An Adaptive Speech Homomorphic Encryption Scheme Based on Energy in Cloud Storage

Qiu-Yu Zhang and Yu-Jiao Ba
(Corresponding author: Qiu-Yu Zhang)

School of Computer and communication, Lanzhou University of Technology
No. 287, Lan-Gong-Ping Road, Lanzhou 730050, China
Email: zhangqylz@163.com, bayujiaolut@163.com
(Received Oct. 26, 2021; Revised and Accepted Apr. 15, 2022; First Online Apr. 23, 2022)

Abstract

Aiming at the problems of the traditional speech encryption scheme in cloud storage, such as security risks, excessive communication consumption, low robustness to resist multiple types of attacks, and low efficiency of the speech homomorphic encryption scheme, an adaptive speech homomorphic encryption scheme based on energy in cloud storage was proposed. Firstly, by comparing the threshold of speech energy, the improved Adaboost algorithm is used to design an adaptive classifier. Then, the speech data is divided into the sound and silent parts according to the energy threshold. Secondly, the BGV homomorphic encryption algorithm is used to encrypt the sound part of the information. Then, the Paillier homomorphic encryption algorithm is used to encrypt the silent part of the information. Finally, the two parts of ciphertext are combined to realize ciphertext domain operation and adaptive decryption. The experimental analysis shows that the proposed scheme has good encryption and decryption efficiency and low ciphertext expansion and can resist various attacks (including statistical, entropy, and chosen-plaintext attacks).

Keywords: Adaptive Classifier; BGV Homomorphic Encryption; Homomorphic Encryption; Paillier Homomorphic Encryption

1 Introduction

With the rapid development of the cloud storage and Internet technology, more and more users choose to store multimedia data uploaded to the cloud. Cloud storage separates the ownership and management rights of data [20,26], which makes the security of multimedia data and the protection of personal privacy in cloud storage attract people's attention [4, 10]. Therefore, ensuring the security of speech content has become an important research issue. As a standard and an effective technology to protect the security of digital multimedia information content, speech encryption plays an important role in the applications of speech retrieval [18].

At present, common speech encryption methods include homomorphic encryption [6, 15, 16], chaotic mapping encryption [1] (including Lorenz mapping, Logistic mapping, Henon mapping, etc.), scrambling encryption, RSA, AES, etc. Since homomorphic encryption can not only protect data privacy, but also allow the operation of encrypted data (such as simple addition, subtraction and multiplication). It can support the extraction of effective features from encrypted data [20]. By analyzing the full homomorphic encryption scheme [17] and the applications of homomorphic encryption, homomorphic encryption is more and more widely used in the field of data encryption [9] and multimedia data encryption (such as image data [25], speech data [6, 15, 16, 20], etc.). Speech homomorphic encryption has become one of the key components of secure speech storage in public cloud computing.

Adaptive control [13] can automatically adjust the processing method, processing sequence, processing parameters, boundary conditions or constraints according to the data characteristics of the processed data to adapt to the statistical distribution characteristics and structural characteristics of the processed data, so as to obtain the best processing effect and realize random optimization. When parameters are estimated from massive speech data according to the optimal scheme, adaptive control can abandon the local optimal in the solution space and tends to the global optimal [8], which is more suitable for massive speech classification in the cloud environment. In recent years, the multimedia data combined with adaptive algorithms, encryption algorithms and other methods have made great achievements in the fields of data encryption [11], image encryption, speech encryption [23] and video encryption. Adaptive technology has also made great achievements in the research of speech processing fields such as semantic analysis, speech retrieval [3], audio watermarking, etc. Adaptive encryption methods can generate speech residual similar to any other speech signal, which can further protect the security of speech data [7, 14, 21]. Therefore, for the practical

applications such as speech data security and ciphertext speech retrieval in cloud storage environment, it is of great significance to research adaptive homomorphic encryption methods suitable for the characteristics of speech signals.

To solve the above problems, an adaptive speech homomorphic encryption scheme based on energy in cloud storage is proposed to ensure the security of cloud data and adaptive encryption for speech signal characteristics. The main contributions of this work are as follows:

1) In order to improve the efficiency of the encryption scheme, the parameters are estimated by comparing the threshold of speech energy, and an adaptive classifier is designed to reduce the complexity of low-energy data encryption and strengthen the robustness against various types of attacks (including statistical attack, entropy attack, and chosen-plaintext attack).

2) By using adaptive parameters to classify speech data based on energy, the original speech data is divided into multiple data blocks according to the energy threshold, and the data blocks are numbered. The strong correlation between adjacent speech blocks is reduced by different homomorphic encryption for the data blocks of the sound and silent part respectively.

3) Parallel adaptive encryption and decryption. Different homomorphic encryption is carried out for the speech data of the sound and silent part after the adaptive selection. Similarly, the parallel decryption is carried out after the adaptive ciphertext differentiation, which minimizes the amount of computation while maintaining a certain computational complexity and improves the encryption efficiency of the encryption scheme.

The rest of the paper is arranged as follows. Section 2 analyzes relevant research work in detail. Section 3 gives the system model of the encryption scheme, and describes the adaptive speech homomorphic encryption algorithm and its processing process in detail. Section 4 analyzes the encryption performance of the encryption scheme. Section 5 gives the experimental results and the performance analysis compared with existing schemes. Finally, we conclude our work in Section 6.

2 Related work

Speech encryption is one of the key steps in ciphertext speech retrieval. In recent years, while exploring homomorphic speech encryption [6, 15, 16], chaotic speech encryption [18] and other encryption schemes, many scholars have proposed adaptive encryption technology for speech data [7, 21] to protect the privacy and security of data in the cloud environment [19]. At present, the combination of multimedia data and adaptive methods has made considerable achievements in image and speech encryption [25]. Adaptive speech encryption is robust to multiple types of attackers (including ciphertext attackers and plaintext attackers), it is not an encryption algorithm, but a combination of speech signal processing and multiple encryption technologies.

Aiming at the security of speech data in cloud storage, Shi [16] proposed a digital speech encryption scheme based on homomorphic encryption by using the symmetric key cryptosystem (MORE-method) with probability statistics and complete homomorphism characteristics to encrypt speech signals, but this scheme has a large ciphertext expansion and cannot resist statistical analysis estimation. In order to solve the above problems, Shi [15] improved the scheme in 2019 and proposed a probability statistics addition homomorphic encryption scheme with small expansion of ciphertext. This scheme limits the expansion of ciphertext data and resists statistical analysis attacks. Imran [6] proposed the El-Gamal speech homomorphic encryption scheme. The security of this scheme is based on computing discrete logarithmic moduli of large prime numbers, which would take thousands of years for attackers to crack. In order to solve the problems of the above schemes, this paper will use the existing efficient homomorphic encryption scheme to encrypt and decrypt the speech data. BGV (Brakerski-Gentry-Vaikuntanathan) homomorphic encryption [2,5] is the most efficient scheme among the current mainstream homomorphic encryption algorithms. Using homomorphic encryption can realize the operation of addition, subtraction and multiplication in the encryption domain, and can further realize the feature extraction operation in the ciphertext domain. Paillier [12] algorithm is the most commonly used and practical additive homomorphic encryption algorithm, which has been applied in many application scenarios. In this paper, BGV algorithm and Paillier algorithm are used for encryption, and the ciphertext can be filtered by a step of multiplication before decryption to support different decryption operations.

