A Speech Homomorphic Encryption Scheme with Less Data Expansion in Cloud Computing

Canghong Shi[1], Hongxia Wang[2*], Yi Hu[3], Qing Qian[1] and Hong Zhao[4]
[1] School of Information Science and Technology, Southwest Jiaotong University
Chengdu, PR 610031 - China
[e-mail: canghongshi@163.com; qianqing_swjtu@163.com]
[2] College of Cybersecurity, Sichuan University
Chengdu, PR 610041 - China
[e-mail: hxwang@scu.edu.cn]
[3] Computer Science Department, Northern Kentucky University
Highland Heights, KY 41099 - USA
[e-mail: huy1@nku.edu]
[4] Department of Electrical and Electronic Engineering, Southern University of Science and Technology
Shenzhen, PR 518000 - China
[e-mail: zh1985444@gmail.com]
*Corresponding author: Hongxia Wang

Received January 15, 2018; revised September 8, 2018; accepted December 6, 2018; published May 31, 2019

Abstract

Speech homomorphic encryption has become one of the key components in secure speech storing in the public cloud computing. The major problem of speech homomorphic encryption is the huge data expansion of speech cipher-text. To address the issue, this paper presents a speech homomorphic encryption scheme with less data expansion, which is a probabilistic statistics and addition homomorphic cryptosystem. In the proposed scheme, the original digital speech with some random numbers selected is firstly grouped to form a series of speech matrix. Then, a proposed matrix encryption method is employed to encrypt that speech matrix. After that, mutual information in sample speech cipher-texts is reduced to limit the data expansion. Performance analysis and experimental results show that the proposed scheme is addition homomorphic, and it not only resists statistical analysis attacks but also eliminates some signal characteristics of original speech. In addition, comparing with Paillier homomorphic cryptosystem, the proposed scheme has less data expansion and lower computational complexity. Furthermore, the time consumption of the proposed scheme is almost the same on the smartphone and the PC. Thus, the proposed scheme is extremely suitable for secure speech storing in public cloud computing.

Keywords: Homomorphic encryption, digital speech signal, secure speech storing, less data expansion, cloud computing

A preliminary version of this paper appeared in ICCCS 2016, July 29-31, Nanjing, China. This version includes a extension and supporting implementation results on speech homomorphic encryption.

http://doi.org/10.3837/tiis.2019.05.020 **ISSN : 1976-7277**

1. Introduction

Lots of speech records containing personal identifying information, such as call centers, banking systems, telecommunication systems etc., are produced every day. However, local computers and mobile devices with limited computational power, memory and storage are not capable to manage these digital speeches. These problems may be solved by various cloud offers, for its unlimited dynamic resources for computation, storage and service provision [1]. Thus more and more digital speech files are stored and processed in the cloud. Nevertheless, once speech data is stored in the cloud, there is no physical access to data servers, and it is only limited digital control by data owners. This raises security concerns.

Meanwhile, public cloud computing and Hadoop system give malicious attackers additional venues for attacking data [2]. Protecting the confidentiality and availability of the data becomes more and more critical and challenging in the cloud [3], which is also the key problems in secure speech storing. Ren *et al*. [4] propose a method to protect the integrity of data in public cloud storage. But how to protect the confidentiality of data has not been solved in public cloud storage [5]. Fortunately, there mainly are two technologies to perform the speech security. One is speech forensics [6-7], another is speech encryption, which is one of the most powerful methods for speech content protection. For speech encryption, most of the methods used to encrypt digital speeches are permutation [8-13]. The disadvantage of those schemes is that these encryption methods may change the mathematical structure of original speeches, thus leading to further operation on encrypted speeches infeasible. In order to conquer these shortcomings, some methods have been proposed, such as in [14], the original audio signal is encrypted by XOR method, the secret data and feature values of the audio are embedded into the encrypted audio by replacing the least significant bit of each sample, enabling identification and classification of the audios via K-NN classifier. Yakubu *et al*. [15] design a secure audio reverberation over cloud based on Shamir's Secret Sharing, which make further processing on the encrypted audio possible. But neither nor they are security yet maintaining the mathematical structure of original audios. While using homomorphic encryption not only protects privacy but also maintains the mathematical structure of original data, thus permitting some operation on encrypted data. So it is extremely suitable for secure speech storing in the public cloud.

