's consent or knowledge [4], [5]. Spyware attack causes the communication device to be untrusted, despite high level security of the communication network.

In spite of security measures employed by established networks such as GSM and Voice over Internet Protocol (VoIP), security vulnerabilities still exist. For instance, the A5 algorithm used in the GSM network is susceptible to hacking [6], [7] and the security vulnerabilities as observed in commercial VoIP communications [8]. Furthermore, the current voice security system is not provided in an end-to-end manner [9]. Thus, the user is enforced to trust mobile operators and third party services. The Department of Homeland Security (DHS) in consultation with the National Institute of Standards and Technology (NIST) in 2017, clearly mentioned that the confidentiality or integrity of the communication should not depend solely on the network protection. Moreover, the use of end-to-end encryption for all communications paths is highly recommended [10].

Performing secure end-to-end voice communication is difficult due to limitations such as limited bandwidth, lossy compression, and noisy channels. Standard digital encryption is not suitable for limited channel bandwidth [11] as encryption adds overheads [3]. The human voice range extends from 100 Hz to 17 kHz, but traditional narrowband telephone calls limit audio frequencies from 300 Hz to 3.4 kHz. Voice channels, such as GSM channel, have maximum bandwidth of 4 KHz, which inherently limits data rates [12]. Wideband audio provides relaxed bandwidth limitations and transmits within the audio frequency range of 50 Hz to 7 kHz [13]. A GSM channel is an example of a voice communication network that uses Automatic Repeat Request (ARQ) for error detection and correction within the limited bandwidth of 300 to 3400 Hz [14]. GSM voice channels, featuring audio codec compression, Discontinuous Transmission (DTX), and Voice Activity Detection (VAD), selectively transmit signals with speech characteristics. This poses a significant challenge in the development of secure end-to-end voice communication [9]. Although standard encryption has a higher encryption strength, it is not suitable for limited bandwidth channel such as analog voice [11].

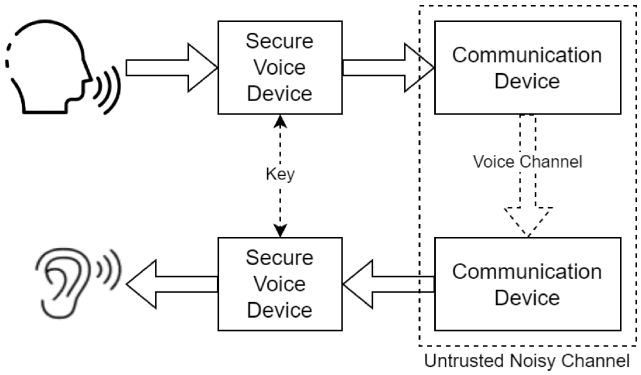

FIGURE 1. The proposed solution approach.

To solve the problem of end-to-end voice communication security and to overcome the limitation, the external secure voice device unit should become a solution approach as shown in Figure 1. Inspired by some seminal ideas from analog communications [15], the security scheme should compatible with a great range of communication networks such as VoIP, cellular, and analog networks [5]. The approach employing lossy encryption aligns with the emerging domain of post-quantum cryptography [16].

In addition, to solve the problem and to overcome the limitation, there are some requirements to be fulfilled. The system should have the throughput capacity 16 kbps or higher. Speech typically requires a high data rate [9]. The robustness against the channel is represented by data error rate and/or by speech quality on the receiver's side. Finally, the high security level should be implemented by the system. Moreover by implementing the standard security algorithms can address this requirement.

The rest of the paper is structured as follows. Section II provides the summary of previous literature review and show the contribution of this literature review paper. Next, Section III includes review methods that are used in this literature review paper. Section IV shows the result by giving the summary narration for each reviewed papers, while Section V presents the analysis for each category. Section VI provides a discussion about the summary of all category and the research opportunities to be developed further by the researchers. Finally, Section VII concludes the paper.

II. PREVIOUS REVIEW WORK ON SECURE VOICE COMMUNICATION

In the literature, there have been some papers reviewing the current state of secure audio, secure communication, and secure voice communication. One of them, Ntantogian et al. [9] published a review paper in 2019 that focuses specifically on security solutions for voice communication, particularly in the context of GSM mobile networks and Voice over IP technology. The paper investigates both research-level and commercial solutions and evaluates them based on three criteria: (i) the level of security provided, (ii)

possible performance issues, and (iii) usability. This paper state that it is the first to categorize and comprehensively evaluate voice encryption schemes for mobile networks. Some of the reviewed papers do not propose solutions for end-to-end voice communication security but rather for end-to-end data communication security. In these cases, the bit-stream does not directly represent voice. The bit-stream is encrypted, modulated into a speech-like signal, then transmitted over the voice channel.

In 2019, Sadkhan et al. [17] conducted a review of previous methods utilizing audio steganography. The study showed that the majority of research employed the Least Significant Bit (LSB) technique to embed confidential messages, in combination with other encryption systems to strengthen the LSB family. Audio steganography was categorized into two types: concealing any type of confidential data in a cover audio file, and concealing audio files within a cover audio file. This paper is limited to discuss about secure voice communication, and the chosen parameters for evaluation are rarely conducted in each reviewed papers.

In 2021, Albahrani et al. [18] conducted a review of audio cryptographic techniques, with a focus on various encryption and decryption techniques based on chaotic maps. The review covers recent contributions to audio encryption, evaluating the algorithms based on security analysis, computational complexity, and quality analysis. It notes that chaotic maps cryptography is a popular technique for securing digital and analog voices. The evaluation in this paper is limited to the level of security and randomness. Moreover this paper does not specifically address encryption algorithms for voice communications.

The most recent review paper is authored by Makhdoom et al. [19] in 2022 provide a coverage of covert communication techniques, including the latest trends, challenges, and future directions. Due to the broad scope of this survey, the discussion on voice communication security is not comprehensive and focused. Furthermore, this paper focuses its coverage to steganography techniques, and does not include cryptography as part of the communication security techniques discussed.

Researchers can take advantages from a new review paper that provide subjective or objective quality assessment of recovered or decrypted voice. The International Telecommunication Union (ITU) has highlighted the importance of evaluating the transmission characteristics of new equipment due to the rapid deployment of digital technologies. According to ITU, in many situations, it is important to determine the subjective effects of a transmission equipment or changes to the transmission characteristics of a telephone network [20]. Moreover, voice quality assessment is a crucial component of both Quality of Service (QoS) and Quality of Experience (QoE).

This review paper provides a comprehensive overview of end-to-end secure voice communication, which involves transmitting a secure voice signal from the sender to the receiver. The paper discusses various methods for securing voice transmissions over communication networks, such as steganography, modem-based cryptography, and chaotic cryptography techniques. It also evaluates these techniques based on broad parameters, including system throughput capacity, transmission error rate, recovered speech quality, and system security. Furthermore, it examines and evaluates the transmission of secure voice signals through simulated models and specific communication networks.

III. REVIEW METHODS

Relevant and recent papers were collected and selected on March 2, 2023, using the search terms ''Secure Voice Communication'' and ''Voice Communication Security''. To obtain a final sample, the inclusion and exclusion criteria is applied. Inclusion criteria included papers that discussed methods of securing voice transmission through specific communication networks or models, were peer-reviewed and accessible to the wider academic community. Exclusion criteria included papers that did not specifically address voice security, papers concerning data security without direct relevance to voice, and papers dealing with Public Key Infrastructure (PKI) secure voice communication. The focus is on symmetric security algorithm developments for secure voice communication. Additionally, literature review papers are excluded to be reviewed. Based on these criteria, a total of 33 proposed solutions were identified across 39 papers.

The papers are classified into three categories: Steganography, Modem-based Cryptography, and Chaotic Cryptography. These three categories have different approaches to improve the voice communication security. The Steganography category comprises papers that propose securing voice communication by concealing the secret voice within the cover voice. The Modem-based Cryptography category consists of papers that propose securing voice communications through the modulation of encryption results to resemble speech, utilizing modem methods such as Codebook Optimization, Optimized Modulation, Parameter Mapping, and Hardware Codec. The Chaotic Cryptography category includes papers that propose securing voice communications by employing mathematical chaos theory in cryptography.

Each proposed solutions are assessed using four parameters: System Throughput Capacity, Transmission Error Rate, Recovered Speech Quality, and the Security of the System. System Throughput Capacity determines the capacity of the method to process data at one time and the performance parameter of the method [9]. Transmission Error Rate and Recovered Speech Quality determine the received speech quality. Even though the transmission error rate does not have a direct correlation to speech quality, increasing the transmission error rate can increase the failure of decryption [9], [14]. Security of the System determines the security level of the proposed method.

The System Throughput Capacity parameter is used to evaluate the proposed solutions by measuring the ability of the system to handle voice signals with higher data rates. Good speech quality typically requires a high data rate [9].

The speech with data rates from 4 kbps to 16 kbps are grouped into 'fair' quality. The speech with data rates above 16 kbps are grouped into 'good' speech quality [21]. A higher system throughput capacity indicates that the system can handle large speech data at one time, while a lower throughput capacity means that the voice signal needs to be compressed to fit the low data rate. System throughput capacity is expressed in terms of samples or data per second.

The Transmission Error Rate is represented by data error rate and/or speech quality on the receiver's side. It is important to consider the ability of a transmitted signal to resist data loss when transmitted over a communication channel. The lower the error rate, the greater the likelihood that the received signal can be accurately reconstructed. The measure of a signal's error rate is represented by Bit Error Rate (BER) or Mean Squared Error (MSE).

In the Received Speech Quality parameter, the fidelity of the recovered speech signal is evaluated in terms of its similarity to the original speech signal. This evaluation is performed by comparing various aspects of the signals such as waveform, spectrum, or subjective evaluation of speech quality by the author. Received Speech Quality is quantified using Signal-to-Noise Ratio (SNR), Perceptual Evaluation of Speech Quality (PESQ), Mean Opinion Score (MOS), or Multiple Stimuli with Hidden Reference and Anchor (MUSHRA).

The Security of the System parameter refers to the ability of the method to implement a particular security algorithm. The security of the system parameter compares the implemented security algorithm, including Advanced Encryption Standard (AES), Triple DES (3DES), Elliptic-curve cryptography (ECC), RC4, Tiny Encryption Algorithm (TEA), as well as novel security algorithms proposed by the authors. The assessment of the security of the system is crucial in ensuring the confidentiality, integrity, and availability of the voice communication. Moreover, by implementing the standard security algorithms, such as AES and 3DES, it can address the security requirement.

In the following sections, a narrative method is used in presenting results analysis of reviewed papers. The narrative method used in this paper summarizes the method, experiment, and result from each papers. The comparisons of each paper are shown in the tables. This approach shows a clear and concise representation of the relevant information obtained from each paper, including the proposed solution, the used evaluation parameters, and the obtained results.

IV. RESULT

A. STEGANOGRAPHY

There are some limitations on the cryptography technique, such as the certainty of the secret information existence and the difficulties of recovering secret information when the signal processing attack or distortion, such as noise addition, compression, cropping, and re-sampling [28], [29]. The steganography technique conceals the existence of the secret information and exhibits good robustness to the signal processing distortion, therefore the secret information can be recovered with minimum error [29].

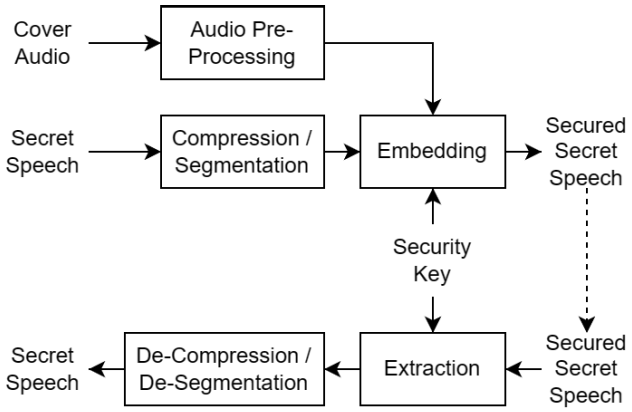

FIGURE 2. Secure voice communication based on steganography technique.

There are a total of eight proposed solutions distributed among ten published papers included in the Steganography category, which can be found in Table 1. Secure voice communication based on steganography technique contain two method to prepare the secret speech for embedding process, that are compression and segmentation, shown in Figure 2. The compression method occur by compressing the secret speech to reduce hiding capacity. The segmentation method occur by segmenting the secret speech into the same segmentation type as the cover audio. The compressed and/or segmented secret speech is then embed to processed cover audio accordingly.

1) DENG et al. [22], [23] IN 2006 AND 2007

In their papers, they proposed a real-time secure communication system based on an information hiding algorithm that adaptively selects embedding locations and employs multi-nary modulation and speech recognition to reduce the bit-rate of secret speech.

a: METHOD

The system is using the the compression method from the Figure 2, by compressing the secret speech using dynamic time warping (DTW) recognition algorithm, followed by manually corrected voice recognition. The secret information is then encoded and embedded in the host speech through Discrete Fourier Transform (DFT) coefficient modification. The proposed algorithm adaptively selects the DFT coefficient corresponding to either the first or second formant of each sonant frame based on an encryption security key. At the receivers, a parallel strategy is implemented to identify potential embedding positions and extract hidden data for displaying them on a screen.

b: EXPERIMENT ENVIRONMENT

In the experiment, the host speech and secret information were both recorded at a sampling rate of 8 kHz, 16 bits per

TABLE 1. Secure voice communication based on steganography.