In order to further improve the data security in the cloud, Shahadi [14] proposed an adaptive speech encryption method, combining the biggest advantages of cryptography and steganography, and adapting the wavelet coefficients of encrypted speech to any other speech signal coefficients. Neither send the encrypted content of the secret speech nor extend the bandwidth of the transmission message. The ciphertext speech retrieved by the scheme shows high speech quality and is robust to both ciphertext-only and plaintext attacks. Jahanshahi [7] designed a robust adaptive control scheme to encrypt speech. The adaptive mechanism was used to estimate the unknown parameters of the system, and the output of the proposed adaptive mechanism was used in the control scheme to achieve a fractional order system of speech encryption. But its efficiency limits the application of the system to a great extent. In order to improve the efficiency of adaptive encryption, Tutueva [21] proposed a new pseudo-random generation method based on the concept of adaptive symmetric chaotic mapping. Adaptive coefficients combined with chaos-based Pseudo-Random

Number Generator (PRNG) are easier to implement than existing chaos-based improved generators, and chaotic maps with adaptive symmetry are suitable for stream ciphers. However, the various attacks on cryptographic systems based on adaptive mapping are not discussed in this paper, which is not enough to prove their security.

In summary, the existing adaptive encryption schemes and homomorphic encryption schemes are mostly used in image fields, and the existing speech adaptive schemes are mostly combined with chaotic encryption and scrambling encryption. There are relatively few studies on sensitive speech data, and the combination of homomorphic encryption and adaptive control is rarely applied to speech data. To solve these problems, an energy-based adaptive speech homomorphic encryption scheme is proposed, which can implement adaptive selective encryption for speech data, making the encryption more efficient and less data expansion, and having strong robustness to a variety of types of attackers.

3 The Proposed Scheme

3.1 System Model

Figure 1 shows the system model in the proposed scheme. The system model consists of four entities: data owner (DO), cloud server (CS), adaptive classifier (AC), and retrieval user (RU).

As shown in Figure 1, the components of the system model accomplish the following:

Data Owner (DO): DO owns local speech data $S = S_1, S_2, \ldots, S_n$. To ensure the privacy and security of speech data, the speech data is encrypted after adaptive classification, and the ciphertext speech data $C = C_1, C_2, \ldots, C_n$ is obtained. Where n represents the number of speech data. Finally, the generated ciphertext speech data C is outsourced to CS for storage.

Adaptive Classifier (AC): AC is an adaptive classifier for speech data generated by threshold estimation. In order to improve the encryption performance, the original speech is classified into sound data $S' = \{S'_1, S'_2, \ldots, S'_n\}$ and silent data $S'' = \{S''_1, S''_2, \ldots, S''_n\}$ by adaptive selection, and the ciphertext speech data $C' = \{C'_1, C'_2, \ldots, C'_n\}$ and $C'' = \{C''_1, C''_2, \ldots, C''_n\}$ are obtained after parallel encryption. Finally, the ciphertext speech data $C = C_1, C_2, \ldots, Cn$ is generated.

Cloud Server (CS): CS stores ciphertext speech data C uploaded by DO and performs ciphertext calculations on C to obtain the new ciphertext C^*. When receiving RU's search request, CS returns the query result C^* to RU.

Retrieval User (DU): DU decrypts the plaintext speech data by using the key sent by DO after receiving the query result from CS.

3.2 Adaptive Classifier

In the process of processing and analysis, adaptive control automatically adjusts the processing method, processing sequence, processing parameters, boundary conditions or constraints according to the data characteristics of the processed data to adapt to the statistical distribution characteristics and structural characteristics of the processed data, so as to obtain the best processing effect. Adaptive control usually uses the adaptive algorithm to generate online estimation of unknown parameters.

The adaptive boosting (Adaboost) algorithm [24] is improved to combine multiple weak classifiers into a strong classifier in this paper. The principle of adaptive classifier is to adjust its parameters according to some criteria and adaptive algorithm to minimize the cost (objective) function of adaptive classifier and achieve the purpose of optimal equilibrium. Figure 2 shows the adaptive classifier (AC) model designed by Adaboost algorithm in the proposed scheme.

The dotted line in Figure 2 represents the iteration effect of different rounds, and each iteration adds a classification structure. The work of the i-th iteration is as follows:

1) Add weak classifier Y_i and weak classifier weight $\mathrm{Alpha}(i)$;

2) The weak classifier Y_i is trained by data set Data and data weight $\mathrm{W}(i)$, and its classification error rate is obtained, so as to calculate its weak classifier weight $\mathrm{Alpha}(i)$;

3) Combine each trained weak classifier Y_i into an adaptive classifier AC. After the training process of each weak classifier is over, if the final error rate is lower than the set threshold (this paper is set to 3%), then the iteration ends; if the final error rate is higher than the set threshold, then update the data weight to get $\mathrm{W}(i+1)$.

The basic principle of the adaptive classifier designed in this paper is to combine multiple weak classifiers (weak classifiers generally use single-layer decision trees) to make them a strong classifier. The algorithm adopts the idea of iteration. Only one weak classifier is trained for each iteration, and the trained weak classifier will participate in the use of the next iteration. That is, in the i-th iteration, there are a total of i weak classifiers, of which $i-1$ are already trained, and their various parameters are no longer changed. This time the i-th classifier is trained. The relationship of the weak classifier is that the i-th weak classifier is more likely to match the data that the first $i-1$ weak classifier did not match, and the final classification output is the comprehensive classification result of the i classifiers.

There are two kinds of weights in the Adaboost algorithm, one is the weight of the data, and the other is the weight of the weak classifier. Where the weight of the data is mainly used for the weak classifier to find the decision

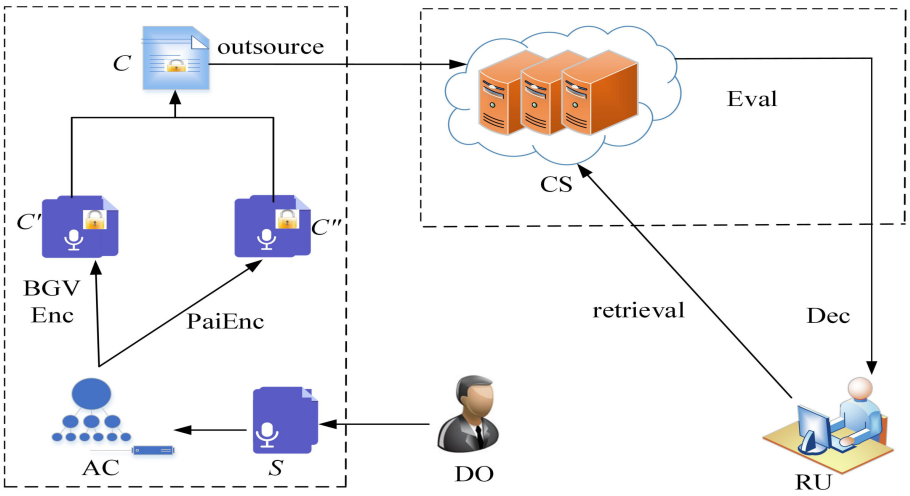


Figure 1: Energy-based adaptive speech homomorphic encryption system model

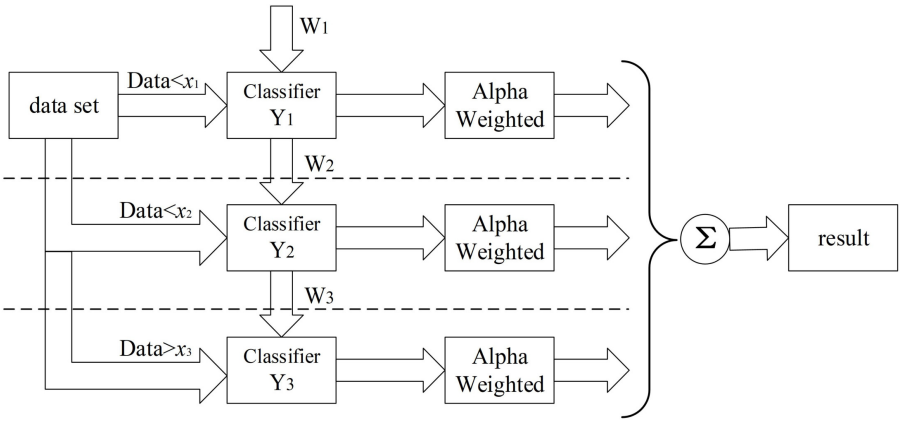


Figure 2: Adaptive classifier (AC) model

point with the smallest classification error. The weight of a weak classifier depends on its error rate. The lower the error rate, the higher the weight. In Adaboost, each weak classifier has its own threshold, and each weak classifier only focuses on a part of the entire dataset, so they must be combined to achieve the final classification.