It is because the homomorphic encryption data can be processed by the third party, a few researches on image homomorphic encryption [16-21] and speech homomorphic encryption [22-23] are proposed. In [16-18], the original images are encrypted by Paillier homomorphic cryptosystem, and secret data is embedded in the encrypted images by using homomorphic and probabilistic properties of Paillier cryptosystem. They efficiently embed and extract secret data to perform reversible data hiding in encrypted domain. At the same time, signal processing in the encrypted domain (SPED) has attracted considerable attention in recent years [24]. The field of SPED was born as a solution to efficiently preserve the privacy on those signal processing scenarios dealing with sensitive data. In order to address these privacy-preserving problems in secure speech storing and SPED of digital speeches, homomorphic encryption like the Paillier cryptosystem [25] has been widely employed for encrypted signal processing primitives. Unfortunately, Paillier homomorphic encryption characteristics make the lengths of cipher-texts increase exponentially as plaintexts are encrypted. Solutions resorting to the Paillier cryptosystem present a very huge cipher data expansion, despite the techniques like packing and unpacking to mitigate this effect [26]. For

instance, when Paillier cryptosystem, with modulus *N* being 1024 bits, is used to encrypt 16-bit speech, the encrypted result is 128 times larger than the original speech. The cipher data expansion becomes a serious problem when dealing with multimedia data like digital speeches. In [22], a method on how a distributed speech enhancement algorithm can be computed in a privacy-preserving manner is illustrated. Speech enhancement algorithms operate on complex, non-integer numbers, while Paillier encryption algorithms work on integer numbers from a limited range. This means that data in a privacy-preserving distributed speech enhancement algorithm based on Paillier homomorphic encryption need to be scaled, quantized, and transformed into real integers. It is very complex for users to perform this work. In [23], Yakubu *et al*. proposed a homomorphic encryption scheme called Shamir's secret sharing encryption scheme which is computationally lighter and gives collective control to perform computation on encrypted signals in cloud. But it is public encryption and is also troublesome to compute keys. It is an expensive computational algorithm for encrypting speeches. Due to the above reasons, it is not desirable for processing digital speeches in public cloud computing. The emphasis is on the fact that certain applications do not require public key cryptographic primitives and burden of long and multiple keys in public key cryptosystems can be reduced by using symmetrical encryption, as pointed out in [27].

To solve the above problems, an efficient speech symmetrical homomorphic encryption scheme for secure speech storing is proposed in this paper. The proposed homomorphic encryption scheme is constructed mainly based on matrix homomorphic encryption [27-28], which is a fully homomorphic encryption scheme. The rest of this paper is organized as follows. The homomorphic encryption based on matrix method is described in Section 2. The details of the proposed speech homomorphic encryption scheme are given in Section 3. Performance analysis of the algorithm is discussed in Section 4. Data expansion and computation time complexity comparison with Paillier cryptosystem are analyzed in Section 5. Experimental results are illustrated in Section 6. The conclusions are drawn in Section 7.

2. Homomorphic Encryption Based on Matrix Method

In [29], Goldwasser *et al*. propose the concept of probabilistic secure cryptosystem. Cryptosystems with homomorphic and probabilistic properties are more secure than cryptosystems with only homomorphic property. One of the homomorphic and probabilistic cryptosystems is symmetrical matrix cryptosystem, which are proposed in [27-28]. The matrix homomorphic cryptosystem [28] is presented as follows:

2.1 The Key Generated Algorithm

1. *N* is a big number. *p* and *q* are prime numbers in *N*. *p* and *q* are mutually prime and their size at less 512 bits.

2. An invertible matrix $K \in M(Z_N)$ is selected. Meanwhile, its inverse matrix can be computed as $K^{-1} \in M(Z_N)$. $M(Z_N)$ is a set of all matrices chosen from Z_N, Z_N is integer field in *N*.

3. For encryption we use $C \leftarrow Enc(x, K)$, where plaintext $x \in Z_N$ is encrypted into a cipher-text $C \in M(Z_N)$.

2.2 The Encryption Algorithm

1. Choose *m* random values, which are defined as $\left\{r_j \in Z_N \left| r_j \neq x, j = 1, 2, \ldots m\right.\right\}$.

2. An *n*+*m* dimension matrix *M* is constructed such that each row has only one element equal to $\left\{x_i \left| i = 1, 2, \ldots, n\right.\right\}$, and other *m* elements equal to r_j.