Solutions	Year	Throughput Cap.	Error Rate	Recovered Speech Quality	Security of The System
Deng, Yang, and Deng [22] [23]	2006, 2007	8 kHz, 16 bit	-	SNR: 47.6625 dB	Information Hiding
Xu, Yang, and Shao [24]	2009	8 kHz, 16 bit	BER: 9%	PESQ: 3.616	Information Hiding based on Scalar Costa Scheme
Qi, Longmei, and jinfu [25]	2018	8 kHz, 16 bit	-	PESQ: 1.3393	Sound Masking Utilize Speech Corpus Selected Frames
Ridha, Jawad, and Kadhim [26]	2018	44.1 kHz	-	FwSNR: 22.44 dB, DMOS: 4.1	Blind Source Separation
Pushpalatha, Ramesh, and Raganna [27]	2019	22.05 kHz, 16 bit	-	The original data can be decrypted without much loss	AES-128 with Watermark Speech Signal
Bharti, Gupta, and Agarwal [28]	2019	16 kHz, 16 bit	-	PESQ: 1.24, MOS: 2.1, The content of the revealed secret audio was understandable	16 kbps Steganography Hiding Capacity
Kumar and Kanhe [30] [29]	2020, 2022	8 kHz, 16-bit	-	PESQ: 3.78	Embed Serial Value (SV) of secret speech & Chaotic Map for Random Embedding
Abdallah and Meshoul [31]	2022	-	-	SNR: -17.39 dB	DWT Combining Signal & Chaotic Baker Map

sample. The performance of the system was evaluated under both non-attacking conditions and PSTN interferences.

c: RESULT

The results showed that it has good robustness against attacks, such as A/U-law Pulse-Code Modulation (PCM) coding/decoding, low-pass filtering, and white noise. Moreover, the subjective test results showed that the watermarked speech is imperceptible to most listeners. The system achieved a SNR preponderance of over 47 dB, and the extracted secret information accuracy remained at 100% after the mixed speech was processed under the attack of A/U-law PCM coding/decoding.

2) XU et al. [24] IN 2009

The paper proposes a novel scheme for speech secure communication that combines information hiding and Compressed Sensing (CS) [32], [33].

a: METHOD

The CS used by this paper to compress the secret speech, indicate that this paper using the compression method as in the Figure 2. The information embedding process includes four steps, which involve encoding the secret speech into a bit stream using a prescribed sensing matrix, calculating embedding transform coefficients, embedding the secret bits based on Scalar Costa Scheme [34], and obtaining the mixed speech. The information extraction process consists of three steps, which involve calculating the coefficients with secret information, extracting the secret bits from the coefficients, and recovering the secret speech via CS decoder.

b: EXPERIMENT ENVIRONMENT

The experiment was conducted on 8 kHz sampled and 16-bit quantized testing speeches with a frame length of 128 samples per frame for both secret and public speeches. The performance of the system was evaluated using two objective indices, ITU-T P.862 perceptual evaluation of speech quality (PESQ) and Segmental Signal-to-Noise Ratio (SNRseg).

c: RESULT

The results revealed that the mixed speech quality achieved a SNRseg of 32.342 and P.862 MOS of 3.616 using the Discrete Cosine Transform (DCT) + Lifting Wavelet Transform (LWT) algorithm. Moreover, the proposed system was found to be robust against various attacks, including A/U-law PCM Coding/Decoding, Additive White Gauss Noise, and Low Pass Filtering. Specifically, the BER for these attacks remained at 0%, below 9%, and below 2.5%, respectively, while the normalized correlation (NC) coefficient remained at 100%, above 91%, and above 97%, respectively.

3) QI et al. [25] IN 2018

In this paper, a new speech privacy protection method is proposed based on sound masking and a speech corpus.

a: METHOD

The proposed method is using the the segmentation method, that is shown in Figure 2, by segmenting input speech as frames and estimates the pitch of each frame. The pitch period and a secret key are used as an index to find the required frames from the speech corpus. Linear additive operations are used to limit masking and maintain pitch consistency. Masked speech is then transmitted and recovered at the receiving end by selecting the corresponding masking frames from the speech corpus using the same index as the transmitting end. The algorithm is tolerant of pitch errors that enhances recovered speech quality.

b: EXPERIMENT ENVIRONMENT

The experiments were conducted on speech signals from 50 Chinese speakers. These speech signals are resampled to a rate of 8000 Hz, and linearly quantized to 16 bits for each sample. These speech signals is compressed by ADPCM

and G.728 LDCELP in order to assess the robustness of the proposed method against speech compression.

c: RESULT

The experimental results show the objective measures of masked signals are PESQ 1.1096, ISD 4.1837, EMBSD 6.2017 and of recovered speech signals are PESQ 3.3294, ISD 0.5223, EMBSD 1.0047. The objective measures of recovered speech signals after the ADPCM codecs is PESQ 3.3300, ISD 0.5224, EMBSD 1.0039 and after G.728 LDCELP is PESQ 1.3393, ISD 3.6551, EMBSD 6.1959. Experiment results show the proposed method has good privacy protection and recovered speech quality. It exhibits robustness against speech compression by waveform coding algorithms and offers benefits to systems utilizing parametric coding algorithms for compression.

4) RIDHA et al. [26] IN 2018

This paper proposed a secured time domain cryptosystem using a modified version of Blind Source Separation (BSS) algorithms for mobile voice communication. The system is designed to provide complete security without modify the existing mobile network infrastructure.

a: METHOD

The encryption process starts by segmenting the original speech signal into R segments and generating R independent pseudorandom key signals. This proposed method is using the segmentation method as in the Figure 2. The encrypted signal is produced by linearly combining speech and key signal segments using a mixing matrix. To limit the bandwidth occupied by the encrypted speech signal, the proposed system uses a modified key generation process.

b: EXPERIMENT ENVIRONMENT

The system was implemented using MATLAB and assessed in real-time conditions. The evaluation was based on three criteria, including bandwidth, speech quality, and residual intelligibility as a measure of security. The experiment utilized speech signals sampled at a rate of 44,100 Hz. The system's performance was compared with other BSS-based systems using the Frequency-Weighted Signal-to-Noise Ratio (fwSNR) and the Degradation Mean Opinion Score (DMOS).

c: RESULT

The fwSNR indices of recovered speech signals calculated for both filtered and non-filtered source signals lie in the range 22.44 - 27.11 dB, which is better than the traditional BSS cryptosystem. The DMOS score for the encrypted signals is 1, and for the recovered signals, it is 4.1, indicating that residual intelligibility in the encrypted signal is low. Running on a an Intel® CoreTM i7 processor with 64-bit Windows 7 operating system personal computers (PC), the encryption of each frame consumed 0.4484 seconds, while decryption required 1.1372 seconds.

5) PUSHPALATHA et al. [27] IN 2019

In this study, an integrated approach to compress, watermark, and encrypt speech signals was proposed for authentication, ownership verification, and security purposes.

a: METHOD

DCT compression used by this paper to reduce data size, indicate that this paper is using the compression method shown in the Figure 2. A spatial domain watermarking technique was used for embedding text data as watermark. AES 128 bits was performed to provide security for watermarked speech signals.

b: EXPERIMENT ENVIRONMENT

The proposed system was implemented in MATLAB with a speech signal of 22,050 Hz, 63,039 total samples, 2.8589 seconds duration, and 16 bit/sample.

c: Result

The results showed that as the compression factor increased, the original signal experienced lossy compression while maintaining the same level of encryption, and the original data could be decrypted without much loss, providing security for the speech data.

6) BHARTI et al. [28] IN 2019

This study proposes a method for generating stego audio by embedding secret audio into cover audio in three steps, segmentation, embedding, and audio generation.

a: METHOD

Firstly, by using the segmentation method as in Figure 2, the cover and secret audio are segmented into two streams containing sign and amplitude information. In the second step, the amplitudes and the sign bits of the secret audio are embedded into the modified cover audio. Finally, stego audio is generated by converting the amplitude stream bits and sign bits into their original forms and multiplying the modified cover audio amplitudes with their respective signs. This method can be utilized for real-time audio transmission with minimal buffering requirements. The secret audio can be extracted by recovering the amplitudes and sign bits, which can then be converted into their original forms and multiplied together to reveal the secret audio.

b: EXPERIMENT ENVIRONMENT

In this experiment, the robustness of steganographic audio transmission was evaluated against attacks such as LSB, resampling, and additive white Gaussian noise (AWGN). The proposed algorithms were used to reveal the secret audio. Hiding capacity was measured as the number of bits that can be hidden into one second of cover audio. Hiding type was classified as deterministic or non-deterministic. Non-deterministic hiding was found to be more secure than deterministic hiding.

c: RESULT

The proposed approach was found to be robust towards attacks as the content of the recovered secret audio was understandable. The hiding capacity of the proposed approach was observed to be maximum when assuming a sampling frequency of 16 kHz. Finally, the experiment concluded that the proposed approach used non-deterministic hiding to ensure higher security.

7) KUMAR AND KANHE [29], [30] in 2020 and 2022

In their papers, they proposed a secure watermarking algorithm that embeds a secret speech signal into a cover audio signal using discrete wavelet transform (DWT) and chaotic mapping.

a: METHOD

The embedding process is divided into four steps. In the first step, the cover audio signal is divided into non-overlapping frames, and the secret speech signal is divided into frames that undergo 1-D DCT. The using of DCT to compress secret speech is an indication that this paper is using the compression method shown in the Figure 2. The resulting DCT coefficients are further divided into segments, and cover audio frames are selected chaotically using random integers generated by a logistic chaotic map. In the second step, second-level DWT is performed on each cover audio frame, and the resulting approximate coefficients are divided into segments. In the third step, singular value decomposition (SVD) is performed on the coefficient matrix to embed the data in the singular matrix. In the fourth step, the DCT coefficients of the secret speech frames are arranged in a matrix, and the watermark is embedded into the singular matrix using a scaling parameter. Finally, another round of the SVD operation is performed to obtain extraction matrices. To optimize the trade-off between imperceptibility and robustness, the scaling parameter is chosen empirically.

b: EXPERIMENT ENVIRONMENT

The proposed watermarking algorithm was tested on the USAC database [35], which contains five music files sampled at 48 kHz of 16 bits, and the NOIZEUS database [36], which contains 30 speech signals sampled at 8 kHz of 16 bits. The algorithm's evaluation involves conducting 150 tests to assess both imperceptibility and robustness. In each test, a secret speech signal is embedded into each cover audio.

c: RESULT

Results show that the algorithm achieved good imperceptibility with an average SNR of 46 dB and ODG of -1.07, respectively. The algorithm is also robust against various signal processing attacks, reconstructing the secret speech signal with an average NCC of 0.95 and an average PESQ score of 4.26 under no attack condition and PESQ score of above 3.0 under various signal processing attacks while preserving the perceptual quality of the reconstructed speech signal.

8) ABDALLAH AND MESHOUL [31] IN 2022

This study proposed a multilayer cryptosystem to encrypt audio communications. The proposed system utilizes three layers of encryption: fusion, substitution, and permutation, in the frequency domain.

a: METHOD

In the fusion layer, either random projection or salting is employed to combine two 2D signals obtained from the cover audio and secret speech input signals prior to DWT processing. The DWT segmentation between the secret speech and the cover audio are combined. This indicates that this study is using the segmentation method as in Figure 2. The resulting signal is then passed to the substitution layer, which uses either DCT or discrete sine transform (DST) techniques to perform a substitution operation on the signal. Next, the resulting signal is passed to the permutation layer, which uses the chaotic baker map to perform a permutation operation. Finally, an inverse DCT or inverse DST is performed to obtain the encrypted signal.

b: EXPERIMENT ENVIRONMENT

The proposed cryptosystems were evaluated in a simulated environment using MATLAB, and various measurements were taken into account during the performance analysis process.

c: RESULT

The results show that the salting-based encryption method and the multilayer DCT/DST cryptosystem offer better levels of security with SNR values of -25 dB and -2.5 dB, respectively. The random projection based on 1D DWT and the three-phase encryption cryptosystem demonstrated the best results, indicating their high resistance to noise attacks. The proposed audio encryption methods were shown to be highly effective, with significant discrepancies between the original and encrypted signals, as evidenced by the values of SNR, SNRseg, Log-Likelihood Ratio (LLR), Spectral Distortion (SD), Correlation Coefficient (r), and Structural Similarity Index Measure (SSIM).

B. MODEM-BASED CRYPTOGRAPHY

The Modem-based Cryptography category consists of papers that propose securing voice communications through the modulation of encryption results to resemble speech, utilizing modem methods such as Codebook Optimization, Optimized Modulation, and Parameter Mapping methods. The list of these papers and methods can be found in Table 2.