3.3 Adaptive Homomorphic Encryption

Figure 3 shows the specific processing flow of the energy-based adaptive speech homomorphic encryption scheme. After the preprocessed original speech is classified by the designed adaptive classifier, the homomorphic encryption of different classes of speech data is performed in parallel. When the speech frame data belongs to the -1 category, it is subjected to BGV homomorphic encryption. When the speech frame data belongs to the $+1$ category, it is subjected to Paillier homomorphic encryption.

The specific processing steps are as follows:

Step 1: Pretreatment. Read the original speech data and perform smoothing processing to obtain the value range of the data.

Step 2: Adaptive classification. The speech data is divided into -1 category and $+1$ category through the trained adaptive classifier.

Step 3: Batch packaging. Use the Chinese remainder theorem (CRT) and single instruction multiple data (SIMD) to pack the classified data into a one-dimensional array to implement parallel encryption operations.

Step 4: Adaptive homomorphic encryption. Perform BGV homomorphic encryption on category -1 sound data; perform Paillier homomorphic encryption on category $+1$ silent data.

The security of the BGV homomorphic encryption algorithm is based on the shortest vector problem (SVP). The algorithm establishes a new method to construct a FHE scheme with a fixed circuit depth without Gentry's bootstrapping (able to evaluate circuits of arbitrary polynomial size). The ciphertext multiplication operation will

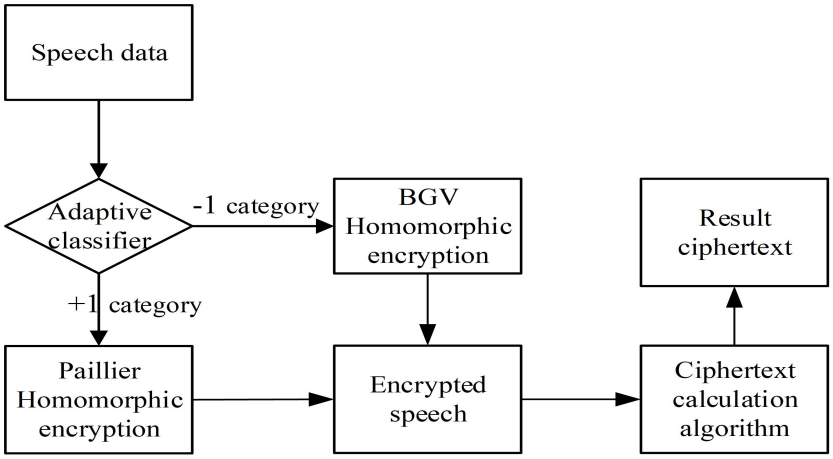


Figure 3: Adaptive speech homomorphic encryption processing flowchart

cause the explosive growth of the ciphertext dimension, resulting in the solution can only perform a constant number of multiplication operations. But the algorithm can use key exchange technology and module exchange technology to solve this problem: the key exchange technology can control the dimensional expansion of the ciphertext vector. After the ciphertext is calculated, the expanded ciphertext dimension is restored to the original ciphertext dimension through key exchange; Modular switching technology can replace the Bootstrapping process in the Gentry scheme to control the noise increase generated by the homomorphic operation of ciphertext. Therefore, after each ciphertext multiplication operation, it is necessary to reduce the dimensionality of the ciphertext through the key exchange technology, and reduce the noise of the ciphertext through the modular exchange technology, so that the next calculation can be continued.

The security of the Paillier homomorphic encryption algorithm is based on the complex remaining difficult problems, and it is a public key encryption algorithm. Before encryption and decryption, public keys n and g that can be used for encryption must be generated. n is the product of two large prime numbers of similar size:$n = p \cdot q$. g is a semi-random number, and its order must be in $Z^*_{n^2}$, that is, the order of g modulo n_2 must be a multiple of n. The public key used for the actual encryption and decryption operation process is (n, g). With the release of the public key, anyone can use the public key to encrypt data and pass the ciphertext to the private key holder.

3.4 Encryption and Decryption Scheme

Figure 4 is the processing flow of this text encryption and decryption scheme, which mainly includes three parts of processing work. First, the adaptive classifier AC is designed, and the classifier model used for homomorphic encryption is trained to perform adaptive homomorphic encryption. Then the data owner DO sends the speech database to AC for adaptive classification, adaptively encrypts the -1 and +1 speech data, stores the encrypted speech in the cloud, and CS performs outsourcing calculation and retrieval. Finally, the authorized user RU decrypts the retrieved speech returned from the cloud to obtain the decrypted speech.

The definitions of symbols used in the proposed scheme are shown in Table 1.

Table 1: Symbol definitions

Symbol	Definitions
$S = S_1, S_2, \ldots, S_n$	Speech data set
$C = C_1, C_2, \ldots, C_n$	Encrypted speech data set
$SK = sk_1, sk_2$	Private key sk_1, sk_2
$PK = pk_1, pk_2$	Public key pk_1, pk_2
$EVK = evk_1, evk_2$	Calculation key evk_1, evk_2
i	Number of iterations
N	Total number of sample data

The proposed adaptive speech homomorphic encryption and decryption scheme consists of 5 algorithms: **Setup**, **GenKey**, **Enc**, **Eval**, **Dec**.

1) Adaptive classifier algorithm. $m_i \leftarrow$**Setup**(S, i). Input speech sample data set $S = S1, S2, \ldots, Sn$ and the number of iterations i, Generate a strong classifier $G(x)$ and output the classification result m_i.

2) Key generation algorithm. $sk, pk, evk \leftarrow$**GenKey** (λ, p, q). This algorithm is a probabilistic key generation algorithm. Enter λ, p, q, and return the private key (sk_1, sk_2), public key (pk_1, pk_2) and ciphertext calculation key (evk_1, evk_2).

3) Speech encryption. $c \leftarrow$**Enc**(pk, m). This algorithm is a probabilistic algorithm. Input the public key pk and the speech data m, and return the ciphertext speech data c.