3. Cipher-text is Eq. (1), where $Enc(M, K)$ is an *n*+*m* dimensional matrix, the diagonal elements $\left\{r_j \left| j = 1, 2, \ldots, m\right.\right\}$ are random numbers, $\left\{x_i \left| i = 1, 2, \ldots, n\right.\right\}$ are plaintext, and *M* is the constructed matrix.

$$C = Enc(M, K) = K \cdot \begin{bmatrix} x_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \ddots & 0 & 0 & \cdots & 0 \\ 0 & 0 & x_n & 0 & \cdots & 0 \\ 0 & 0 & \cdots & r_1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \cdots & r_m \end{bmatrix} \cdot K^{-1} \tag{1}$$

2.3 The Decryption Algorithm

Analogously, decryption inverts the similarity transformation and returns the first element of matrix obtained as $M \leftarrow Dec(C, K)$. This function returns plaintext $\left\{x_i \left| i = 1, 2, \ldots, n\right.\right\}$ corresponding to cipher-text *C* using only the invertible matrix *K* as follows:

$$M = Dec(C, K) = K^{-1} \cdot \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1(n+m)} \\ c_{21} & c_{22} & \cdots & c_{2(n+m)} \\ \vdots & \vdots & & \vdots \\ c_{(n+m)1} & \cdots & & c_{(n+m)(n+m)} \end{bmatrix} \cdot K \tag{2}$$

3. The Proposed Speech Homomorphic Encryption Scheme

The proposed scheme mainly includes three parts: *construction of speech matrix unit, digital speech encryption, and the digital speech decryption*. In the construction of speech matrix unit, a series of fundamental unit used for encryption is created. In the digital speech encryption, a homomorphic encryption scheme is proposed to encrypt the matrix units. In the digital speech decryption, the speech cipher-texts are decrypted by the symmetrical key of this scheme. We choose an original speech signal $M = \left\{m_i \left| i = 1, 2, \cdots, L\right.\right\}$ with *L* samples, and the sampling value is within the range of [-32768, 32767]. The process of proposed speech homomorphic encryption scheme is elaborated as following:

3.1 Construction of Speech Matrix Unit

Firstly, the original speech signal *M* is divided into *V* non-overlapping frames and denoted as $\left\{M(v) \left| v = 1, 2, \cdots, V\right.\right\}$. The length of each speech frame is *J*, that is $V = L/J$. Secondly, two

speech frames should be chosen at least, since the speech is a "short stationary signal". If speech signals are dealt with one frame or less, the processed results still retain some characteristics of original speech signals. So S $(2 \le S \le V)$ frames are chosen and transformed into a matrix unit $\{M_f \mid f = 1, 2, \cdots, \lfloor V/S \rfloor\}$ shown as Eq. (3). There are *n*-*k* rows and *n* columns in every constituted matrix M_f, where $(n-k) \cdot n = S \cdot J$ and *n*, *k* are randomly chosen. If the last few frames of the original speech do not have $J \cdot S$ samples, we can complement $J \cdot S$ samples with some random numbers.

$$M_f = \begin{bmatrix} m_1 & m_2 & \cdots & m_n \\ m_{n+1} & & \cdots & m_{2n} \\ \vdots & \vdots & & \vdots \\ m_{S \cdot J - n + 1} & & \cdots & m_{S \cdot J} \end{bmatrix}_{(n-k) \times n} \tag{3}$$

3.2 Digital Speech Encryption

Definition 1: For a matrix $A \in R^{h \times n}$, *R* is behalf of field of real number, if there is one matrix $B \in R^{n \times h}$ satisfies that:

$$B \cdot A = I_n \tag{4}$$

where the matrix *A* is left invertible, and *B* is the left inverse matrix of *A*, where I_n is identity matrix.

For large primes *p* and *q*, let $p \cdot q = N$. Two random matrices $P_1, P_2 \in M_{h \times n}(Z_N)$ are selected, their dimension are $h \times n$. Both P_1^{-1} and P_2^{-1} are the left inverse matrix of P_1 and P_2 respectively, and their dimensions are $n \times h$. The matrices P_1, P_2 are constructed as Eq. (5) and Eq. (6), p_{ij} and p'_{ij} are random numbers in Z_N. If row or column vectors of P_1, P_2 are linearly independent vectors (It makes sure that the matrices P_1, P_2 are invertible), then P_1, P_2 are used as the symmetrical key pair of the speech homomorphic encryption system. Otherwise, the matrices P_1, P_2 should be re-selected.

$$P_1 = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{h1} & p_{h2} & \cdots & p_{hn} \end{bmatrix}_{h \times n} \bmod N, \quad 0 \le p_{ij} \le N, \tag{5}$$

and

$$P_2 = \begin{bmatrix} p'_{11} & p'_{12} & \cdots & p'_{1n} \\ p'_{21} & p'_{22} & \cdots & p'_{2n} \\ \vdots & \vdots & & \vdots \\ p'_{h1} & p'_{h2} & \cdots & p'_{hn} \end{bmatrix}_{h \times n} \bmod N, \quad 0 \le p'_{ij} \le N \tag{6}$$

With chosen $k \cdot n$ random numbers, we place these random numbers at the last *k* rows of M_f to construct matrix unit as follows:

$$M'_f = \begin{bmatrix} m_1 & m_2 & \cdots & \cdots & m_n \\ m_{n+1} & & \cdots & \cdots & m_{2n} \\ \vdots & \vdots & & \cdots & \vdots \\ m_{S\cdot J-n+1} & & \cdots & \cdots & m_{S\cdot J} \\ r_{S\cdot J+1} & \cdots & & \cdots & r_{S\cdot J+n} \\ \vdots & \vdots & & \cdots & \vdots \\ r_{S\cdot J+(k-1)\cdot n+1} & & & \cdots & r_{S\cdot J+k\cdot n} \end{bmatrix}_{n\times n} \tag{7}$$

where M'_f is the constructed matrix unit, which is used for encryption in the next step.