Modem-based Cryptography technique consist of encoder, encryption, error correction, and modulator modules. The encoder module is used to encode the secret speech signal into a compressed signal or map the secret speech signal into speech parameters. The encoded signal is then encrypted by

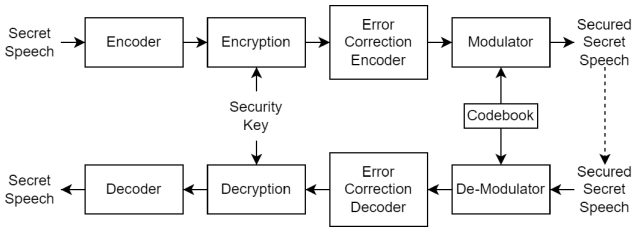

FIGURE 3. Secure voice communication using modem-based cryptography technique.

encryption module using the secret key. To ensure the data integrity of the encrypted signal, error correction is applied. The modulator module works as speech synthesizer to convert the encrypted signal into speech-like signal as secured secret speech, based on certain codebook. The codebook can be a finite alphabet of predefined symbols or waveforms, a well-known digital modulation form, or a speech parameters of commonly used speech models. Then the secured secret speech is transmitted. Through the demodulator, error correction, decryption, and decoder modules, the secured secret speech recovered into the original secret speech. This scheme is shown in Figure 3.

1) CODEBOOK OPTIMIZATION

Codebook Optimization is a technique that encodes data through speech codecs into a finite alphabet of predefined symbols or waveforms optimized for transmission through the voice codec. This method is featured in three published papers, which present three proposed solutions utilizing the Codebook Optimization method.

a: ABRO et al. [3], [55] IN 2019

A new framework for secure end-to-end voice communication was proposed by using Linear Predictive Coding (LPC) compression as encoder module, AES-256 as encryption module, and the single codebook-based method as a modulator module, as shown in Figure 3.

Method: The speech signal sampled at 8 kHz of 13 bits are first compressed using 10th order LPC analysis. The resulting digital data is encrypted using AES 256 bits with 128 bits block size, and Elliptic Curve 25519 (EC25519) with Diffie-Hellmann (DH) key exchange protocol is used to exchange the secret key. The encrypted data is then converted to a speech-like waveform using a modulator that employs a codebook trained on human voice using Linde-Buzo-Gray (LBG) algorithm [56], [57] with a total of 1024 codewords. The transmitter encodes the bits through the codebook, while the receiver selects the best possible index to recover the encoded signal.

Experiment environment: The system was implemented on two Windows 7 Operating system PCs, each equipped with an Intel Core i3-8100 processor and 8 GB RAM. Two GSM mobile phones were connected to the PCs using a hands-free cable, and a real-time GSM-to-GSM call was established via the PCs.

Result: The proposed method achieved a data rate of 2 kbps and 0% bit error rate, indicating its suitability for real-time applications involving the transmission of encrypted voice over the GSM network.

b: KRASNOWSKI et al. [37] IN 2022

In their recent paper, they introduced a novel Data over Voice technique for sending encrypted data or speech as pseudo-speech in the audio domain over existing voice communication infrastructures, like 3G cellular network and VoIP. speech compression, Codec2, is used as encoder module, AES as encryption module, Reed-Solomon codes as error correction module, and codebook-based method as modulator module. This paper employ all modules from the Figure 3.

Method: The proposed system involves speech encoding, encryption, error correction, and data modulation blocks. The input and output signals of the processing chain are analog, but all internal processing is performed digitally. The scheme uses low-bitrate speech compression by Codec2 [58], [59], followed by encryption with AES in the counter mode of operation, and protection against channel errors by shortened Reed-Solomon (RS) codes with erasures and 6-bit symbols. The modulation block generates a discrete-time audio signal that is played to the digital audio input of a voice channel.

Experiment environment: The system was implemented and tested in GNU Octave environment with real voice calls up to 6.4 kbps over VoIP voice channels using 4G wireless network and 2.4 kbps over 3G cellular calls.

Result: The proposed system enables safe voice transmission with an effective binary error rate significantly below 1% and can correctly decode the signal as long as the harmonic structure is preserved.

c: CUBRILOVIC et al. [38] IN 2022

The authors proposed a secure voice transmission system based on speech-like waveform (symbols) codebook modulation technique over different voice communication channels. The speech signal is encoded by compression method, then encrypt the compressed signal, apply the Forward Error Correction code, and modulate it into a ''speechlike'' sequence using a codebook. All modules from the Figure 3 is employed.

Method: The transmitter's operation mode consists of several steps. First, the microphone-recorded signal is compressed and digitalized with a low bit-rate speech codec. Next, the digital data is encrypted and protected from noise with a Forward Error Correction (FEC) code. Then, the resulting bit stream is modulated into a ''speechlike'' sequence using a codebook, with a synchronization sequence of shaped pseudo-noise sequences added for demodulation. Offline, pre-trained symbol waveforms are collected, with each symbol having its own cluster. These waveforms are collected by repeating each symbol several times with different predecessors, and experiments are carried out on different

TABLE 2. Secure voice communication modem-based cryptography.

Codebook Optimization					
Solutions	**Year**	**Throughput Cap.**	**Error Rate**	**Recovered Speech Quality**	**Security of The System**
Abro, Rauf, Mobeen-ur-Rehman, *et al.* [3]	2019	8 kHz, 13 bit	BER: 0%	-	AES-256 ECDH (EC25519)
Krasnowski, Lebrun, and Martin [37]	2022	6.4 kbps over VoIP	BER: $< 1\%$	Harmonic structure of the signal is preserved	AES-256 (CTR Mode)
Cubrilovic, Mandic, and Krstic [38]	2022	1.6 kbps	BER: 0.12%	-	AES-256 or similar
Optimized Modulation					
Solutions	**Year**	**Throughput Cap.**	**Error Rate**	**Recovered Speech Quality**	**Security of The System**
Tseng and Chiu [39]	2007	8 kHz, 8 bits	-	The descrambled speech is very similar to the original speech	93 factorials possible scrambling keys OFDM Permutation
Andrade, Campos, and Apolinario [40]	2008	8 kHz, 16 bits	-	PESQ: 2.3-2.9	13,000 analog permutation
Islam, Ajmal, Ali, *et al.* [41]	2009	-	-	The decryption block successfully recovers the original signal back	256-QAM Binary-Coded
Chen and Guo [42]	2011	2.4 kbps	BER: 0.12%	Original speech has been recovered successfully	-
Biancucci, Claudi, and Dragoni [43]	2013	8 kHz, 13 bits	BER: 7.1%	-	AES-256 128 bit block cipher or an SSSC block
Sheikh, Akhtar, Parah, *et al.* [44]	2017	-	Haar BER: 0.81%, Db10 BER: 0.65%	The speech signals compressed with Db10 wavelet have a better sound quality than the one compressed with Haar wavelet	Gold Code Multilayer Encryption
Rehman, Adnan, Batool, *et al.* [14]	2021	0.7 kbps	BER: 0.00%	Speech quality greatly impacted by the peculiarities of the service provider and physical location	FPE Algorithm
Parameter Mapping					
Solutions	**Year**	**Throughput Cap.**	**Error Rate**	**Recovered Speech Quality**	**Security of The System**
Ozkan and Berna Ors [46] [45]	2011, 2015	1.6 kbps	BER: 0.00%	-	AES-128 ECB mode
Yarman, Ulger, and Aslan [48] [47]	2017	16 kbps	-	MOS: 82%	AES-256
Chang and Woźniak [49]	2020	-	Loss data rate: 0.33%	Recovered Speech Quality: Good	168-bit key 3Des-ECC
Krasnowski, Lebrun, and Martin [5]	2022	8 kHz	RMSE: 1493.30 for energy, 525.70 for pitch, 0.12 for spectral envelope parameter	MUSHRA MOS: 49%	256 bits key PRNG
Hardware Codec					
Solutions	**Year**	**Throughput Cap.**	**Error Rate**	**Recovered Speech Quality**	**Security of The System**
Chumchu, Phayak, and Dokpikul [50]	2012	2 kbps	-	Voice communication is acceptable performance	RC4
Boruchinkin [51]	2015	-	-	R-factor: 83-93	AES & RSA
Chouhan and Singh [52]	2015	48 kHz, 16 bits	-	The decryption block successfully recovers the original signal back	Tiny Encryption algorithm (TEA) 128-bit key
Lin, Yang, Duanmu, *et al.* [53]	2015	-	-	SNR: 15.1 dB	Four Stage Multilayer Feedforward Neural Network for Neural Cryptography
Mondal and Sharma [54]	2019	15 kHz, 32 bits	-	Superb voice quality	AES-128
Ge, Sun, Zheng, *et al.* [11]	2021	48 kHz, 16 bits	-	Subjective evaluation: 80% voice quality level with noise	12,960 scrambling encryption mode

communication channels. Next, the waveforms are clustered in clusters using k-nearest neighbor algorithm for each of 2^p symbols in every communication setup, and the codebooks trained in this way are added to the device's database.

Experiment environment: The proposed solution was tested in real-time, and its applicability was extended to VoIP services. The audio connection was established via a 3.5 mm audio cable, Bluetooth, or the Inter-IC Sound (I2S) interface in scenarios where a GSM module was integrated into the encryption device. The system was tested on various voice communication channels, including GSM 3G, GSM Voice over Long-Term Evolution (VoLTE), WhatsApp Voice Call, and Google Meet.

Result: The results of the experiments yielded a range of Symbol Error Rate (SER) between 0.25% to 9.11% and BER between 0.12% to 6.8%.

2) OPTIMIZED MODULATION

The Optimized Modulation method refers to a method based on well-known digital modulation techniques. The primary

objective of this technique is to optimize the modulation parameters involved to ensure reliable communication over the voice codec. A total of seven proposed solutions are available, which have been published in seven scientific papers belonging to this method.

a: TSENG AND CHIU [39] IN 2007

They proposed a speech-scrambling technique to prevent residual intelligibility from scrambled speech. The method involved a combination of an appropriate QAM mapping method and an Orthogonal Frequency Division Multiplexing (OFDM) scheme. The QAM scheme is used for parameter mapping encoder, then the encryption mmodule is using permutation, finally the Inverse Fast Fourier Transform (IFFT) is used to transform a time domain signal. The modules of encoder, encryption, and modulator are used as shown in Figure 3.

Method: The proposed OFDM speech scrambler converts an original speech signal (PCM format at 8 bits per sample) into a binary data stream, which is then mapped into complex-valued symbols using a 64-QAM scheme. The frequency components are permuted by a permutation unit and transformed into time-domain signals using the IFFT unit. To ensure no bandwidth expansion, only the 93 frequency components corresponding to subcarriers numbered 12 to 104 are permuted, resulting in a frequency range of 375 - 3250 Hz, which is within the bandwidth of the original speech signal. The scrambling key generated by a permutation has a key space of 93 factorials. A seed is required to be securely transmitted over a communication channel for the generation of scrambling keys using the scrambling key generator on the receiver side.

Experiment environment: The channel noise was additive white Gaussian noise with an SNR of 25 dB, and the synchronization between the transmitter and receiver was aligned.

Result: The scrambling key generated had 93 factorials possible scrambling keys, making brute force attack on the system impractical. The scrambled speech was similar to white noise, indicating that no residual intelligibility. Both formant and pitch information were completely wiped out in the scrambled speech. After descrambling, the resulting speech was very similar to the original speech.

b: DE ANDRADE et al. [40] IN 2008

This study examines the effectiveness of different voice scrambling techniques for mobile communication vocoders. This paper use the encoder, encryption, and modulator modul based on Figure 3. This paper compares two different encoder, DCT TDS and UDFT FDS, then scrambling as encryption, and transform back to time domain as the modulator.

Method: The Bi-Dimensional Scrambler and Transform-Domain Scrambler are investigated. Around 13,000 out of the total 40,320 (8 factorial) possibilities of subbands permutation to produce scrambled speech signals with low residual intelligibility.

Experiment environment: The experiment tested two scrambling schemes, DCT TDS and UDFT FDS, using 8 kHz signals with 16-bit precision and a frame length of 20 ms. The experiment was conducted in Brazilian Portuguese.

Result: Results showed that scrambled speech after the AMR codec had PESQ scores ranging from 2.3 to 2.9 for FDS and a bit rate from 4.75 kbps to 12.2 kbps. While the PESQ scores for scrambled speech were lower than for clear speech, descrambled signals were still intelligible, indicating that FDS could be used for low-cost voice privacy mobile phones.

c: ISLAM et al. [41] IN 2009

They proposed a real-time end-to-end secure communication system based on QAM modulation developed in MATLAB Simulink. According to Figure 3, this paper utilized the quantization encoder, scrambling module, and QAM modulation module.

Method: The speech input is 8-bit quantized and modulated with Binary-coded QAM before being demodulated via a user-defined QAM scheme to create a highly randomized signal that still retains a speech-like waveform.

Experiment environment: The hardware platform includes two personal computers, each with two sound playback and recording devices, connected to headphones and a Nokia 1100 GSM handset. The MATLAB 7.4.0.287 (R2007a) Simulink 6.6 is used for real-time simulation.

Result: The encryption algorithm completely distorts the human speech, but the decryption block successfully recovers the original signal. The encrypted speech is incomprehensible and conveys nothing without decryption, ensuring a secure and protected GSM channel.

d: CHEN AND GUO [42] IN 2011

This paper introduced an OFDM-based method for secure data communication by modulating encrypted data onto speech-like waveforms. The method utilized the digitzation encoder, encryption module, and OFDM modulator accordance to Figure 3.

Method: The proposed scheme utilized orthogonal subcarriers for parallel transmission and Fast Fourier Transform (FFT) for computational complexity reduction. Synchronization was achieved by using a predefined speech sequence, which can be used for secure voice transmission.

Experiment environment: The proposed scheme was evaluated on a GSM-to-GSM connection and a sample Chinese voice.