4) Ciphertext calculation algorithm. $c' \leftarrow$ **Eval** (evk, C, c). Input the ciphertext calculation key evk,

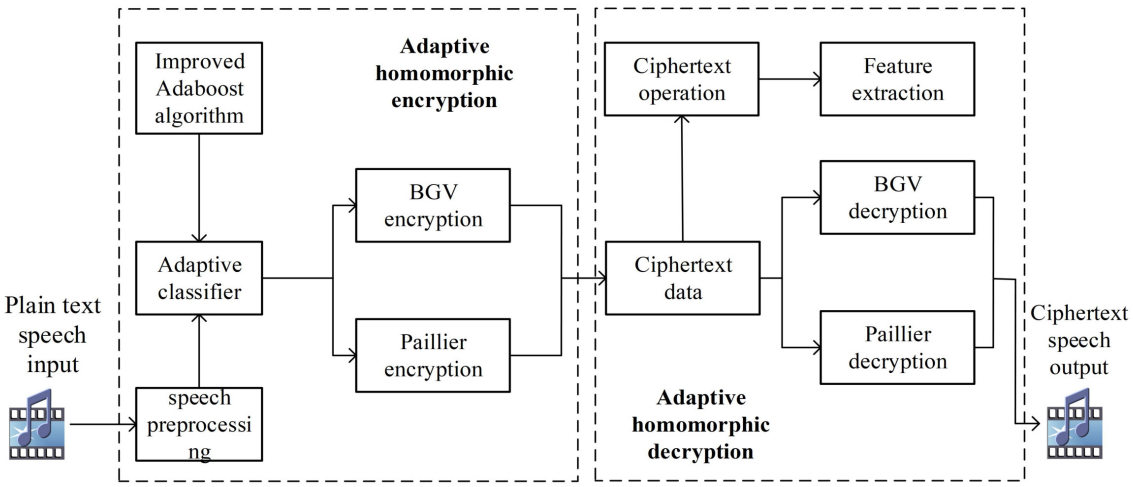


Figure 4: Adaptive speech homomorphic encryption and decryption processing flow

circuit C and ciphertext c, and output the ciphertext calculation result c'.

5) Speech decryption. $m' \leftarrow \mathbf{Dec}(sk, c')$. This algorithm is a deterministic algorithm. Enter the private key sk and the ciphertext calculation result c', and return the decrypted speech m'.

The processing process of the adaptive speech homomorphic encryption and decryption system is as follows:

Step 1: Adaptive classification. The parameter adaptive algorithm is combined with the discrete system to obtain the estimated parameters.

After 3 iterations, the process of implementing adaptive classification is as follows:

1: Initialize the weight distribution of the training data (each sample). Each training sample is initialized with the same weight $W_1 = 1/N$.

2: Perform multiple iterations, $i = 1, 2, 3$, i represents the number of iterations. The adaptive classifier designed in this paper iterates 3 times in total.

1) Use the training sample set with the weight distribution $w_i (i = 1, 2, 3)$ for learning, and get the weak classifier $\mathrm{Y}_i(x)$. The criterion is shown in Equation (1). The error function of the weak classifier is the smallest, that is, the sum of the weights corresponding to the wrong samples is the smallest.

$$\varepsilon_i = \sum_{n=1}^{N} w_n^{(i)} I\left(\mathrm{Y}_i\left(x_n\right) \neq t_n\right) \tag{1}$$

$$\mathrm{Y}_i\left(x\right) : \chi \rightarrow \{-1, +1\} \tag{2}$$

2) Calculate the weight of the weak classifier $\mathrm{Y}_i(x)$, and the weight $w(i)$ indicates the importance of $\mathrm{Y}_i(x)$ in the final classifier.

$$w(i) = \frac{1}{2} \log \frac{1 - \varepsilon_i}{\varepsilon_i} \tag{3}$$

This value increases as ε_i decreases. That is, a classifier with a small error rate is more important in the final classifier.

3) Update the weight distribution of the training sample set. Used in the next iteration. The weight of the misclassified samples will increase, while the weight of the correct score will decrease.

3: After the iteration is completed, the combined weak classifier is the final adaptive classifier AC.

$$\mathrm{AC} = \sum_{i=1}^{3} w(i) \mathrm{Y}_i(x) \tag{4}$$

Step 2: Key generation. This scheme is a multi-key homomorphic encryption system, and the key generation algorithm is composed of two key generation functions.

1: BGV key generation

Randomly select an element on χ as the private key: $sk_1 = s \leftarrow \chi$, take $\mathbf{a} \leftarrow R_q, \mathbf{e} \leftarrow \chi$, and get the public key $pk_1 = \left([-(\mathbf{a}s + \mathbf{e})]_q, \mathbf{a}\right)$. Thus, the key pair (sk_1, pk_1) is obtained.

2: Paillier key generation

In the case of the same key length, the keys $g = n + 1, \lambda = \varphi(n), \mu = \varphi(n) - 1 \mathrm{mod} n$ can be quickly generated, where $\varphi(n)$ refers to the Euler function, and its value is $(p-1) \times (q-1)$. Get the key pair (sk_2, pk_2), the public key $pk_2 = (n, g)$, and the private key $sk_2 = (\lambda, \mu)$.

Step 3: Speech encryption. Use BGV homomorphic encryption to encrypt the sound part of the speech information; Paillier homomorphic encryption to encrypt the silent part of the speech information.

1: BGV encryption

Encryption algorithm of BGV homomorphic encryption: $c_1 = ([\Delta \cdot m + p[0]\mathbf{u} + \mathbf{e}_1]_q, [p[1]\mathbf{u} + \mathbf{e}_2]_q)$. Get the ciphertext c_1.

2: Paillier encryption

1) The plaintext m is a positive integer greater than or equal to 0 and less than n.
2) Randomly select r to satisfy $0 < r < n$ and $r \in Z_{n^2}^*$ (a sufficient condition is that r and n are relatively prime). $r \in Z_{n^2}^*$ means that r has a multiplicative inverse element in the remainder of n^2. Get the ciphertext $c_2 = g^m r^n \bmod n^2$.

Step 4: Ciphertext calculation. The ciphertext calculation algorithm for homomorphic encryption.

1: BGV ciphertext calculation

Satisfy full homomorphisms. Input the ciphertext calculation key evk_1, the circuit C and the ciphertext c_1, and the output is the ciphertext calculation result c_1'.

2: Paillier ciphertext calculation

It satisfies add homomorphism, that is, the multiplication of ciphertext is equal to the addition of plaintext: $D(E(m1) \cdot E(m2)) = m1 + m2$. Since it supports additive homomorphism, Paillier algorithm can also support multiplication homomorphism, that is, multiplication of ciphertext and plaintext. The ciphertext calculation is as follows:

$$\begin{cases} c1 \equiv g^{m1} \cdot r_1^n \bmod n^2 \\ c2 \equiv g^{m2} \cdot r_2^n \bmod n^2 \end{cases}$$
$$\Rightarrow c1 \cdot c2 \equiv g^{m1} \cdot g^{m2} \cdot {r_1}^n \cdot {r_2}^n \bmod n^2$$
$$\Rightarrow c1 \cdot c2 \equiv g^{m1+m2} \cdot (r_1 \cdot r_2)^n \bmod n^2$$

$c_1 \cdot c_2$ can be regarded as $m = m1 + m2$ encrypted ciphertext, and the decryption result of $c_1 \cdot c_2$ is m.

Step 5: Decryption steps. Use adaptive ciphertext data selection to filter the ciphertext and perform homomorphic decryption.

1: BGV decryption

Input private key $s = sk_1$, ciphertext c_1, output decrypted plaintext $m_1' = \left[\frac{t}{q}[c_1[0] + c_1[1] \cdot s]_q\right]_t$.

2: Paillier decryption

Input private key $s = sk_2$, ciphertext c_1, output decrypted plaintext $m_2 = L\left({c_2}^\lambda \bmod n^2\right)\mu \bmod n$.