The left inverse matrix of P_2 is calculated and denoted as P_2^{-1}, and the speech plaintext matrices M'_f are encrypted using the key P_1 and P_2^{-1}, obtaining the speech cipher-text matrices C_f as follows:

$$C_f = P_1 \cdot M'_f \cdot P_2^{-1} = \sum_{j=1}^{h}\left(\sum_{i=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{h} p_{ji}\cdot m_{ij}\right)\cdot p'_{ji} \bmod N = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1h} \\ c_{21} & c_{22} & \cdots & c_{2h} \\ \vdots & \vdots & & \vdots \\ c_{h1} & c_{h2} & \cdots & c_{hh} \end{bmatrix}_{h\times h} \tag{8}$$

In Eq. (8), the key matrices P_1, P_2 and plain matrix M'_f may or may not be square matrix. Here $h \neq n$, n is fixation, the encrypted speech matrices C_f is $h\times h$ dimension, the speech cipher-texts are more random than the situation when $h = n$. It is more difficult to guess the size of the plaintext matrix M'_f from the encryption results. Natural speech-frames generally have some samples close to zero. In other words, M'_f is redundant. When $h < n$ and plaintext matrix M'_f is sparsity, the redundancy is reduced in the speech cipher-text frames, but it cannot be decrypted when plaintext matrix M'_f is not sparsity. When $h > n$, the redundancy is increased in the speech cipher-text frames. For large C_f, there is no efficient algorithm to factorize a large matrix into its matrix factors. Thus this speech encryption scheme is a secure cryptosystem. In order to obtain better security, we can add more random numbers in the matrix unit M'_f to render the encryption results more random, or we can disrupt the order of elements in plain speech matrix unit M'_f.

3.3 Homomorphism

This proposed speech homomorphic encryption scheme allows users to perform operations scaling and addition on the plain data by manipulating the encrypted data. Let $E[\cdot]$ and $D[\cdot]$ denote the encryption and decryption operations in this scheme, respectively. This scheme is addition homomorphic according to Eq. (9) as follows:

$$C_1 + C_2 = P_1 \cdot M_1 \cdot P_2^{-1} + P_1 \cdot M_2 \cdot P_2^{-1} = P_1 \cdot (M_1 + M_2) \cdot P_2^{-1} \tag{9}$$

where M_1, M_2 are plain speech frames, and C_1, C_2 are encrypted speech frames.

If we consider two plaintexts m_i and m'_i, the addition homomorphic property of proposed cryptosystem ensures that

$$D\left[\left(E\left(m_i, r_j\right)+E\left(m_i', r_j'\right)\right) \bmod N\right]=\left(m_i+m_i'\right) \bmod N, \quad \left(1 \le i \le (n-k)\times n, 1 \le j \le k \times n\right), \tag{10}$$

and

$$D\left[k \cdot E\left(m_i, r_j\right) \bmod N\right]=k \cdot m_i \bmod N, \quad \left(1 \le i \le (n-k)\times n, 1 \le j \le k \times n\right) \tag{11}$$

where $N = p \cdot q$ is a product of two large prime numbers p and q, $(n-k)\times n$ is the dimension of matrix M_f, and $k \times n$ is the numbers of random numbers.

3.4 Lossless Compression of Encrypted Speech

In the compression procedure, all of the encrypted speech samples are converted to two parts of smaller cipher-texts by removing the mutual information in each encrypted sample to reduce data size. In the compression process, every encrypted speech sample is divided into two parts: the first part made up of the cipher-text $S=\left\{s_{ij} \mid i=1,2,\ldots,h;\ j=1,2,\ldots h\right\}$ as Eq. (12), the second part is the cipher-text $C''=\left\{c_{ij}'' \mid i=1,2,\ldots,h;\ j=1,2,\ldots h\right\}$ as Eq. (13). The information in the second part will be reserved and the information redundancy in the first part will be reduced. The details are shown as follows:

1) The first part speech cipher-text S and the second part speech cipher-text C'' of an encrypted speech C can be acquired by calculating s_{ij} and c_{ij}''. Namely,

$$s_{ij}=\left\lfloor \frac{c_{ij}}{2^{15}} \right\rfloor, \tag{12}$$

and

$$c_{ij}''=c_{ij}-2^{15} \cdot s_{ij} \tag{13}$$

Here, the value 2^{15} is information redundancy in encrypted speech, c_{ij} is the encrypted speech samples.