Result: Results showed that the proposed scheme can achieve a throughput capacity of 2.4 kbps with a BER of 0.12% in the GSM-to-GSM connection, indicating its suitability for real-time secure data communications. The recovered speech was very similar to the original speech, indicating the proposed scheme's effectiveness in recovering the original speech.

e: BIANCUCCI et al. [43] IN 2013

The proposed device enables secure data and voice transmission over GSM voice channel network.

Method: The input speech signal is first compressed and then encrypted by an AES-256 128 bit block cipher or a Self-Synchronizing Stream Cipher (SSSC) block before being converted into a suitable waveform for GSM codec and network requirements by the FM-based data modulator. As shown in Figure 3, encoder module, encryption module, and modulator module is used in this method.

Experiment environment: The input signal was sampled at 8 kHz with 13 bit resolution. The device was implemented using a pair of x86 PCs.

Result: The results showed that the modulation algorithm could bypass compression errors introduced by GSM compression. The BER after demodulation was 0.068%, and after decryption, it was 7.1%.

f: SHEIKH et al. [44] IN 2017

This paper proposed a system for secured digital data compression and modulation for robust data transmission over GSM voice channels.

Method: The proposed system comprises several stages according to Figure 3. The input speech signal is generated from text-to-speech conversion. Wavelet compression as the encoder module, encoding with Gold Code [60] as the encryption module, and Quadrature Phase Shift Keying (QPSK) modulation as the modulator module. Then the secured speech signal is transmitted. At the receiving end, the signal is first decoded, demodulated, and decompressed using inverse wavelet transform.

Experiment environment: The experiment involved converting a desired text into speech using a text-to-speech conversion system in MATLAB. The speech signal was then encoded with a gold code sequence, modulated using QPSK modulator and sent over the AWGN and GSM channel. At the receiving end, the signal was decoded, demodulated, and decompressed using inverse wavelet.

Result: An adaptive FIR filter was used to avoid noise caused by the AWGN channel. Among the four compression and denoising techniques tested, Db10 with hard thresholding gave the best results with 98% retained signal energy, 87 peak signal-to-noise ratio and less bit error rate. Coded speech signals had higher PSNR and lower BER than those without coding. The secured compressed speech signals encoded with gold code sequence had less BER than those without coding. Db10 wavelet with hard thresholding provided better sound quality than Haar wavelet.

g: REHMAN et al. [14] IN 2021

The study proposes a secure end-to-end voice communication system over GSM using codec2 700 bps compression [61] as the encoder module. The encryuption module is using the Format Preserving Encryption (FPE) algorithm. Then coherent PSK as a modulator module. This scheme corresponds to Figure 3.

Method: The proposed flow involves human speech as input that undergoes specific voice processing and goes to the encoder block. Codec2 compresses the speech and converts the analog data to digital data, which is then encrypted using the FPE algorithm. The encrypted data is then further processed to generate a modulated signal using a modulation scheme that utilizes Frequency division multiplex (FDM) and QPSK. The modulated signal is transmitted over the GSM voice channel and is demodulated on the receiver side. Decryption is carried out on the demodulated bits, and the decrypted bits undergo the process of the decoder, which converts the bits to an analog signal and decompresses the data. The output analog signal undergoes post-processing and leads to regenerated input human speech.

Experiment environment: The proposed system was evaluated for performance and speech quality on two different platforms: computer-based and embedded systems, in a real-time GSM communication scenario. The i7 computer with 12 GB RAM was used for the computer-based system, while the NVIDIA Jetson Tx2 is used for the embedded system. NISQA v0.4, which was a speech quality assessment tool proposed by Mittag et.al., was used to evaluate the speech quality of the system.

Result: The study emphasizes that the effectiveness of any secure voice communication technique over GSM networks could be significantly influenced by the specific characteristics of the service provider and the physical locations of both the transmitter and receiver. The proposed methodology has shown viable results on both the computer-based system and the embedded system.

3) PARAMETER MAPPING

The Parameter Mapping method involves mapping the encrypted bit stream onto speech parameters of commonly used speech models, which are then utilized to synthesize an audio signal with speech-like characteristics. Six published papers within this method presented four proposed solutions.

a: OZKAN AND ORS IN 2015 [45] AND 2011 [46]

This paper proposed a secure voice communication system based on a modem that transmits data over GSM voice channel. The system is using digitization encoder, encryption, and modulator module as shown in Figure 3.

Method: The system first digitizes the speech signal and encrypts the bit streams using Electronic Code Book (ECB) mode 128 bit AES implemented on an FPGA board. Then, the encrypted bit streams are coded as synthetic speech signals using a codec module that uses an LPC model to synthesize speech-like signals, which are coded by the GSM Full Rate (GSM-FR) codec. Communication bit streams indicate index numbers of speech parameter codebooks including energy, vocal filter coefficients, and pitch as the modulator module. In the 2015 research, a novel approach to produce an artificial database is created using LPC from the constraints of GSM

instead of filtering a speech database, which overcomes the challenge of producing a database consisting of stable sets of LPC parameters by using Line Spectrum Frequencies (LSF) instead of LPC parameters.

Experiment environment: The system was tested in a computer environment using the 13 kbps GSM-FR low bit rate vocoder 06.10 source code and the SOX tool without the line effects.

Result: The experiment resulted in achieving a 1.6 kbps simulation data rate and zero BER for wireless communication errors.

b: YARMAN et al. [47] IN 2017

The study proposes a secure VoIP system, corresponding to Figure 3, by encoding voice using the SYMPES coding technique [48], [62] as the encoder module and the encryption module using an open standard encryption algorithm.

Method: The SYMPES coding technique involves creating a database of sample voice recordings for each speaker, encoding and encrypting the voice in the transmitter end, transmitting the encrypted data over an IP network, and then decoding and decrypting the voice in the receiver end. The database of sample voice recordings is used to generate an index for each frame of voice data, which is then encrypted and transmitted to the receiver. The receiver uses the index to decode and decrypt the voice data.

Experiment environment: The experiment was conducted using a SYMPES terminal with an interface software application, VoIP software, VPN user software, and a speech database encrypted with a 256-bit AES encryption algorithm. Real-time audio coding and decoding using SYMPES is planned to be carried out using the graphics processing unit (GPU) and the main processor (CPU) with heterogeneous processing logic.

Result: The quality of the SYMPES encoded voice was evaluated using the MOS on a limited user population, achieving 4.1 for 16 kbps SYMPES.

c: CHANG AND WOŹNIAK [49] IN 2020

As shown in Figure 3, this paper proposed a method for encrypting voice transmission using Multiband-Excitation (MBE) model as the encoder and modulator module. The encryption module is using a combination of 3DES and ECC algorithms.

Method: The method involves changing the voice signal through a low-pass filtering method, synthesizing the voice through A/D conversion, and splitting the voice frame to encrypt it. The output speech is the inverse process of encryption, where the speech frames are decrypted, encapsulated, synthesized, and then amplified through D/A conversion to obtain higher quality speech signal. The voice collector uses a modem as the communication component and a telephone line to connect the public telephone exchange network to transmit data. The MBE model divides the frequency spectrum of a frame speech into several harmonic bands to judge the voiced or unvoiced of each band respectively. The encoder quantizes these model parameters, adds error correction code, and then transmits them in a data stream of 2.4 - 9.6 kbps. The decoder receives the bit stream, reconstructs the parameters of the model, and uses these parameters to generate synthetic speech information.

Experiment environment: The hardware composition of the voice collector includes a microcontroller C8051F120 as the control core, voice codec chip AMBE-1000, A/D-D/A conversion chip CSP1027, audio amplification filter chip TLC2272, and voice signal amplification output chip LMX358.

Result: The results showed that the proposed method can effectively encrypt the voice transmission process with fast encryption speed and good encryption effect. The data integrity was high, with a data integrity rate of 93% when the data encryption amount reached 50GB. The method also showed strong anti-attack ability, with a low data loss rate of only 0.33% when the attack strength coefficient was 1.5.

d: KRASNOWSKI et al. [5] IN 2022

This paper presents a novel encryption scheme for securing voice communications over 3G/4G or VoIP voice calls that uses the encoder, encryption, and modulator module as shown in Figure 3. The encoder module of this scheme uses a perceptually-oriented method to represent speech timbre as points on a 16-dimensional hypersphere of parameters. The encryption module uses a scrambling scheme based on commutative group codes on hyperspheres in even dimensions. Additionally, the modulator module uses a novel pseudo-speech synthesis technique adapted to data transmission over voice channels with voice compression and voice activity detection, and leveraged a Machine Learning (ML) approach to enchance the reconstruction of vocal sounds that were enciphered, transmitted and distorted by the voice channel.

Method: The encryption process is divided into two stages: speech encoding and enciphering. In the speech encoding stage, a harmonic speech encoder models speech signals as a combination of amplitude-modulated harmonics, and maps 20 ms speech frames into a sequence of vocal parameters. The enciphering stage then uses a pseudo-random number generator (PRNG) to independently scramble the vocal parameters of every frame into a new set of parameters defined over a new space of pseudo-speech parameters. The scrambled sequence is then processed by the pseudo-speech synthesizer, which produces a synthetic signal resembling pseudo-speech. The encrypted signal duration is the same as the duration of the initial speech, which is an essential requirement in real-time operation. In the receiver side, the descrambling process reverses enciphering operations using the same vector of random integers produced by the PRNG. The output of the descrambling process is a sequence representing the harmonic parameters of the speech frames. Finally, the authors improve upon harmonic speech synthesis by introducing a

narrowband modification of the LPCNet [63], a ML based synthesizer.

Experiment environment: In the simulation experiments to evaluate the robustness of pseudospeech against channel distortion, the secure PRNG was implemented using the NumPy PCG64 random sequence generator. The robustness of deciphering was evaluated by compressing the encrypted signal with Opus-Silk 1.3.1 or inserting AWGN into an encrypted signal. The first experimental part about intelligibility was inspired by the speech intelligibility rating (SIR), where the participants were asked to estimate the intelligibility of 10 English sentences of about 10 seconds each. The second experimental part was a quality assessment based on MUSHRA methodology adapted for the medium quality speech signals perceptual evaluation.

Result: The simulations showed that the errors on energy and spectral envelope are non-negligible and they gradually rise with the growing level of distortion. The SIR experiment showed that the participants recognized about 12% fewer words from the synthesized speech samples than in the reference. The MUSHRA tests showed that the introduction of distortion into encrypted signals resulted in degraded speech quality. The rating of speech quality decreased with the increasing level of distortion. A small channel distortion, like the one introduced by AWGN at SNR 20 dB, has a relatively minor impact on perceived speech quality. The excerpts decrypted from signals sent over FaceTime were rated at 49% ('Fair').

4) HARDWARE CODEC

The Hardware Codec method includes proposed solutions that employ audio codecs for the modulation of encrypted stream-bits into audio signals. There are six proposed solutions published in six papers that included in this method.

a: CHUMCHU et al. [50] IN 2012

They presented a framework for voice encryption over GSM-based networks that utilized the RC4 algorithm to protect digital voice. Based on the Figure 3, the proposed framework contain the ABME-2000 Codec as a encoder and modulator modules, and the RC4 algorithm as the encryption module that implemented in PIC18f4553 microprocessor.

Method: The communication process involved three phases: parameter reading, CSD call making, and digital voice transmitting. In the parameter reading phase, the prototype reads a special SMS message stored in the mobile phone, which includes the secret key, receiver mobile phone number, and status. The CSD call making phase involves the prototype commanding the mobile phone to make a CSD call using standard AT commands sent via Bluetooth link. In the digital voice transmitting phase, the prototype uses various components such as audio amplifiers to amplify the analogue signal from the microphone and the signal from D/A before sending to the speaker. The A/D and D/A are used to transform between analog and digital signals, with the analog device chip, AD73311, performing this function. The codec is used to compress and decompress the digital information of voice to a low bit rate, and in the project, the ABME-2000 codec is utilized. The two microprocessors, PIC18f4553, are used for the RC4 algorithm encryption and control system. Finally, the very small EDS200 Bluetooth module is used to facilitate communication between the prototype and mobile phone.

Experiment environment: The experimental scenario consisted of two identical voice encryption prototypes and two Motorola mobile phones connected to different networks via PSTN.

Result: The experimental results demonstrated acceptable performance, and the prototypes could be commercialized with a low cost of less than 40 US dollars.

b: BORUCHINKIN [51] In 2015

The study presents a communication system with hardware encryption, consisting of four components: an encrypting center, a switching server, a mobile application, and a hands-free headset. This system is using Vocoder ML7029 as the encoder and modulator module, then the AES and RSA algorithm as the encryption module implemented in MCU STM32F215, as shown in Figure 3.

Method: The headset generates session keys using symmetric encryption (via AES), asymmetric encryption and electronic signs (for RSA). The mobile application sends messages to the switching server, which forwards them to the recipient after successful authentication.

Experiment environment: The study conducted tests on a communication system with hardware encryption, comprising a switching server using Openfire and a hands-free headset with three main blocks: vocoder ML7029, microcontroller MCU STM32F215, and Bluetooth-module SPBT2632C2A.AT2 compatible with Bluetooth v3.0.