Step 6: Speech reconstruction. The decrypted speech matrix is restored to a complete decrypted speech.

4 Encryption Performance Analysis

In order to evaluate the performance and efficiency of the proposed speech encryption scheme, some unequal-length speeches in the Chinese speech database THCHS-30 [22] opened by Tsinghua University were selected as the test speech for this experiment for encryption and decryption. Among them, S1.wav is a 8s speech, S2.wav is a 6s speech, and S3.wav is a 4s speech. Use PyCharm platform tools to perform speech preprocessing to obtain adaptively classified speech data as a database to realize adaptive homomorphic encryption of speech data.

Experimental hardware environment: Intel(R) Core (TM) i5-8250U CPU, 1.80GHz, RAM 12GB.

Software Environment: Windows 10, PyCharm, Matlab R2017b.

4.1 Correlation Analysis

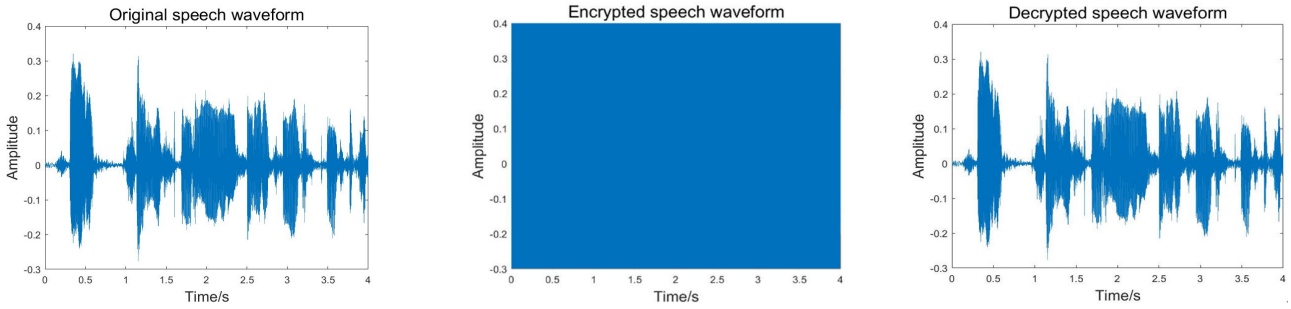
Correlation analysis [15], as a data statistical method, is widely used in the performance evaluation of speech encryption algorithms. If the value of the correlation coefficient is around +1 or −1, it indicates that the two speech signals are highly correlated; if the value of the correlation coefficient of the two speech signals is around 0, it means that the correlation between the two speech signals is extremely poor. The correlation coefficient expression is shown in Equation (5), and its correlation expression is as follows:

$$r_{xk} = \frac{C(x,k)}{\sqrt{V(x)}\sqrt{V(k)}} \tag{5}$$
$$C(x,k) = \frac{\sum (x-\bar{x})(k-\bar{k})}{N-1} \tag{6}$$
$$V(x) = \frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2 \tag{7}$$
$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{8}$$

where $\bar{x}$, $\bar{k}$ are the mean values of the original speech signal x and the encrypted speech signal k respectively; $C(x,k)$ is the covariance between the original speech signal x and the encrypted speech signal k; $V(x)$ and $V(k)$ represent the variance between the original speech x and the encrypted speech k.

Figure 5 shows the waveform of original speech, encrypted speech and decrypted speech of the S3.wav as an example. Figure 5(a), Figure 5(b) and Figure 5(c) are waveform diagrams of original speech, encrypted speech and decrypted speech respectively.

It can be seen from Figure 5(b) that no speech waveform features can be seen in the encrypted speech, so the encryption effect is good. In order to further reflect the anti-statistical analysis attack performance of the encryption algorithm, the correlation coefficient before and after the speech encryption is analyzed, and the Pearson correlation coefficient is calculated by Equation (5) to measure the correlation between the signals. Table 2 shows



(a) Original speech (b) Encrypted speech (c) Decrypted speech

Figure 5: Waveform diagram of original speech, encrypted speech and decrypted speech

the correlation analysis of the original speech, encrypted speech and decrypted speech of different durations. Figure 6 shows the correlation comparison before and after speech encryption.

Table 2: Correlation analysis of speech

File	Original & Encrypted	Original & Decrypted
S1.wav	-0.0043	0.9890
S2.wav	-0.0032	0.9920
S3.wav	0.0018	0.9983
Average	0.0031	0.9931

If the correlation coefficient is around $+1$ or -1, it indicates that the two speech signals are highly correlated. If the correlation coefficient between two speech signals is around 0, it means that the correlation between the two speech signals is extremely poor. It can be seen from Table 2 and Figure 6 that the correlation coefficient between the original speech and the encrypted speech is close to 0, indicating that the original speech and the encrypted speech are unrelated and the speech encryption performance is good. The correlation coefficient between the original speech and the decrypted speech is between -1 and $+1$, indicating that the recovery and reconstruction performance of the speech is extremely strong, and basically can achieve lossless recovery.

4.2 SNR and SNRseg

Signal-to-noise ratio (SNR) [16], as one of the most common and direct methods to verify the performance of data encryption algorithms. It is mainly used to measure the noise content and distortion degree of signals in encrypted data, and is widely used in multimedia data encryption. The expression of SNR is shown in Equation (9).

$$SNR=10\times\log_{10}\frac{\sum_{i=0}^{L}x_i^2}{\sum_{i=1}^{L}[x_i-y_i]^2} \tag{9}$$

where L represents the number of samples; x_i stands for the original speech signal; y_i stands for encrypted speech signal. The higher the SNR is, the less noise is generated. Generally speaking, the smaller the SNR is, the greater the noise in the encrypted signal is, the higher the distortion degree is and the better the encryption quality is.

Segmented SNR (SNRseg) [16] is the average of short-frame SNR. This is one of the widely used objective evaluation measurement methods, which can be used to estimate the quality of the speech signal. The lower the SNRseg value, the higher the encryption noise and the better the encryption effect. The expression of the SNRseg function is shown in Equation (10):

$$SNR_{seg}=\frac{10}{M}\times\sum_{m=0}^{M-1}\log_{10}\frac{\sum_{n=Lm}^{Lm+L-1}x_i^2}{\sum_{n=Lm}^{Lm+L-1}[x_i-y_i]} \tag{10}$$

where M represents the number of frames in the speech signal.

Table 3: SNR and SNRseg of encrypted speech

File	SNR (dB)	SNRseg (dB)
S1.wav	-40.1540	-41.1022
S2.wav	-38.4883	-40.5567
S3.wav	-51.0261	-53.6930
Average	-44.8495	-45.1173

The proposed scheme performs SNR and SNRseg tests on encrypted speech data of different durations. It can be seen from Table 3 that the SNR and SNRseg values of the encrypted speech obtained from the experimental results are lower, indicating that the proposed encryption scheme has higher encryption quality and stronger security.

4.3 Security Analysis

In general, the security analysis must be satisfied when a new encryption scheme is proposed. A perfect encryption scheme should be robust against all kinds of cryptanalysis attacks (i.e., statistical attack, entropy attack, chosen-plaintext attack, etc). Therefore, this paper conducts some security analyses to demonstrate the effectiveness of the proposed algorithm.