2) All the data generated from s_{ij} and c_{ij}'' are collected in the first part and the second part, which are the encrypted speech and stored in the cloud. Reducing the information redundancy 2^{15} in Eq. (13) is aim at producing the lossless compressed data of encrypted speech.

The proposed scheme is linear to its encrypted process. The cipher-texts and plaintexts can be formation linear functional relationship. The compression results also have the same nature. The linearity process of compression does not influence homomorphism of the proposed scheme.

3.5 Speech Reconstruction

With the compressed data and the value 2^{15}, a receiver can perform the following process to reconstruct the original encrypted speech. It can satisfy different requirements by selecting one part compressed data S or two parts compressed data S and C''.

1) Decompose the received compressed data, i.e., S and C'', and obtain the values of all samples s_{ij} and c_{ij}'', with the value 2^{15}, the receiver calculates c_{ij} by using each sample of s_{ij} and c_{ij}'' at the same positions such as

$$c_{ij}=c_{ij}''+2^{15} \cdot s_{ij} \tag{14}$$

All the data of samples c_{ij} are collected as original encrypted speech.

2) The compressed data S provides more detailed information about the original encrypted speech to produce a final reconstructed result. When required quality of decrypted speech is

not too high and storage resources are limited, only *S* is needed. When it is required that speech in special applications should be decrypted completely, then the two parts compressed, i.e., data *S* and C'' need to be reconstructed lossless from original encrypted speech. The original speech cipher-text *C* will be decrypted by the process of homomorphic decryption illustrated in the next section. Since no information is lost in the reconstruction process, it is a lossless reconstruction.

3.6 Digital Speech Decryption

The decryption process of the speech cipher-texts is similar to the encryption and it involves running the algorithm in reverse. The process of decryption speech scheme is as follows:

With the symmetrical key pair P_1, P_2 of encryption system, the left inverse matrix P_1^{-1} of P_1 is calculated, and the cipher-texts $C = \{c_{ij} \mid i = 1, 2, \ldots, h;\ j = 1, 2, \ldots h\}$ can be decrypted according to Eq. (15).

$$M' = P_1^{-1} \cdot C \cdot P_2 = \sum_{i=1}^{n} \left(\sum_{j=1}^{h} \sum_{j=1}^{h} \sum_{i=1}^{n} \dot{p}_{ij} \cdot c'_{ji} \right) \cdot p''_{ij} \bmod N = \begin{bmatrix} m'_{11} & m'_{12} & \cdots & m'_{1n} \\ m'_{21} & m'_{22} & \cdots & m'_{2n} \\ \vdots & \vdots & & \vdots \\ m'_{n1} & m'_{n2} & \cdots & m'_{nn} \end{bmatrix}_{n \times n} \tag{15}$$

where M' is defined as the decrypted result containing random numbers, $\dot{p}_{ij}$ and p''_{ij} are the elements of P_1^{-1} and P_2, respectively.

3.7 Restore Decrypted Speech

The matrix units $M(v)$ of original speech $M = \{M(v) \mid v = 1, 2, \ldots, V\}$ can be reconstructed by discarding the last *k* rows random numbers of decrypted matrix units M', and the reconstructed result is shown as follows:

$$M_f = \begin{bmatrix} m_1 & m_2 & \cdots & m_n \\ m_{n+1} & & \cdots & m_{2n} \\ \vdots & \vdots & & \vdots \\ m_{S \cdot J - n + 1} & & \cdots & m_{S \cdot J} \end{bmatrix}_{(n-k) \times n} \tag{16}$$

From Eq. (16), the speech matrices M_f can be restored into *S* frames of speech signal, thus the original speech signal samples are obtained. Then, we collect all of the decrypted frames to restore original speech signal.

4. Performance Analysis

4.1 Security Analysis

The proposed speech homomorphic encryption scheme is secure under two conditions: the plain speech is regarded as “random data” for attackers, and the assumption that the plain speech is consisted of random large-numbers in *N*, which is a product of two large prime numbers *p* and *q*. In this scheme, the random numbers r_j are used for randomizing the plain speech samples to make the plain speech as “random data”. The proposed speech homomorphic encryption scheme with $C_f = P_1 \cdot M'_f \cdot P_2^{-1}$ linearly encode a plain speech

M'_f into a cipher speech C_f. The security of the speech homomorphic encryption scheme is based on the difficulty of dividing big prime number.