Result: All architecture elements were tested separately and as a whole. The switching server was successfully tested for cluster functioning and system scalability through load testing. During voice-encrypted calls, the time delay did not exceed 300 milliseconds, and the R-factor signal quality parameter was 93 with a stable fast-speed internet connection. In networks with lower speed connections, such as EDGE, 3G, and LTE, the R-factor was rated at satisfied and very satisfied levels of user satisfaction (83, 87, and 91, respectively).

c: CHOUHAN AND SINGH [52] IN 2015

This study proposed a method for secure voice communication over GSM network using the TEA [64]. According to Figure 3, this method utilized the AKM AK4642EN stereo codec as an encoder module and 128-bit key TEA algorithm as encryption module implemented in ATMEGA 328 Microcontroller Arduino.

Method: The input speech signal is first amplified using a variable gain amplifier and digitized using an AKM AK4642EN stereo codec. The output bit stream is encrypted using a 128-bit key TEA algorithm embedded in an

ATMEGA 328 Microcontroller Arduino board and converted into a synthesized speech signal using a Digital to Analog Converter (DAC), which is suitable for the GSM codec and network requirements.

Experiment environment: The study implemented a real-time prototype of the system, demonstrating full duplex secure voice communication on GSM-to-GSM voice calls.

Result: The encrypted waveforms were completely different from the original speech signal, and the TEA algorithm completely distorted the human speech after encryption. The decryption block in the receiver's side successfully recovered the original signal back.

d: LIN et al. [53] IN 2015

The paper proposes a secure voice communication based on neural cryptography implementation using custom instructions to achieve very low resource devices for real-time performance.

Method: The proposed method involves preprocessing the voice signal, followed by a 512-point Short-Time Fourier Transform (STFT) as the encoder module, then used a four-stage multilayer feedforward neural network for encryption and decryption module. This method corresponds to Figure 3. The synaptic weights of the first and last two stages serve as the encryption and decryption keys, respectively. The interface between the base processor and the custom functional unit (CFU) only includes the control signals for the custom instruction (CI) encoding and the synchronization of multi-cycle custom instructions.

Experiment environment: The paper reports experiments conducted using MATLAB to determine the topology parameters of the feedforward neural networks and to analyze the feasibility of cryptography. They also used C++ programs to normalize the MATLAB results to the correct range of fixed point values. Finally, they implemented the proposed neural cryptography on an embedded processor with custom instructions to speed up the execution of secure voice communication on very low resource embedded processors. The Altera Nios II/e was chosen as the very low resource target base processor for the DE2-70 board running at 100 MHz.

Result: The MATLAB experiments showed that the MSE result was in the range of 0.0473% to 16.4%, the total entropy result was in the range of 0.841 to 493, and the SNR result was in the range of −0.627 to 15.2 with 16 number of neurons showing the best result. The larger numbers of neurons or bits resulted in more hardware cost and execution time. From these experiments, the authors suggest using the number of bits to determine the desired sound quality and the minimum number of neurons to decide the desired security level.

e: MONDAL AND SHARMA [54] IN 2019

The paper describes an application of AES on a real-time secured voice communication system using FPGA. The system consists of Micro-Blaze, UART control, 128-bit key encryption, and GPIO. The Analog to Digital Converter (ADC) and Pulse Density Modulation (PDM) is used as encoder module, then the AES-128 is used as encryption module, such as shown in Figure 3.

Method: The transmitter side module processes human voice data through several steps, including capturing voice data using a microphone sensor and converting the analog data to digital format using an ADC. The data is then modulated using PDM and divided into frames for encryption. AES-128 is used to encrypt the data, which is then wirelessly transmitted using an RF-based Pmod transceiver module. The receiver side module decrypts the received signal using the public key provided during encryption, reconstructs the signal using interpolation techniques, and converts it to analog signal using PWM before playing it through the mono audio port.

Experiment environment: The design was implemented on Nexys4 DDR FPGA kits using VHDL in Vivado 2018.3 Design Suite.

Result: The results showed high security and superb voice quality in real-time secure voice communication.

f: GE et al. [11] IN 2021

The paper proposes a voice encryption device that uses a composite encryption method to divide speech into frames, rearrange them in the time domain, and encrypt the content of the frames. Based on the Figure 3, the proposed method contain the WM8731 audio chip as an encoder and modulator modules, and the 12,960 scrambling mode as the encryption module that implemented in FPGA.

Method: The voice data is collected by the WM8731 audio chip, stored in the SDRAM, grouped, and then encrypted by the FPGA. The collected voice data is stored in the SDRAM, with sequential data points forming a frame, and subsequent frames being aggregated into an encryption period. The device shuffles the order of these encrypted information segments and protects its important content from eavesdropping. The receiver can recover the original voice information according to the pre-arranged scrambled order. To restore the scrambled voice to its original form, the predetermined scrambling order is executed.

Experiment environment: To verify the effectiveness of the proposed method, the paper conducted time delay, voice quality, and encryption strength analysis tests on the voice encryption effect when using mobile phones for calls.

Result: Increasing the frame size and the number of frames can effectively increase the encryption strength, followed by an increase in delay of the product. The voice quality results show that the respondents could hear the original voice message clearly and the encryption did not affect the normal call, with an average of voice quality assessment of 80%. The design has a wide range of applications and strong portability. For various voice call scenarios, it offers good call quality. There are up to 12,960 encryption modes in total, and the time required to crack the encryption by testing all parameters should be at least 200 hours.

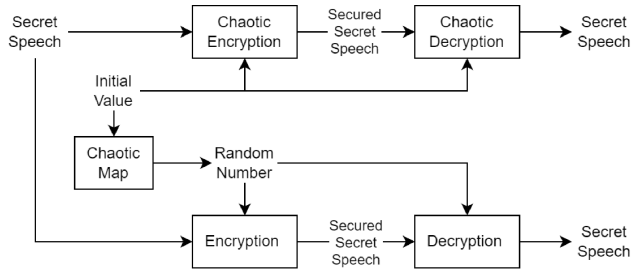

FIGURE 4. Secure voice communication based on chaotic cryptography technique.

C. CHAOTIC CRYPTOGRAPHY

The Chaotic Cryptography category comprises published papers proposing voice communication security using mathematical chaos theory in cryptography. Seven papers belonging to this category feature five proposed solutions. The complete list of these solutions can be found in Table 3.

Secure voice communication method utilized chaotic algorithm to generate encryption key and for encryption algorithm such as shown in Figure 4. Chaotic algorithm can be used to become pseudo random number generator from the initial values by using the certain chaotic map, then the generated random number is used as the secret key for the encryption and decryption algorithm. Furthermore, the chaotic algorithm can be used as the encryption method to permute the secret speech.

a: LEI et al. [65] IN 2004

The paper proposes a secure voice communication system based on DSP and a chaotic voice encryption algorithm combining Cat map and Logistic map.

Method: The analog voice signal is first transformed into PCM voice codes, which are then encrypted by the DSP chip. The chaotic algorithm in this paper are used for fast block encryption algorithm as shown in Figure 4. The block encryption algorithm is consist of Cat map permutation algorithm and Logistic map diffusion algorithm. The encoded data is transmitted through the TLI6C554 ASE, and the received data is decoded and transformed back into PCM codes by the DSP, which are then transformed back into the analog voice signal.

Experiment environment: The experiment involved two devices connected by a 5-meter long cable of RS-232 serial communication standard. The two devices had microphones and speakers, and two persons communicate with each other using the system.

Result: The result of the experiment showed that the proposed system was able to provide safe and real-time voice communication based on about 20 seconds speech waveform of the original sound wave, the encoded sound wave, and the decoded sound wave. Moreover, the authors noted that there was still much work to be done to make the system practicable in the future.

b: TANG AND TANG [66] IN 2005

The paper proposes a voice signal encryption scheme based on a fast chaos-based random number generator with a cascade approach to resolve the quantization effect of a finite precision with fixed-point arithmetic.

Method: The paper proposes a fast chaos-based encryption method for voice communication using a stream cipher. The encryption process uses a discrete version of a skew Tent map as the random number generator, with an additional mixing process based on a high-dimensional Cat map to resolve the quantization effect of a finite precision to generate 32-bit random key stream. The encryption function proceeds with the original voice data in 32-bit combined with the key stream, and the resulting encrypted signal is sent over the Internet using a UDP/IP channel. Thae chaotic algorithm in this paper are used for pseudo random number generator, as shown in Figure 4, with modulo method as encryption algorithm. The decryption function applies the same password and decrypts the received signal promptly, even if there is packet loss.

Experiment environment: The experiment was carried out using a sine signal as the input signal.

Result: The encrypted signal looked like noise with a flat spectrum with no additional bandwidth. The enciphering/deciphering rate was 18.51 Mbps on a Windows XP Professional with 2.6 GHz Pentium-4 processor and 512 MB RAM PC platform, which is about 3 times faster than DES.

c: LIU AND CHENG [67] IN 2017

The article proposed a chaos encryption algorithm based on a conversion table that utilizes an improved Logistic mapping as a pseudo random number generator, as shown in Figure 4. The algorithm was designed to improve real-time and security aspects.

Method: A new condition is added to the encryption algorithm based on the improved Logistic mapping to evenly update control parameters: $\mu \times k = 2300$. This increases the key space from (μ, x) to (μ, x, k). Before encrypting or decrypting voice signals, the algorithm sets initial values for important variables and calculates the first chaotic state. The encryption process involves calculating several variables and the ciphertext using the conversion table. If certain conditions are met, the encryption algorithm updates some parameters and calculates the number of iterations of the chaotic system. Decryption is similar and uses the same initial conditions. It adds a variable representing the value of a table cell and a formula for calculating the plaintext.

Experiment environment: The experimental environment involved running tests on MATLAB and STM32F407 Discovery, and the voice signals were output by MATLAB.

Result: The results showed that the encryption algorithm effectively hid and covered up the statistical information of sampling values on the original voice signals, resulting in improved security. The encrypted voice signals were decrypted and restored to the original voice signals with nearly identical waveform and frequency spectrum.

TABLE 3. Secure voice communication based on chaotic cryptography.

Solutions	Year	Throughput Cap.	Error Rate	Recovered Speech Quality	Security of The System
Lei, Zhao, Dai, *et al.* [65]	2004	8 kHz, 8 bit	-	-	Chaotic Encryption Algorithm Combining Cat Map and 256 Parts Logistic Map
Tang and Tang [66]	2005	-	-	The original signal is promptly received from the decryption results	Fast 32-bit Chaos-Based PRNG
Liu and Cheng [67]	2017	22.05kHz, 8-bit	-	Voice restored can be recognized its contents	PRNG Chaotic Encryption 2600 Control Parameters
Riyadi, Pandapotan, Khafid, *et al.* [69] [68]	2018	25 kHz, 8 bits	MSE: 0.2048	Audio clear and intelligible	128-bit Chaotic Cryptographic CFB Mode
Hayati, Suryanto, Ramli, *et al.* [70] [71]	2019	16 kHz	-	FwSNR: 3.620 dB, PESQ: 1.251	Chaotic Permutation Multicircular Shrinking and Expanding 256 bits

d: RIYADI et al. [68], [69] IN 2018

The authors proposed a secure voice channel prototype that utilizes a 128-bit Chaotic cryptographic algorithm with Cipher Feedback as an alternative for simple but reliable algorithm.

Method: The algorithm was applied for secure voice channel using a 128-bit key, with plaintext divided into 8-bit blocks. The secret key was directly used for encryption and decryption and divided into 8-bit blocks called session keys. The discrete chaotic algorithm uses the logistic function with an initial condition in the value range of 0 to 1, the number of iterations, and the system parameters in the chaotic range of 3.75 to 4.00. Real and pseudo random numbers are generated using the chaotic algorithm, logistic function, as shown in Figure 4. A session key is selected randomly, and the initial value and number of iterations are modified accordingly. The system parameters are determined in the range of chaos, and the logistic functions are iterated using initial conditions, number of iterations, and system parameters generated in the previous step. The final value is used for encryption and decryption of ciphertext using modulo method. The system was implemented using two Xilinx Spartan-3 FPGA boards and managed serial communication between the FPGAs using Xilinx ISE Design Suite 14.6 software.

Experiment environment: The experiment used two Xilinx Spartan-3 FPGA boards for encryption and decryption.

Result: The encryption block produced different ciphertext for the same input value due to the feedback cipher mode, but the decryption process successfully recovered the original binary input with similar value. The resulting decryption signal was clear and intelligible with MSE result of 0.2048 V^2 and THD-N result of 4.41%.

e: HAYATI et al. [70] IN 2019

This study proposed a platform called Encryption Decryption Device (EDD) for securing voice data over wireless mobile communication through end-to-end encryption.

Method: The platform consists of three functions: ADC, encryption/decryption, and DAC. Encryption and decryption are carried out using a digital dynamic complex permutation module that rotates through a set of expanded keys. The chaotic algorithm in this paper are used for encryption algorithm as shown in Figure 4. At the transmitter, the voice is encrypted using Chaotic Permutation Multicircular Shrinking, and at the receiver, it is decrypted with Chaotic Permutation Multicircular Expanding. The speech is then enhanced using a noise estimator that combines the Wiener filter and q-spectral subtraction, and the enhanced speech is transformed back to the time domain using inverse discrete Fourier transform (IDFT) and overlap and add (OLAP) [71].

Experiment environment: The experiment environment involved passing selected utterances through an encryptor before being transmitted using various communication channels, including GSM, IMA-ADPCM, Microsoft ADPCM, and PCM. The received speech was then decoded using a decryptor process and fed through a speech enhancement method.