1) Statistical Attack

If the encryption performance of a speech encryption system is well, the encrypted speech statistics

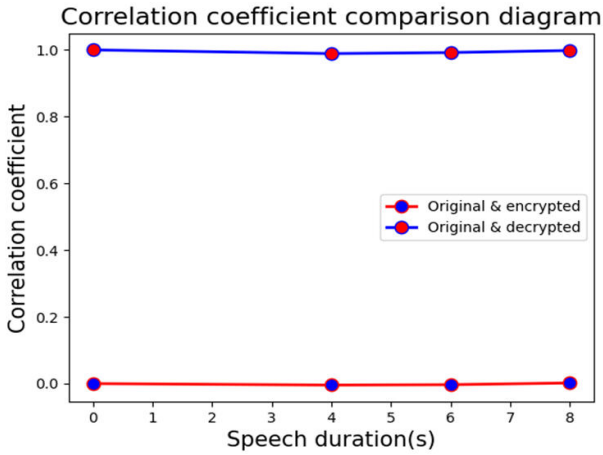


Figure 6: Correlation comparison before and after speech encryption

histogram [18] should be evenly distributed. This section takes S3.wav as an example for encryption performance analysis. Figure 7 shows the amplitude histogram of the original speech and the encrypted speech.

It can be seen from Figure 7 that the amplitude histogram of the original speech in Figure 7(a) has irregular statistical features. The amplitude histogram distribution of encrypted speech in Figure 7(b) is relatively stable without large ups and downs, which has a good masking effect on the statistical features of the speech data. It can be seen from Figure 7(b) that the encrypted speech data has poor correlation and little statistical information, indicating that the proposed encryption scheme is sufficient to resist statistical attack.

2) Entropy Attack

Information entropy analysis [10] is mainly used for the error rate of encrypted speech data. Generally, the value of information entropy is proportional to the error rate of speech. The higher the information entropy of encrypted speech data, the better the effect of speech encryption. The calculation expression of information entropy is shown in Equation (11):

$$H = -\sum_{K=0}^{S} p(k)\log_2 p(k) \tag{11}$$

where $p(k)$ is the input speech data, S represents the number of sampling points.

For each speech file, if the entropy value of the encrypted speech data is close to 16, it indicates that the speech encryption system has better encryption effect and higher security. Table 4 shows the calculation of information entropy for speech data of different durations.

It can be seen from Table 4 that the information entropy of encrypted speech data is basically close to the expected value of 16, indicating that the proposed encryption scheme has high security and is sufficient to resist entropy attack.

Table 4: Information entropy analysis of speech

File	Original speech	Encrypted speech
S1.wav	11.6241	15.4649
S2.wav	11.7700	15.6166
S3.wav	12.2390	15.7770
Average	11.5956	15.6594

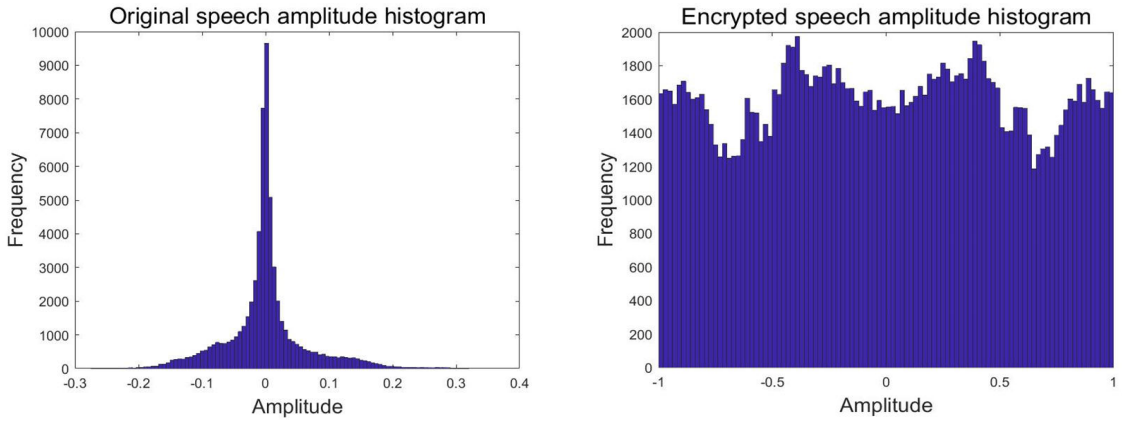
3) Chosen-plaintext Attack

The number of samples change rate (NSCR) [16] is an evaluation index of chosen-plaintext attack, which is widely used in the field of speech encryption. It reflects the proportion of the data points in the same position of two speech data to the whole data point. NSCR reflects the proportion of data points that are not equal in the same position of two speech data that are not equal to the entire data point. If NSCR is approximately equal to 100%, it is considered that the encryption algorithm has high performance and can resist various plaintext attacks. Table 5 shows the sample rate of change for different durations of speech.

Table 5: NSCR of speech

File	NSCR (%)
S1.wav	100
S2.wav	99.999
S3.wav	99.996
Average	99.998

It can be seen from Table 5 that the NSCR values obtained by the proposed scheme are all close to 100%, indicating that the encrypted speech data



(a) Original speech (b) Encrypted speech

Figure 7: Amplitude histogram of original speech and encrypted speech

sample points are diametrically opposite to the original speech, and the proposed scheme can effectively resist differential attacks.

5 Experimental Analysis

5.1 Adaptive Classification

The adaptive classifier proposed in this paper has obtained 3 decision points after 3 iterations. The decision point 1 that is less than (equal to) -1.0124 is divided into $+1$ category, the rest are divided into -1 category, and the weight of classifier Y_1 is 0.5. The decision point 2 is greater than (equal to) 1.0124 and divided into $+1$ category, the rest are divided into -1 category, and the weight of classifier Y_2 is 0.3. The decision point 3 is less than (equal to) 2.4492 is divided into $+1$ category, the rest are divided into -1 category, the weight of classifier Y_3 is 0.4.

This study loads the speech data as a sequence object, i.e. one-dimensional arrays, each with a time label. Confirm that the datasets have been loaded correctly with the summary data in Figure 8 and visualize these data as the dataset waveforms as shown in Figure 9.

By observing the density of the training data set, we can further understand the residual error of the data structure analysis model. Ideally, the distribution of residuals should follow a Gaussian distribution with zero mean. The residual is calculated by subtracting the predicted value from the actual value. Figure 10 shows the energy distribution of the sample data set used in this article, and Figure 11 shows the residual density of the designed classifier training model.

It can be seen from Figure 11 that the residual error of the adaptive classifier designed in this paper is basically Gaussian, the model has a small deviation, and the mean value basically tends to zero. If there is any autocorrelation in the residuals, it means there is an opportunity to improve the model. Ideally, if the model fits properly, no autocorrelation should be retained in the residuals. When training the classifier, first extract the first 50 data of the sample for autocorrelation analysis, and the autocorrelation coefficient is shown in Figure 12; when all the data of a sample is input to train the classifier, the autocorrelation coefficient is shown in Figure 13.

It can be seen from Figure 12 and Figure 13 that the autocorrelation coefficient of training a sample of all data is much lower than that of training a sample of a small amount of data. The more samples in the training set, the more the autocorrelation tends to zero. Figure 13 shows that all autocorrelations have been captured in the training model, and there is no autocorrelation in the residuals. Therefore, the training model has passed all the standards, and this model can be saved as an adaptive classifier for subsequent use.

5.2 Speech Encryption and Decryption Efficiency Analysis

The complexity of the speech encryption system and the efficiency of speech encryption and decryption are mutually restricted. Existing algorithms often ignore the speech encryption and decryption time when ensuring key security, and are not suitable for massive speech encrypted data. Table 6 shows the time efficiency analysis of encryption and decryption of speech data with different durations.