The threat of this scheme is known as plaintext attack. This attack is based on the availability of some plaintext-cipher-text pairs. Attackers can extract useful information in the encryption process, which can be used to decrypt speech cipher-texts. However, there are random numbers in the encrypted scheme, so the encrypted results are different when one speech matrix unit of plain speech is encrypted twice. Thus, the encryption algorithm is semantically secure and the plaintext attack cannot take effect.

In the proposed scheme, when users send an encrypted speech signal to the cloud. The users choose a key matrix pair P_1, P_2 and encrypt the speech signal via matrix multiplication $C = P_1 \cdot M \cdot P_2^{-1}$. Only the encrypted *C* is transmitted to the cloud. It is very difficult to find the key matrix and the plain speech, if the attackers do not know the key which was used to encrypt the speech signal. In practice, the size *n* of speech signal is very large; it is difficult to determine the plain speech without knowledge of the secret key.

The speech encryption scheme is secure under the assumption that the key matrix pair P_1, P_2 is uniformly random and independent for arbitrary dimension. The attackers cannot learn the key P_1, P_2. Suppose an attacker learns some key pairs of original speeches and their corresponding encrypted speeches in a Known Plaintext Attack model, the information available to the attacker under the Known Plaintext Attack model is that the attacker can derive the original speech from encrypted speech, which is impossible for encrypting speeches once. This speech encryption scheme can encrypt speeches once safely, but it is not secure when the same key is used more than twice.

4.2 Characteristics of the Encrypted Speech

Short-time average energy and short-term average zero-crossing rate are speech characteristics in time domain. Short-time average energy can describe the energy variation of the segment of unvoiced or voiced speech with time-varying. The length of two speech frames of our proposed scheme is $2 \cdot J$, where *J* is the length of one frame from part 3.1 in Section 3. The short-time average energy for one frame of original speech and two frames can be defined as follows:

$$E_{k_1} = \sum_{k_1=1}^{J} m_{k_1}^2 , \tag{17}$$

and

$$E_{k_1} + E_{k_2} = \sum_{k_1=1}^{J} m_{k_1}^2 + \sum_{k_2=1}^{J} m_{k_2}^2 \tag{18}$$

Short time average energy of encrypted speech of two frames is given by Eq. (19), where $n \times n, h \times h > 2 \cdot J$. The energy derivation of this statement is also given by Eq. (19), from which it can be seen that the short-time average energy of cipher-texts is much larger than the plaintexts. Comparing Eq. (18) with Eq. (19), the short-time average energy of two speech frames of encrypted speech have been hugely changed, which means that the characteristics of speech cipher-texts is greatly different from the plaintexts'.

$$
\begin{aligned}
E'_n &= \sum_{k=1}^{2\cdot J}\left[\sum_{i=1}^{2\cdot J}\sum_{j=1}^{2\cdot J} c_{ij}\right]_k^2 = \sum_{k=1}^{2\cdot J}\left[\sum_{j=1}^{2\cdot J}\sum_{i=1}^{2\cdot J}\sum_{i=1}^{2\cdot J}\sum_{j=1}^{2\cdot J} c_{ij}\cdot c_{ij}\right]_k \\
&= \sum_{k=1}^{2\cdot J}\left[\left(\sum_{j=1}^{2\cdot J}\left(\sum_{i=1}^{2\cdot J}\sum_{i=1}^{2\cdot J}\sum_{j=1}^{2\cdot J} p_{ij}\cdot m_{ji}\right)\cdot p'_{ji}\right)\times\left(\sum_{j=1}^{2\cdot J}\left(\sum_{i=1}^{2\cdot J}\sum_{i=1}^{2\cdot J}\sum_{j=1}^{2\cdot J} p_{ij}\cdot m_{ji}\right)\cdot p'_{ji}\right)\right]_k \\
&>> \sum_{k=1}^{2\cdot J}\left[\sum_{j=1}^{2\cdot J}\sum_{i=1}^{2\cdot J} m_{ij}^4\right]_k \\
&> \sum_{k_1=1}^{J}\left[\sum_{j=1}^{J}\sum_{i=1}^{J} m_{ij}^2\right]_{k_1} + \sum_{k_2=1}^{J}\left[\sum_{j=1}^{J}\sum_{i=1}^{J} m_{ij}^2\right]_{k_2} \\
&= \sum_{k_1=1}^{J} m_{k_1}^2 + \sum_{k_2=1}^{J} m_{k_2}^2
\end{aligned}
\tag{19}
$$

where $[\cdot]_k$ is on behalf of "matrix", k is the location of elements in the matrix, the numbers of k are equal to the total numbers of k_1 and k_2, p_{ij} and p'_{ji} are elements in Eq. (5) and Eq. (6), respectively.