Result: The results of the test showed that the quality of the decrypted speech had decreased in quality and was not as clear as when on the transmitter side. With speech enhancement, the average log spectral distances (LSD) in dB for various communication channels were improved, with GSM having an LSD of 40.00, I-ADPCM having an LSD of 36.17, M-ADPCM having an LSD of 43.93, and PCM having an LSD of 36.17. The PESQ test results also showed an improvement in the quality of the decrypted speech for all encoding schemes, with GSM having a PESQ score of 1.251, I-ADPCM having a score of 2.654, M-ADPCM having a score of 2.649, and PCM having a score of 2.654. For most methods, the enhanced speech tended to have lower FwSNR. The FwSNR test results showed that GSM had an FwSNR of 3.620, I-ADPCM had an FwSNR of 13.415, M-ADPCM had an FwSNR of 12.799, and PCM had an FwSNR of 13.405. Overall, the study showed that the speech enhancement method could improve the quality of decrypted speech transmitted through various communication channels.

V. ANALYSIS

A. STEGANOGRAPHY

There are a total of eight proposed solutions distributed among ten published papers included in the Steganography category, which can be found in Table 1. According to the

TABLE 4. The papers included based on the year of the category group.

Year	Steganography	Cryptography using Modem Method				Chaotic Cryptography
		Codebook Optimization	Optimized Modulation	Parameter Mapping	Hardware Codec	
2004						[65]
2005						[66]
2006	[22]			[48]		
2007	[23]		[39]			
2008			[40]			
2009	[24]		[41]			
2010						
2011			[42]	[46]		
2012					[50]	
2013			[43]			
2014						
2015				[45]	[51], [52], [53]	
2016						
2017			[44]	[47]		[67]
2018	[25], [26]					[69], [68]
2019	[27], [28]	[3]			[54]	[70], [71]
2020	[30]			[49]		
2021			[14]		[11]	
2022	[29], [31]	[37], [38]		[5]		

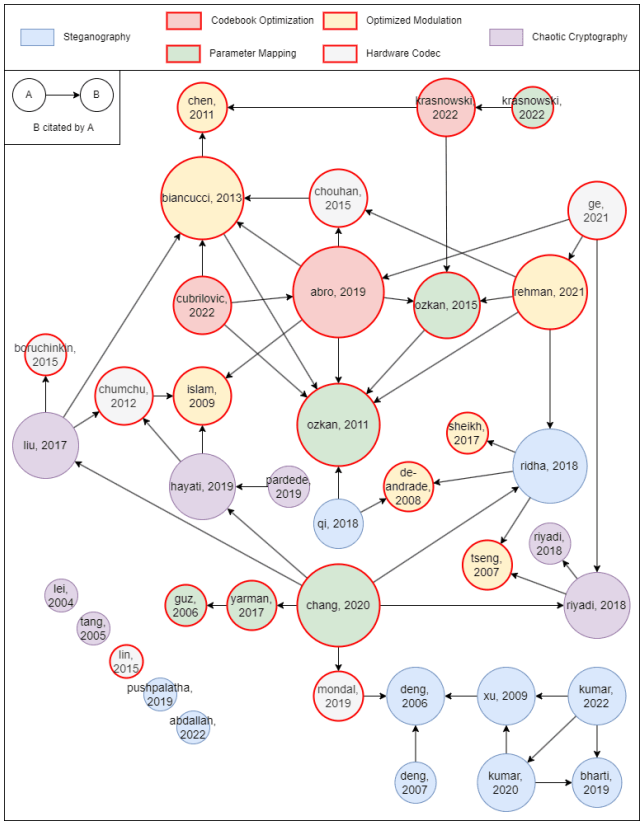

FIGURE 5. Paper network map based on citations.

Paper Network Map in Figure 5, the total citation for this category by the papers included in this review paper are seven. The most popular technique used in this category is DCT.

Above all proposed solutions in the Steganography category, a proposed solution by Ridha, Jawad, and Kadhim [26] has the best performance by the evaluation parameters. This paper was published by IEEE Communications Letters in 2018. The proposed system is based on a modified version of BSS technique. This proposed solution has the highest throughput capacity with 44.1 kHz. Furthermore, this proposed solution evaluate the speech quality using objective assessment by FwSNR and subjective assessment by DMOS. The subjective assessment result is better than the other proposed solution assessments such as PESQ and MOS. The experiment setup was built in MATLAB and tested in real-time environment, where the encrypted speech was transferred to the decryption part of the system over a Wi-Fi local area network. The experiment setup did not quite represent the real condition of voice channel communication with the limitations that were explained in section I. Moreover, the residual intelligibility is the only parameter to measure security of the system, which is not enough to assure a strong security. Although the BSS technique is compatible to scrambling the voice signal, but the security level is far below the standard security algorithm.

In the last five years, from 2018 to 2022, there are seven published papers included in this category, which is the most papers in the last five years above other category, as shown shown in Table 4. The throughput capacity parameter shows that six proposed solutions had the speech bit rate above 16 kbps. This category has the most paper with high throughput capacity above all categories. For the quality evaluation, there are seven proposed solutions using the objective assessment and two proposed solutions using the subjective assessment. This category uses the most objective assessment above other category.

There are some drawbacks of this technique such as low imperceptibility by hiding the secret voice within the cover voice. This is contrary to the purpose of Steganography techniques to achieve high imperceptibility. Moreover, most proposed solutions do not provide voice channel environment testing, but rather use general transmission interference

such as AWGN. Only two proposed solutions are using voice channel environment testing, PSTN interference and G.728 LDCELP, with poor PESQ results. This technique is more suitable for asynchronous communication than real-time communication. Improving the imperceptibility level and adjusting for voice channel environment testing are concerns for further development.

B. MODEM-BASED CRYPTOGRAPHY

There are 29 published paper featuring 25 proposed solutions in this category. The papers included in the Modem-based Cryptography mostly feature robustness evaluations. This amount of paper is the most among the other categories. Eight of these papers provide a promising level of robustness. Among ten proposed solutions that utilized the standard security algorithm over all categories, nine of them belonged to the Modem-based Cryptography category. Overall, the most popular technique used by the Modem-Based Cryptography category are LPC, MBE, Codec2, OFDM and QPSK. Based on the ability to use standard security algorithms, the Linear Prediction Vocoder and Multi Excitation techniques are more compatible. The list of these papers and methods are comprehensively listed in Table 2. The detailed analysis for each method are explained below.

1) CODEBOOK OPTIMIZATION

This method is featured in three published papers, which present three proposed solutions belonging to the Codebook Optimization method. According to the Paper Network Map in Figure 5, the total citation for this method by the papers included in this review paper is three. The proposed solutions in this method are using LPC technique and Mixed Excitation Coding based technique Codec2 [61].

Based on the throughput capacity of the system parameter, the solution proposed by Krasnowski et al. [37] has the highest value with 6.4 kbps tested over VoIP voice channels using 4G wireless network and 2.4 kbps tested over 3G cellular calls. Based on the error rate parameter, the solution proposed by Abro et al. [3] has the lowest BER value with 0% tested over the real GSM network using two GSM mobile phones. Both of these papers implemented the standard security algorithm, AES 256 bit and Diffie-Hellmann key exchange protocol with EC25519.

All papers in this method were published between 2018 to 2022. All proposed solutions implement the standard security algorithm, AES-256. It can be deduced that this method is the most promising method to implement the standard security algorithm. Finally, all proposed solutions are tested in real-time voice communication channel, such as GSM and VoIP. Thus, this method is suitable for real-time communication.

There are some drawbacks of this method. The throughput capacity parameter shows that only one proposed solution had throughput capacity above 4 kbps and one proposed solution had throughput capacity above 16 kbps. For the quality evaluation, the proposed solutions neither used objective nor subjective assessment. Improving throughput capacity and using standard quality assessments are concerns for further development.

2) OPTIMIZED MODULATION

A total of seven proposed solutions are available, which have been published in seven scientific papers belonging to this method. According to the Paper Network Map in Figure 5, the total citation for this method by the papers included in this review paper is 14, which makes this method the most citated category. The OFDM and QPSK are the most popular technique in this method. Most of the proposed solutions were tested in voice channel environment. Four solutions are tested in simulation of GSM voice communication channel and two solutions are tested in real-time GSM voice communication channel.

There are three proposed solutions that have high throughput capacity with speech bit rate above 16 kbps. Among them, the proposed solution by Andrade et al. [40] has the highest throughput capacity with 8 kHz of 16 bits, or equivalent to 128 kbps. This proposed solution is the only solution that used standard quality assessment, the objective assessment PESQ. Even though the PESQ result is between 2.3 to 2.9, it is considered to be almost 'fair' quality. Included among the high throughput capacity, another proposed solution is by Biancucci et al. [43] with 8 kHz of 13 bits, or equivalent to 104 kbps. By developing the method based on FM modulation, this proposed solution is the only solution that implement the standard security algorithm, AES-256 128 bit block cipher. Lastly, the another proposed solution included among the high throughput capacity is by Tseng and Chiu [39] with 8 kHz of 8 bits, or equivalent to 64 kbps. Even though this proposed solution does not implement the standard security algorithm, but it has many possible scrambling keys with 93 factorials possible scrambling keys OFDM Permutation. In two recent publications, the QPSK modulation method is utilized.

In the last five years, from 2018 to 2022, there was one published paper included in this method, which is the least papers in the last five years than other category, as shown shown in Table 4. Only three proposed solutions provide throughput capacity 16 kbps or higher. There was one proposed solutions using the objective assessment and none of them used the subjective assessment. Only one proposed solution implement the standard security algorithm, AES-256. This method is the most basic method to improve voice communication security and the popularity of this method is declining.

3) PARAMETER MAPPING

Six published papers within this method presented four proposed solutions. According to the Paper Network Map in Figure 5, the total citation for this method by the papers included in this review paper is eight. The paper authored by Ozkan et al. [46] included in this method has the most citation above all papers, with six citations. The proposed solutions in this method are using LPC technique, LPCNet the

AI-based LPC, MBE, and custom techniques such as SYMPES [48], [62].

Above all proposed solutions in the Parameter Mapping method, a proposed solution by Yarman et al. [47], using SYMPES voice codec, has the best value of the evaluation parameters. This solution provide high throughput capacity with 16 kbps, good speech quality with MOS 82%, and implement the standard security algorithm AES-256. Nevertheless, the implementation of SYMPES voice codec requires a GPU and a CPU to provide large computing power.

This method has good capability for implementing standard security algorithm. There are three proposed solutions that implement the standard security algorithm: 3Des-168, AES-128 and AES-256. For the quality evaluation, none of the proposed solutions used the objective assessment, but there are two proposed solutions that used the subjective assessment. Thus, this method used the most subjective assessment above other category. All proposed solutions were tested in voice channel environment, real-time and simulation testing.

In the last five years, from 2018 to 2022, there were two published papers included in this method. The throughput capacity parameter shows that two proposed solutions had speech bit rate above 16 kbps. Thus, improving throughput capacity is concern for further development.

4) HARDWARE CODEC

Seven papers belonging to this method feature six proposed solutions. According to the Paper Network Map in Figure 5, the total citation for this method by the papers included in this review paper is six. There are various hardware codec techniques, such as PCM, PDM, STFT, and MBE.

The proposed solution by Mondal and Sharma [54] is the only solution that provides high throughput capacity with 15 kHz of 32 bits, or equivalent to 480 kbps and implements standard security algorithm AES-128. Nevertheless, the speech quality assessment was not conducted in this proposed solution. Furthermore, the testbed of this proposed solution was using RF based wireless communication, which does not represent the real condition of voice channel communication with the limitations that were explained in the section I.

The throughput capacity parameter shows that three proposed solutions had the speech bit rate above 16 kbps and one above 4 kbps. For the quality evaluation, there are two proposed solutions using the objective assessment and one proposed solution using the subjective assessment. Most of the proposed solutions are tested in voice channel environment. Two of them were tested in real-time GSM voice communication channel, one of them were tested in real-time PSTN voice communication channel and one of them were tested in VoIP simulation.

In the last five years, from 2018 to 2022, there were two published paper included in this method. There are two proposed solutions that implement the standard security algorithm, AES-128 and AES-RSA. The Hardware Codec category utilized the DSP chip to encode and decode the speech signal. There is a lack of hardware codecs utilization with standard security algorithms in recent years, indicating that the majority of methods developed for end-to-end voice communication security have not reached the hardware level implementation.

C. CHAOTIC CRYPTOGRAPHY

Seven papers belonging to this category feature five proposed solutions. The complete list of these solutions can be found in Table 3. According to the Paper Network Map in Figure 5, the total citation for this category by the papers included in this review paper is four. Some of the proposed solutions in this category are using PCM and DFT technique, but most of them simply use the custom chaotic cryptography techniques.

The proposed solution by Hayati et al. [70] is the only solution that provided standard quality assessment, with FwSNR: 3.620 dB and PESQ: 1.251 as objective assessments. This proposed solution has high system throughput capacity with 16 kHz. The encryption method implemented by this solution is 256 bits Chaotic Permutation Multicircular Shrinking and Expanding. For this proposed solution, the combination of permutation and substitution should be developed to increase the security level of the encryption.

In the last five years, from 2018 to 2022, there were four published paper included in this category. The throughput capacity parameter shows that three proposed solutions had the speech bit rate above 16 kbps. For the quality evaluation, there is one proposed solution using the objective assessment and none of them used the subjective assessment. None of this category implement the standard security algorithm, because the security algorithm developed in this category are based on the chaotic cryptography. Only two proposed solutions are tested in voice channel environment, which is all of them are using Arnold Cat Map for the chaotic algorithm. Thus, the development of chaotic cryptography is still a challenge for end-to-end secure voice communication. Chaotic cryptography based on Arnold Cat Map is the most promising for voice security development.