Table 6: Efficiency of speech encryption and decryption

File	Encryption time(s)	Decryption time(s)
S1.wav	15.7421	8.8643
S2.wav	11.3554	5.9435
S3.wav	8.1106	4.1652
Average	1.9560	1.0541

It can be seen from Table 6 that the encryption algorithm use about 1.9560 s to encrypt speech per second, and the decryption algorithm takes about 1.0541 s to encrypt speech per second, indicating that the proposed

```
count     64000.000000
mean       9036.312578
std        2038.179647
min           0.000000
25%        8356.000000
50%        9013.000000
75%        9525.000000
max       19552.000000
Name: 0, dtype: float64
```

Figure 8: Summary data

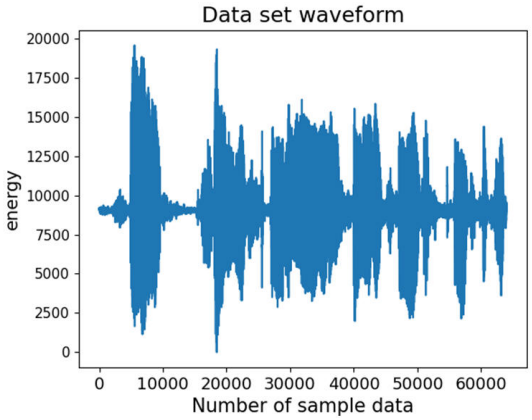


Figure 9: Dataset waveform

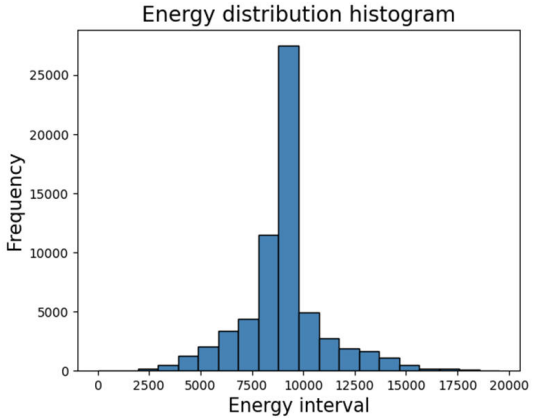


Figure 10: Histogram of energy distribution

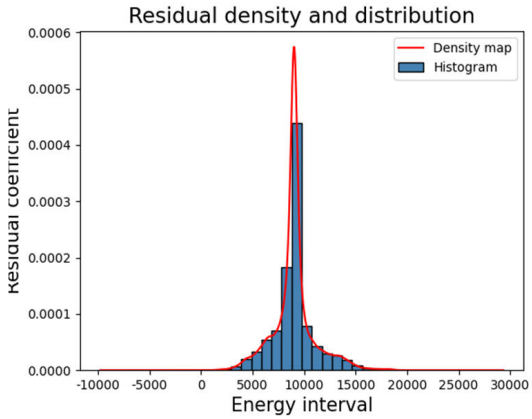


Figure 11: Residual error density and distribution

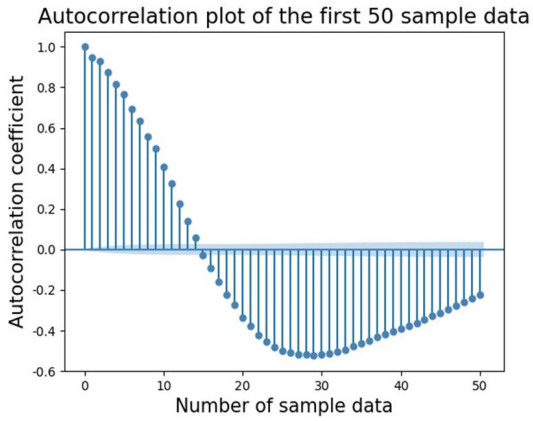


Figure 12: Autocorrelation plot of the first 50 sample data

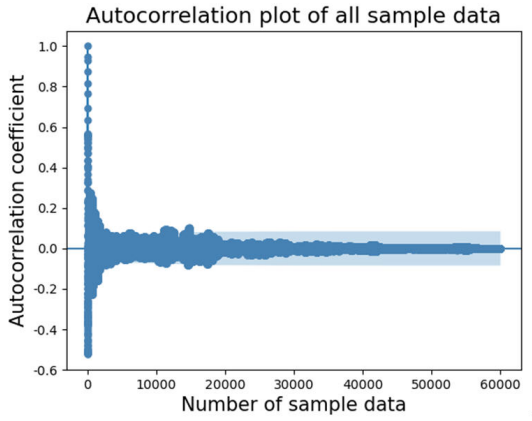


Figure 13: Autocorrelation plot of all sample data

scheme has good encryption and decryption efficiency.

5.3 Comparative Analysis with Existing Encryption Schemes

This section compares the experimental results with the improved probabilistic statistics addition homomorphic algorithm, El-Gamal, probabilistic homomorphic speech encryption algorithm in the existing scheme [6,15,16] and BFV speech homomorphic encryption scheme to objectively and accurately evaluates the proposed scheme. The experimental comparisons all adopt the speech data with a duration of 4s, and take the average value of each item for comprehensive evaluation as shown in Table 7.

It can be seen from Table 7 that compared with Ref. [6, 15, 16] and BFV homomorphic encryption sys-

Table 7: Performance comparison

Evaluation index	Proposed	Ref. [15]	Ref. [6]	Ref. [16]	Ref. [1]	Ref. [14]	BFV
Key size	2×256	-	-	196	-	$L \times 16$	128
key space	$2^{2\times256}$	-	-	4×2^{196}	-	$2^{L\times16}$	2^{128}
SNR (dB)	-44.8495	-29.96	-35.0224	-45.0206	-	-47.4953	-
SNRseg (dB)	-45.1173	-	-38.0201	-45.2222	-	-	-
Correlation coefficient original & encrypted	0.9931	0.9438	-	0.7386	0.9901	0.9981	-
Correlation coefficient original & decrypted	0.0031	0.0027	-	0.0115	0.0008	0.0032	-
Encryption time (s)	8.1106	25.3075	-	5.7208	2.3005	-	13.0878
Decryption time (s)	4.1652	24.2481	-	29.0230	2.8775	-	4.1264
Cipher expand	24.7519	6.6667×10^6	-	-	-	-	26.5397
Statistical attack	✓	✓	×	✓	✓	✓	✓
Entropy attacks	✓	×	×	×	×	×	×
Chosen-plaintext attack	✓	✓	×	×	✓	✓	✓

tem, the proposed scheme is generally superior than other speech homomorphic encryption algorithms. Compared with the Ref. [14], the adaptive speech homomorphic encryption scheme proposed in this paper has a larger key space than adaptive speech encryption model. This is mainly because the proposed scheme performs adaptive classification on the speech data at first, and then the classified data of different types are separately encrypted with different homomorphic encryption in parallel. While ensuring good encryption performance, the computational complexity and the ciphertext expansion are reduced. Compared with the Ref. [1], the encryption and decryption efficiency is lower than that of the speech chaotic encryption algorithm, but the proposed scheme has higher security, and can realize the subsequent ciphertext operation to support the ciphertext speech retrieval system.