The frequency information of digital speech can be revealed by short-term average zero-crossing rate to some extent. Short-term average zero-crossing rate can be decided by the numbers of plus-minus sample values. In this scheme, the key matrix must contain minus samples. When encrypting happens, the plus-minus of samples $c_w(i)$ for the cipher-texts is different from the plaintexts' $m_w(i)$, so the short-term average zero-crossing rate for two frames is different. Short-term average zero-crossing rate of two frames of plain speech and cipher speech are shown as follows:

$$
Z_s = \sum_{i=s}^{s+N-1}\left|\operatorname{sgn}\left[m_w(i)\right]-\operatorname{sgn}\left[m_w(i-1)\right]\right| + \sum_{j=s+N}^{s+2\cdot N-1}\left|\operatorname{sgn}\left[m_w(j)\right]-\operatorname{sgn}\left[m_w(j-1)\right]\right| \tag{20}
$$

where $m_w(i)$ are samples of one speech frame, $m_w(j)$ are samples of the other speech frame.

$$
Z'_s = \sum_{i=s}^{s+k\cdot n+2\cdot N-1}\left|\operatorname{sgn}\left[c_w(i)\right]-\operatorname{sgn}\left[c_w(i-1)\right]\right| \tag{21}
$$

where $c_w(i)$ are samples of cipher-texts for two speech frames.

Also, $\operatorname{sgn}[\cdot]$ is symbolic function and defined as follows:

$$
\operatorname{sgn}\left[x(n)\right] = \begin{cases} 1, & x(n)\geq 0; \\ -1, & x(n)<0 \end{cases} \tag{22}
$$

From the above analysis, we can see that the short-time average energy and short-term average zero-crossing rate of the encrypted speech are greatly different from those of the original speech. These features of original speech are changed or missing after encrypting, so the attackers cannot obtained these features from the encrypted speech, and these features of encrypted are useless for the attackers.

5. Data Expansion Comparison and Computational Complexity

5.1 Data Expansion Analysis

The proposed speech homomorphic encryption method and Paillier cryptosystem are both probabilistic and addition homomorphic cryptosystem. However, due to working with large integer numbers in the encrypted domain, operations like exponentiation of Paillier cryptosystem become significantly expensive in terms of run-time compared to operations in the non-encrypted domain. Compared with Paillier cryptosystem, operations such as addition and multiplication of the proposed method are cheaper in terms of computation. It has high efficiency, low-complexity and less data expansion. It also can encrypt more samples one time than Paillier cryptosystem due to its matrix function structure.

Be realistic, fifteen samples of original speeches are taken and encrypted for example. The cipher-texts samples are produced by proposed method multiplied by one extend factor $Q = 10^{35}$ (the factor Q is aimed at making the encrypted cipher-texts approach to Paillier's), and the result is defined as CQ, the speech cipher-texts produced by Paillier cryptosystem are defined as CP. In order to make the speech cipher-texts smaller, the public key g of Paillier can be 2. The comparison of data expansion between CQ and CP is shown in **Fig. 1**, from which we can know that the Paillier cryptosystem has much more data expansion than the proposed method, and the data expansion of Paillier cryptosystem is much bigger than that proposed with the samples enlargement.