VI. DISCUSSION

A. SUMMARY

After analyzing the 39 papers, a total of 33 proposed solutions for secure voice communication are identified and classified into three category Steganography, Modem-based Cryptography, and Chaotic Cryptography. Overall, standard quality assessment such as subjective and objective assessment are rarely conducted. The utilization of Artificial Intelligent for this research is limited, only four published papers utilize it.

Steganography category is the most popular technique from 2018-2022. The rise in popularity of Steganography techniques is due to their ability to conceal the existence of secret information and exhibit good robustness against signal processing attacks. This technique holds promise for achieving a high throughput capacity, and its evaluation adheres to

standard quality assessment procedures. Imperceptibility and compatibility for real-time voice communication should be addressed in further development.

Modem-based Cryptography is the most promising category to implement standard security algorithms than steganography and chaotic cryptography categories. Linear Prediction Vocoder and Multi Excitation techniques are compatible for implementing standard security algorithms. There are four methods included in this category: Codebook Optimization, Optimized Modulation, Parameter Mapping and Hardware Codec.

1) Codebook Optimization method shows promise in implementing standard security algorithm. All proposed solutions of his method implement AES-256 as security algorithm. Moreover, by testing in real-time voice communication channel, this method shows it is suitable for real-time voice communication. Low throughput capacity and lack of standard quality assessment should be a concern in further development.
2) With high number of citations and least publications in recent years, concluded that Optimized Modulation method is the most basic method for end-to-end secure voice communication. Moreover, lack of standard quality assessment and standard security algorithm implementation, inflict the declining of this method popularity for future development.
3) Parameter Mapping method has a good capability to implement standard security algorithm, such as 3Des-168, AES-128 and AES-256. This method uses the most subjective assessment than the other categories. Moreover, this method is compatible for real-time voice communication, since all proposed solutions are tested in voice channel environment. Low throughput capacity of this method should be a concern for further development.
4) Hardware Codec method based solutions have a good number of high throughput capacities, standard quality assessments, and testings in voice channel environment. Lack of standard security algorithms implementations and publications in recent years, indicate that the majority of methods developed for end-to-end voice communication security have not reached the hardware level implementation.

Most of Chaotic Cryptography category publications were published in the last five years. Some of this category have high throughput capacity. This category has a lack of standard quality assessments and testings in voice channel environment. Thus, the development of chaotic cryptography is still challenging for end-to-end secure voice communication. Testing in voice channel environment was only done for Arnold Cat Map based method, indicating that this method is the most promising method for voice security development based on chaotic cryptography.

B. RESEARCH OPPORTUNITIES

In this section the research opportunities are presented.

- Some techniques can fulfill several parameters. To achieve high throughput capacity and good standard quality assessment result, Steganography technique can fulfill it. To be able to implement standard security algorithm and suitable for real-time voice communication, Codebook Optimization and Parameter Mapping of Modem-based Cryptography can fulfill it, especially by utilizing Linear Prediction Vocoder and Multi Excitation. There is no technique or even a single paper that can fulfill all parameters, so research to fulfill all parameters needs to be developed in the future. In addition, the implementation of hardware codec for secure end-to-end voice communication method should be further developed. This development is expected to provide fast processing for real-time voice communication.
- Signal synchronization is concerned for further research as stated by [5], [37], and [45]. Synchronization of the received signals is necessary to successfully decrypt the received signals.
- The quality of speech is included as the QoS and QoE. Speech quality improvement should be further researched. Some proposed solutions implement the speech enhancement such as [5] and [71]. To evaluate the speech quality, The standard quality assessment, such as objective assessment, and especially subjective assessment should be required. Some of the objective assessments should be considered, such as Frequency-Weighted Segmental SNR (fwSNRseg), PESQ and Deep Learning based assessments [72]. Some of the subjective assessments such as Comparison Category Rating (CCR) assessments included Comparison Mean Opinion Score (CMOS) and Absolute Category Rating (ACR) assessments included MOS, Diagnostic Acceptability Measure (DAM) and ITU-T P.835 Standard [73], should be considered more frequently.
- The utilization of Artificial Intelligent has a lot of space to be developed for this research area. The development of AI-based secure end-to-end voice communication system, AI-based quality assessment and AI-based quality improvement should be considered for future research. The development of speech recognition based voice security is a research opportunity, since capability of Neural Network for speech recognition is fascinating. Deep Learning based assessments for speech quality is also a research opportunity [72]. AI-based quality improvement such as LPCNet [5] should be further developed. Deep learning algorithm can be utilized for extracting the important features from audio signal and for authentication [31]. Machine learning algorithm can be developed to predict the channel condition to select the suitable method for a specific channel [14].

VII. CONCLUSION

Secure end-to-end voice communication is crucial for various stakeholders, but limitations like limited bandwidth, lossy compression, and noisy channels hinder its performance. Researchers in this field can learn from this review paper whose very comprehensive evaluation parameters. This paper offers an overview of end-to-end secure voice communication and evaluates various techniques for securing voice transmissions over communication networks, including Steganography, Modem-based Cryptography, and Chaotic Cryptography, based on different parameters. It covers also encryption algorithms for voice communication.

From The 39 papers reviewed in this paper, 33 proposed solutions were identified and classified into three categories: Steganography, Modem-based Cryptography, and Chaotic Cryptography. Steganography is a promising technique to provide high throughput capacity and good standard quality assessment results, with the drawbacks being low imperceptibility and low compatibility for real-time voice communication. Modem-based Cryptography is the most promising category to implement standard security algorithms and is suitable for real-time voice communication, with the drawback being low throughput capacity. Chaotic Cryptography has high throughput capacity but low compatibility for real-time voice communication.

Lack of standard quality assessment, such as subjective and objective assessment, become a concern. In addition, the speech quality improvement should be considered. The implementation of standard security algorithm is challenging for the end-to-end secure voice communication. The Linear Prediction Vocoder and Multi Excitation techniques have a good capability for implementing standard security algorithms. Artificial Intelligent based method has a lot of space for development research.

REFERENCES

[1] *Measuring the Information Society Report*, International Telecommunication Union, Geneva, Switzerland, 2018.

[2] *The State of Broadband 2020: Tackling Digital Inequalities—A Decade for Action*, International Telecommunication Union and United Nations Educational, Scientific and Cultural Organization, Geneva, Switzerland, 2020.

[3] F. I. Abro, F. Rauf, Mobeen-ur-Rehman, B. S. Chowdhry, and M. Rajarajan, "Towards security of GSM voice communication," *Wireless Pers. Commun.*, vol. 108, no. 3, pp. 1933–1955, Oct. 2019, doi: 10.1007/s11277-019-06502-y.

[4] J. Scott-Railton, B. Marczak, B. A. Razzak, M. Crete-Nishihata, and R. Deibert, "Reckless exploit: Mexican journalists, lawyers, and a child targeted with NSO spyware," Citizen Lab Res. Report, Univ. Toronto, Mexico, Tech. Rep., Jun. 2017.

[5] P. Krasnowski, J. Lebrun, and B. Martin, "A novel distortion-tolerant speech encryption scheme for secure voice communication," *Speech Commun.*, vol. 143, pp. 57–72, Sep. 2022, doi: 10.1016/j.specom.2022.06.007.

[6] A. Greenberg, "Codebreaker Karsten Nohl: Why your phone is insecure by design," Tech. Rep., 2011.

[7] P. K. Gundaram, A. N. Tentu, and S. N. Allu, "State transition analysis of GSM encryption algorithm A5/1," *J. Commun. Softw. Syst.*, vol. 18, no. 1, pp. 36–41, 2022, doi: 10.24138/jcomss-2021-0104.

[8] *WhatsApp Voice Calls Used to Inject Israeli Spyware on Phones*, Financial Times, London, U.K., May 2019.

[9] C. Ntantogian, E. Veroni, G. Karopoulos, and C. Xenakis, "A survey of voice and communication protection solutions against wiretapping," *Comput. Electr. Eng.*, vol. 77, pp. 163–178, Jul. 2019, doi: 10.1016/j.compeleceng.2019.05.008.

[10] *Study Mobile Device Security: DHS*, D. O. H. Security, Washington, DC, USA, Apr. 2017.

[11] X. Ge, G. Sun, B. Zheng, and R. Nan, "FPGA-based voice encryption equipment under the analog voice communication channel," *Information*, vol. 12, no. 11, p. 456, Nov. 2021, doi: 10.3390/info12110456.

[12] M. Boloursaz, R. Kazemi, D. Nashtaali, M. Nasiri, and F. Behnia, "Secure data over GSM based on algebraic codebooks," in *Proc. East-West Design Test Symp. (EWDTS)*, Sep. 2013, pp. 1–4, doi: 10.1109/EWDTS.2013.6673148.

[13] R. Cox, S. De Campos Neto, C. Lamblin, and M. Sherif, "ITU-T coders for wideband, superwideband, and fullband speech communication [Series editorial," *IEEE Commun. Mag.*, vol. 47, no. 10, pp. 106–109, Oct. 2009, doi: 10.1109/MCOM.2009.5273816.

[14] M. U. Rehman, M. Adnan, M. Batool, L. A. Khan, and A. Masood, "Effective model for real time end to end secure communication over GSM voice channel," *Wireless Pers. Commun.*, vol. 119, pp. 1643–1659, Mar. 2021, doi: 10.1007/s11277-021-08299-1.

[15] D. Kahn, *The Codebreakers: The Comprehensive History of Secret Communication From Ancient Times to the Internet*. New York, NY, USA: Scribner, Dec. 1996.

[16] A. S. Bhatia and A. Kumar, "Post-quantum cryptography," in *Emerging Security Algorithms and Techniques*, K. Ahmad, M. N. Doja, N. I. Udzir, and M. P. Singh, Eds., 1st ed. Boca Raton, FL, USA: Taylor & Francis, 2019, pp. 139–158, doi: 10.1201/9781351021708-9.

[17] S. B. Sadkhan, A. A. Mahdi, and R. S. Mohammed, "Recent audio steganography trails and its quality measures," in *Proc. 1st Int. Conf. Comput. Appl. Sci. (CAS)*, Dec. 2019, pp. 238–243, doi: 10.1109/CAS47993.2019.9075778.

[18] E. A. Albahrani, T. K. Alshekly, and S. H. Lafta, "A review on audio encryption algorithms using chaos maps-based techniques," *J. Cyber Secur. Mobility*, vol. 11, no. 1, pp. 53–82, Nov. 2021, doi: 10.13052/jcsm2245-1439.1113.

[19] I. Makhdoom, M. Abolhasan, and J. Lipman, "A comprehensive survey of covert communication techniques, limitations and future challenges," *Comput. Secur.*, vol. 120, Sep. 2022, Art. no. 102784, doi: 10.1016/j.cose.2022.102784.

[20] *Methods for Subjective Determination of Transmission Quality*, document ITU, P.800, Aug. 1996.

[21] R. V. Cox and P. Kroon, "Low bit-rate speech coders for multimedia communication," *IEEE Commun. Mag.*, vol. 34, no. 12, pp. 34–41, Dec. 1996, doi: 10.1109/35.556484.

[22] Z. Deng, Z. Yang, and L. Deng, "A real-time secure voice communication system based on speech recognition," in *Proc. Int. Conf. Syst. Netw. Commun. (ICSNC)*, 2006, p. 22, doi: 10.1109/ICSNC.2006.13.

[23] Z. Deng, Z. Yang, X. Shao, N. Xu, C. Wu, and H. Guo, "Design and implementation of steganographic speech telephone," in *Advances in Multimedia Information Processing—PCM 2007*, vol. 4810. Berlin, Germany: Springer, 2007, pp. 429–432, doi: 10.1007/978-3-540-77255-2_52.

[24] T. Xu, Z. Yang, and X. Shao, "Novel speech secure communication system based on information hiding and compressed sensing," in *Proc. 4th Int. Conf. Syst. Netw. Commun.*, Sep. 2009, pp. 201–206, doi: 10.1109/ICSNC.2009.71.

[25] D. Qi, N. Longmei, and X. Jinfu, "A speech privacy protection method based on sound masking and speech corpus," *Proc. Comput. Sci.*, vol. 131, pp. 1269–1274, Jan. 2018, doi: 10.1016/j.procs.2018.04.342.

[26] O. A. L. A. Ridha, G. N. Jawad, and S. F. Kadhim, "Modified blind source separation for securing End-to-End mobile voice calls," *IEEE Commun. Lett.*, vol. 22, no. 10, pp. 2072–2075, Oct. 2018, doi: 10.1109/LCOMM.2018.2864146.

[27] G. S. Pushpalatha, D. S. Ramesh, and A. Raganna, "Voice encryption with watermarking for secure speech communication," *J. Emerg. Technol. Innov. Res.*, vol. 6, no. 1, pp. 406–414, 2019.

[28] S. S. Bharti, M. Gupta, and S. Agarwal, "A novel approach for audio steganography by processing of amplitudes and signs of secret audio separately," *Multimedia Tools Appl.*, vol. 78, no. 16, pp. 23179–23201, Aug. 2019, doi: 10.1007/s11042-019-7630-4.