6 Conclusions

In this paper, an adaptive speech homomorphic encryption scheme based on energy in cloud storage is proposed, which solves the risks of data privacy exposure of traditional speech encryption methods, the low efficiency of existing speech homomorphic encryption schemes, and the poor adaptability of speech encryption schemes. The proposed scheme combines adaptive technology and homomorphic encryption technology to perform energy-based adaptive data classification and batch encryption of speech data, reducing the amount of the data encrypted by FHE to improve the efficiency of speech homomorphic encryption. Using the ciphertext calculation function supported by homomorphic encryption to realize speech data calculation operations in the ciphertext domain can greatly improve the security and calculation efficiency of speech retrieval and speech recognition systems. The analysis of the encryption/decryption capabilities of different test speech signals through multiple performance indicators such as correlation coefficient, SNR, and SNRseg shows that the proposed scheme can provide encrypted speech signals with low residual intelligibility. By comparing the performance with the existing scheme, the proposed scheme effectively improves the efficiency of speech homomorphic encryption, and can effectively improve the adaptability of the encryption scheme under the premise of keeping the algorithm complexity unchanged. It has higher security and lower ciphertext expansion, and can resist statistical attack, entropy attack, and chosen-plaintext attack.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 61862041, 61363078). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

References

[1] O. M. Al-Hazaimeh, "A new speech encryption algorithm based on dual shuffling henon chaotic map," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 3, pp. 2203–2210, 2021.

[2] H. Chen, W. Dai, M. Kim, "Efficient multi-key homomorphic encryption with packed ciphertexts with application to oblivious neural network inference," in *In the 2019 ACM SIGSAC Conference*, pp. 395–412, London, UK, November 2019.

[3] X. Feng, Y. Zhou, "English audio language retrieval based on adaptive speech-adjusting algorithm," *Complexity*, vol. 2021, pp. 2762180(1)–2762180(12), 2021.

[4] M. S. Hwang, E. F. Cahyadi, S. F. Chiou, C. Y. Yang, "Reviews and analyses the privacy-protection system

for multi-server," *Journal of Physics: Conference Series*, vol. 1237, no. 2, p. 022091, IOP Publishing, 2019.

[5] I. Iliashenko, V. Zucca, "Faster homomorphic comparison operations for bgv and bfv," *Proceedings on Privacy Enhancing Technologies*, vol. 2021, no. 3, pp. 246–264, 2021.

[6] O. A. Imran, S. F. Yousif, I. S. Hameed, "Implementation of el-gamal algorithm for speech signals encryption and decryption," *Procedia Computer Science*, vol. 167, no. 3, pp. 1028–1037, 2020.

[7] H. Jahanshahi, A. Yousefpour, J. M. Munoz-Pacheco, "A new fractional-order hyperchaotic memristor oscillator: Dynamic analysis, robust adaptive synchronization, and its application to voice encryption," *Applied Mathematics and Computation*, vol. 383, no. 2, pp. 125310, 2020.

[8] H. Leng, H. Chen, "Adaptive hdg methods for the brinkman equations with application to optimal control," *Journal of Scientific Computing*, vol. 87, no. 46, pp. 2–34, 2021.

[9] Z. Y. Li, X. L. Gui, Y. J. Gu, X. S. Li, H. J. Dai, X. J. Zhang, "Survey on homomorphic encryption algorithm and its application in the privacy-preserving for cloud computing (in chinese)," *Ruan Jian Xue Bao/Journal of Software* (in Chinese), vol. 29, no. 7, pp. 1830–1851, 2018.

[10] L. Liu, Z. Cao, C. Mao, "A note on one outsourcing scheme for big data access control in cloud," *International Journal of Electronics and Information Engineering*, vol. 9, no. 1, pp. 29–35, 2018.

[11] S. Najam, M. U. Rehman, J. Ahmed, "Data encryption scheme based on adaptive system," in *Global Conference on Wireless and Optical Technologies (GCWOT)*, pp. 1–4, University of Malaga, Spain, October 2020.

[12] C. Orlandi., P. Scholl, S. Yakoubov, "The rise of paillier: Homomorphic secret sharing and public-key silent ot," in *Advances in Cryptology – EUROCRYPT 2021*, pp. 678–708, Zagreb, Croatia, October 2021.

[13] A. Siswanto, C. Y. Chang, S. M. Kuo, "Multirate audio-integrated feedback active noise control systems using decimated-band adaptive filters for reducing narrowband noises," *Sensors*, vol. 20, no. 22, pp. 6693–6705, 2020.

[14] H. I. Shahadi. "Covert communication model for speech signals based on an indirect and adaptive encryption technique," *Computers and Electrical Engineering*, vol. 68, no. 4, p. 425–436, 2018.

[15] C. Shi, H. Wang, Y. Hu, "A speech homomorphic encryption scheme with less data expansion in cloud computing," *KSII Transactions on Internet and Information Systems*, vol. 13, no. 5, pp. 2588–2609, 2019.

[16] C. Shi, H. Wang, Q. Qian, H. Wang, "Privacy protection of digital speech based on homomorphic encryption," in *Cloud Computing and Security (ICCCS)*, pp. 365–376, Nanjing, China, July 2016.

[17] Y. Tang, B. Zhu, X. Ma, "Decoding homomorphically encrypted flac audio without decryption," in *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 675–679, Brighton, UK, May 2019.

[18] L. Teng, H. Li, J. Liu, "An efficient and secure ciphertext retrieval scheme based on mixed homomorphic encryption and multi-attribute sorting method under cloud environment," *International Journal of Network Security*, vol. 20, no. 5, pp. 872–878, 2018.

[19] L. Teng, H. Li, S. Yin, "IM-Mobishare: An improved privacy preserving scheme based on asymmetric encryption and bloom filter for users location sharing in social network," *Journal of Computers (Taiwan)*, vol. 30, no. 3, pp. 59–71, 2019.

[20] P. Thaine, G. Penn, "Extracting mel-frequency and bark-frequency cepstral coefficients from encrypted signals," in *INTERSPEECH 2019*, pp. 3715–3719, Graz, Austria, September 2019.

[21] A. V. Tutueva, E. G. Nepomuceno, A. I. Karimov, "Adaptive chaotic maps and their application to pseudo-random numbers generation," *Chaos, Solitons & Fractals*, vol. 133, no. 3, p. 109615, 2020.

[22] D. Wang, X. Zhang, "Thchs-30: A free chinese speech corpus," *arXiv preprint arXiv:1512.01882*, 2015.

[23] Q. Wu, M. Wu, "Adaptive and blind audio watermarking algorithm based on chaotic encryption in hybrid domain," *Symmetry*, vol. 10, no. 7, p. 284, 2018.

[24] A. Ww, C. Dsb, "The improved adaboost algorithms for imbalanced data classification," *Information Sciences*, vol. 563, no. 7, pp. 358–374, 2021.

[25] S. Yin, J. Liu, L. Teng, "Improved elliptic curve cryptography with homomorphic encryption for medical image encryption," *International Journal of Network Security*, vol. 22, no. 3, pp. 419–424, 2020.

[26] X. Zhu, T. K. Han, M. Kim, "Privacy-preserving multimedia data analysis," *The Computer Journal*, vol. 64, no. 7, pp. 991-992, 2021.

Biography

Qiu-yu Zhang Researcher/Ph.D. supervisor, graduated from Gansu University of Technology in 1986, and then worked at school of computer and communication in Lanzhou University of Technology. He is vice dean of Gansu manufacturing information engineering research center, a CCF senior member, a member of IEEE and ACM. His research interests include network and information security, information hiding and steganalysis, multimedia communication technology.

Yu-jiao Ba is currently a master student of the School of Computer and Communication, Lanzhou University of Technology, China. She received the BS degrees in computer science and technology from Lanzhou University of Technology, Gansu, China, in 2019. Her research

interests include network and information security, audio signal processing and application, multimedia data security.