[29] K. P. Kumar and A. Kanhe, "Secured speech watermarking with DCT compression and chaotic embedding using DWT and SVD," *Arabian J. Sci. Eng.*, vol. 47, no. 8, pp. 10003–10024, Aug. 2022, doi: 10.1007/s13369-021-06431-8.

[30] P. K. Kasetty and A. Kanhe, "Covert speech communication through audio steganography using DWT and SVD," in *Proc. 11th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2020, pp. 1–5, doi: 10.1109/ICCCNT49239.2020.9225399.

[31] H. A. Abdallah and S. Meshoul, "A multilayered audio signal encryption approach for secure voice communication," *Electronics*, vol. 12, no. 1, p. 2, Dec. 2022, doi: 10.3390/electronics12010002.

[32] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006, doi: 10.1109/TIT.2006.871582.

[33] R. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007, doi: 10.1109/MSP.2007.4286571.

[34] J. J. Eggers, R. Bauml, R. Tzschoppe, and B. Girod, "Scalar Costa scheme for information embedding," *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 1003–1019, Apr. 2003, doi: 10.1109/TSP.2003.809366.

[35] *VoiceAge: Unified Speech and Audio Database (USAC)*. Accessed: Apr. 9, 2023. [Online]. Available: https://voiceage.com/Audio-Samples-AMR-WB.html

[36] Y. Hu and P. C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," *Speech Commun.*, vol. 49, nos. 7–8, pp. 588–601, Jul. 2007, doi: 10.1016/j.specom.2006.12.006.

[37] P. Krasnowski, J. Lebrun, and B. Martin, "Introducing a novel data over voice technique for secure voice communication," *Wireless Pers. Commun.*, vol. 124, no. 4, pp. 3077–3103, Jun. 2022, doi: 10.1007/s11277-022-09503-6.

[38] S. Cubrilovic, D. Mandic, and A. Krstic, "Evaluation of improved classification of speech-like waveforms used for secure voice transmission," in *Proc. 21st Int. Symp. INFOTEH-JAHORINA (INFOTEH)*, Mar. 2022, pp. 1–5, doi: 10.1109/INFOTEH53737.2022.9751308.

[39] D. C. Tseng and J. H. Chiu, "An OFDM speech scrambler without residual intelligibility," in *Proc. TENCON IEEE Region 10 Conf.*, Oct. 2007, pp. 1–4, doi: 10.1109/TENCON.2007.4428903.

[40] J. F. de Andrade, M. L. R. de Campos, and J. A. Apolinario, "Speech privacy for modern mobile communication systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2008, pp. 1777–1780, doi: 0.1109/ICASSP.2008.4517975.

[41] S. Islam, F. Ajmal, S. Ali, J. Zahid, and A. Rashdi, "Secure end-to-end communication over GSM and PSTN networks," in *Proc. IEEE Int. Conf. Electro/Inf.ation Technol.*, Jun. 2009, pp. 323–326, doi: 10.1109/EIT.2009.5189636.

[42] L. Chen and Q. Guo, "An OFDM-based secure data communicating scheme in GSM voice channel," in *Proc. Int. Conf. Electron., Commun. Control (ICECC)*, Sep. 2011, pp. 723–726, doi: 10.1109/ICECC.2011.6066715.

[43] G. Biancucci, A. Claudi, and A. F. Dragoni, "Secure data and voice transmission over GSM voice channel: Applications for secure communications," in *Proc. 4th Int. Conf. Intell. Syst., Modeling Simulation*, Jan. 2013, pp. 230–233, doi: 10.1109/ISMS.2013.10.

[44] J. A. Sheikh, S. Akhtar, S. A. Parah, and G. M. Bhat, "A new method of Haar and Db10 based secured compressed data transmission over GSM voice channel," in *Intelligent Techniques in Signal Processing for Multimedia Security*, vol. 660, N. Dey and V. Santhi, Eds. Cham, Switzerland: Springer, 2017, pp. 401–426, doi: 10.1007/978-3-319-44790-2_18.

[45] M. A. Özkan and S. Berna Örs, "Data transmission via GSM voice channel for end to end security," in *Proc. IEEE 5th Int. Conf. Consum. Electron. Berlin (ICCE-Berlin)*, Sep. 2015, pp. 378–382, doi: 10.1109/ICCE-Berlin.2015.7391285.

[46] M. A. Ozkan, B. Ors, and G. Saldamli, "Secure voice communication via GSM network," in *Proc. 7th Int. Conf. Electr. Electron. Eng. (ELECO)*, Dec. 2011, pp. II-288–II-292.

[47] B. S. Yarman, C. Ulger, and A. B. Aslan, "SYMPES technique encoded IP-based secure voice communication system," in *Proc. Int. Symp. Signals, Circuits Syst. (ISSCS)*, Jul. 2017, pp. 1–3, doi: 10.1109/ISSCS.2017.8034870.

[48] Ü. Güz, H. Gürkan, and B. S. Yarman, "A new method to represent speech signals via predefined signature and envelope sequences," *EURASIP J. Adv. Signal Process.*, vol. 2007, no. 1, Dec. 2006, Art. no. 056382, doi: 10.1155/2007/56382.

[49] Z. Chang and M. Woźniak, "Encryption technology of voice transmission in mobile network based on 3DES-ECC algorithm," *Mobile Netw. Appl.*, vol. 25, no. 6, pp. 2398–2408, Dec. 2020, doi: 10.1007/s11036-020-01617-0.

[50] P. Chumchu, A. Phayak, and P. Dokpikul, "A simple and cheap end-to-end voice encryption framework over GSM-based networks," in *Proc. Comput., Commun. Appl. Conf.*, Jan. 2012, pp. 210–214, doi: 10.1109/ComComAp.2012.6154800.

[51] A. Y. Boruchinkin, "Secure voice communication system with hardware encryption of data on hands-free headset," in *Proc. 8th Int. Conf. Secur. Inf. Netw.*, Sep. 2015, pp. 76–79, doi: 10.1145/2799979.2800030.

[52] A. Chouhan and S. Singh, "Real time secure end to end communication over GSM network," in *Proc. Int. Conf. Energy Syst. Appl.*, Oct. 2015, pp. 663–668, doi: 10.1109/ICESA.2015.7503433.

[53] C. H. Lin, B. C. Yang, C. B. Duanmu, B. W. Chen, D.-S. Chen, and Y. Wang, "Neural cryptography for secure voice communication using custom instructions," in *Proc. Int. Conf. Embedded Syst. Appl. (ESA)*, 2015, pp. 57–61.

[54] S. Mondal and R. K. Sharma, "Application of advanced encryption standard on real time secured voice communication using FPGA," in *Proc. 10th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2019, pp. 1–6, doi: 10.1109/ICCCNT45670.2019.8944857.

[55] F. I. Abro, F. Rauf, M. Batool, B. S. C. Dhry, and S. Aslam, "An efficient speech coding technique for secure mobile communications," in *Proc. IEEE 9th Annu. Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON)*, Nov. 2018, pp. 940–944, doi: 10.1109/IEMCON.2018.8614855.

[56] C.-C. Chang and Y.-C. Hu, "A fast LBG codebook training algorithm for vector quantization," *IEEE Trans. Consum. Electron.*, vol. 44, no. 4, pp. 1201–1208, Nov. 1998, doi: 10.1109/30.735818.

[57] A. K. Pal and A. Sar, "An efficient codebook initialization approach for LBG algorithm," 2011, *arXiv:1109.0090*.

[58] Rowe and Valin. *Codec 2 Rowetel*. Accessed: Apr. 9, 2023. [Online]. Available: https://www.rowetel.com/?pageid=452

[59] S. Erhardt, T. Kurin, F. Lurz, R. Weigel, and A. Koelpin, "An open-source speech codec at 450 bit/s with pseudo-wideband mode," in *Proc. 16th Eur. Radar Conf. (EuRAD)*, Oct. 2019, pp. 413–416, doi: 10.23919/EuMC.2019.8910691.

[60] R. Gold, "Optimal binary sequences for spread spectrum multiplexing (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 4, pp. 619–621, Oct. 1967, doi: 10.1109/TIT.1967.1054048.

[61] D. G. Rowe, "Techniques for harmonic sinusoidal coding," Ph.D. dissertation, School Phys. Electron. Syst. Eng., Fac. Inf. Technol., Univ. South Australia, 1997.

[62] B. S. Yarman, Ü. Güz, and H. Gürkan, "On the comparative results of," *AEU Int. J. Electron. Commun.*, vol. 60, no. 6, pp. 421–427, Jun. 2006, doi: 10.1016/j.aeue.2005.08.003.

[63] J.-M. Valin and J. Skoglund, "A real-time wideband neural vocoder at 1.6kb/s using LPCNet," in *Proc. Interspeech*, Sep. 2019, pp. 3406–3410, doi: 10.21437/Interspeech.2019-1255.

[64] V. Andem, "A cryptanalysis of the tiny encryption algorithm," Ph.D. dissertation, Dept. Comput. Sci., Univ. Alabama, Tuscaloosa, AL, USA, Jan. 2003.

[65] H. Lei, Y. Zhao, Y. Dai, and Z. Wang, "A secure voice communication system based on DSP," in *Proc. 8th Control, Autom., Robot. Vis. Conf.*, Jun. 2004, pp. 132–137, doi: 10.1109/ICARCV.2004.1468811.

[66] K. W. Tang and W. K. S. Tang, "A chaos-based secure voice communication system," in *Proc. IEEE Int. Conf. Ind. Technol.*, May 2005, pp. 571–576, doi: 10.1109/ICIT.2005.1600703.

[67] J. Liu and Y. Cheng, "The design and simulation of real-time encryption algorithm for mobile terminal voice source," in *Proc. Int. Conf. Comput. Syst., Electron. Control (ICCSEC)*, Dec. 2017, pp. 1016–1021, doi: 10.1109/ICCSEC.2017.8446839.

[68] M. A. Riyadi, N. Pandapotan, M. R. A. Khafid, and T. Prakoso, "FPGA-based 128-bit chaotic encryption method for voice communication," in *Proc. Int. Symp. Electron. Smart Devices (ISESD)*, Oct. 2018, pp. 1–5, doi: 10.1109/ISESD.2018.8605446.

[69] M. A. Riyadi, M. R. A. Khafid, N. Pandapotan, and T. Prakoso, "A secure voice channel using chaotic cryptography algorithm," in *Proc. Int. Conf. Electr. Eng. Comput. Sci. (ICECOS)*, Oct. 2018, pp. 141–146, doi: 10.1109/ICECOS.2018.8605229.

[70] N. Hayati, Y. Suryanto, K. Ramli, and M. Suryanegara, "End-to-end voice encryption based on multiple circular chaotic permutation," in *Proc. 2nd Int. Conf. Commun. Eng. Technol. (ICCET)*, Apr. 2019, pp. 101–106, doi: 10.1109/ICCET.2019.8726890.

[71] H. Pardede, K. Ramli, Y. Suryanto, N. Hayati, and A. Presekal, "Speech enhancement for secure communication using coupled spectral subtraction and Wiener filter," *Electronics*, vol. 8, no. 8, p. 897, Aug. 2019, doi: 10.3390/electronics8080897.

[72] M. Torcoli, T. Kastner, and J. Herre, "Objective measures of perceptual audio quality reviewed: An evaluation of their application domain dependence," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 1530–1541, 2021, doi: 10.1109/TASLP.2021.3069302.
[73] P. C. Loizou, "Speech quality assessment," in *Multimedia Analysis, Processing and Communications*, vol. 346, J. Kacprzyk, W. Lin, D. Tao, J. Kacprzyk, Z. Li, E. Izquierdo, and H. Wang, Eds. Berlin, Germany: Springer, 2011, pp. 623–654, doi: 10.1007/978-3-642-19551-8_23.

ALBERTUS ANUGERAH PEKERTI received the bachelor's and master's degrees in electrical engineering from the Institut Teknologi Bandung (ITB), in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree in voice communication security with ITB. He is an Academic and a Researcher with ITB. Additionally, he contributes to projects related to smartphones and laptops, actively advancing the electronics industry in Indonesia.

ARIF SASONGKO received the bachelor's and master's degrees from Institut Teknologi Bandung (ITB), Bandung, Indonesia, in 1998 and 2001, respectively, and the Ph.D. degree in nano/microelectronics from Joseph Fourier University, Grenoble, France, in 2005. He is currently an Accomplished Academic, a Researcher, and a Lecturer with ITB. Notably, he has contributed to projects related to efficient and multi-standard cryptography. As the Secretary of the Innovation Development and Entrepreneurship Institute (LPiK-ITB), he augment the impact of research and development in higher education. His research interests include hardware accelerators for cryptography, embedded systems, and VLSI architecture. His research articles cover topics, such as FPGA-based controllers, smart cards, communication systems, and cryptography.

ADI INDRAYANTO received the bachelor's degree from the Institut Teknologi Bandung (ITB), Bandung, Indonesia, in 1988, the master's degree (M.Sc.) from the University of Manitoba, Canada, in 1992, and the Ph.D. degree from The University of Manchester, U.K., in 2005. He joined the Electronics Research Group and the Microelectronics Center. Beyond academia, he has actively contributed to the electronics industry in Indonesia. He is currently an Accomplished Academic, a Researcher, and a Lecturer with ITB. His expertise extends to development projects related to smartphones and laptops. As part of his commitment to advancing technology, he collaborates with industry partners and government agencies to drive innovation and enhance the country's electronics ecosystem.