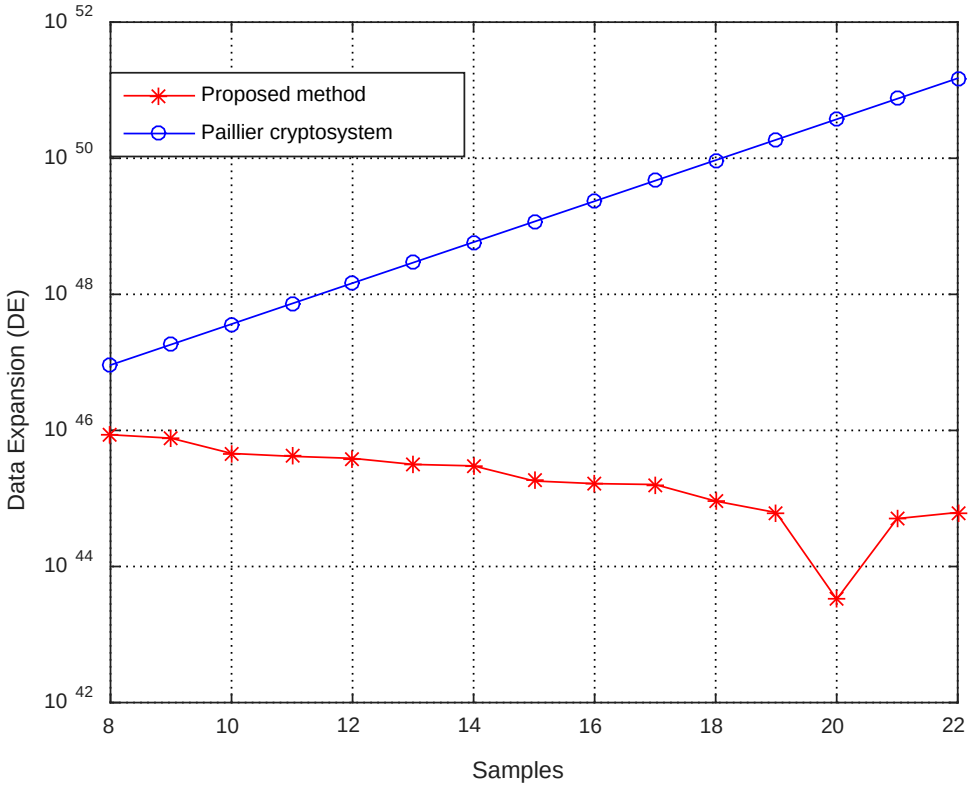


Fig. 1. Data expansion comparison of cipher-text between Paillier cryptosystem and the proposed method

5.2 Computational Complexity Analysis

Generally speaking, the computational complexity of an algorithm should be expressed as $T(n)=\Theta(f(n))$, where $T(n)$ is time frequency, $f(n)$ is the algorithm function of *n*, and Θ is asymptotic notation used for describing asymptotic behavior of function $f(n)$.

According to the proposed speech homomorphic encryption scheme, its encryption function $C=P_1\cdot M\cdot P_2^{-1}$ has two times $n\times n$ dimension matrices multiplication, which is two triple loops, each loop is from 1 to *n*, all of calculation times are $2\times n\times n\times n=2\cdot n^3$. The decryption algorithm function $M=P_1^{-1}\cdot C\cdot P_2$ has the same function structure as the encryption mapping. Their computational complexities are all $\Theta(n^3)$, which is the cubic order computation time complexity. And computational complexities of both encryption and decryption function are

$$\begin{aligned} T(n)&=\sum_{j=1}^{h}\left(\sum_{i=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{h}p_{ij}\cdot m_{ji}\right)\cdot p'_{ij}=2\cdot\sum_{i=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{h}\left(p_{ij}\cdot m_{ji}\right)\\ &=2\cdot\sum_{i=1}^{n}\sum_{j=1}^{h}\Theta\left(p_{ij}\cdot m_{ji}\right)=\Theta\left(n^3\right)\end{aligned} \tag{23}$$

However, the encryption speech algorithm based on Paillier cryptosystem [16-18] is $C=g^{m_i}\cdot r^N \bmod N^2$, where m_i denotes the speech samples. Its encryption speech computational complexity is more than $\Theta(g^n)$ (*g* is the public key in Paillier cryptosystem), which is linear exponent computation time complexity. The encryption computational complexity of Paillier cryptosystem is

$$T(n)=\sum_{i=1}^{n}\Theta\left(g^{m_i}\cdot r^N\right)\bmod N^2=r^N\cdot\sum_{i=1}^{n}\Theta\left(g^{m_i}\right)\bmod N^2>\Theta\left(g^n\right) \tag{24}$$

Paillier decryption algorithm is $m_i=\left(\left(c_i^{\lambda}\bmod N^2-1\right)/\left(g^{\lambda}\bmod N^2-1\right)\right)\bmod N$, and its computational complexity is more than $\Theta(n^{\lambda})$ (λ is the private key in Paillier cryptosystem, which is much bigger than 3), which is linear λ*-th* power order computation time complexity. The decryption algorithm computational complexity can be calculated as

$$\begin{aligned} T(n)&=\sum_{i=1}^{n}\Theta\left(\frac{c_i^{\lambda}\bmod N^2-1}{g^{\lambda}\bmod N^2-1}\right)\bmod N\\ &=\left(\frac{1}{g^{\lambda}\bmod N^2-1}\right)\cdot\sum_{i=1}^{n}\Theta\left(c_i^{\lambda}\bmod N^2-1\right)\bmod N>\Theta\left(n^{\lambda}\right)\end{aligned} \tag{25}$$

Through the above analysis, it is known that the proposed scheme has much lower computational complexity than Paillier cryptosystem, which means the former outperforms the latter asymptotically. In practice, when *n* is bigger, the Paillier cryptosystem will has much higher computational complexity than the proposed scheme.

To evaluate the time consumption of encryption and decryption of the proposed scheme, two types (Chinese for type 1 and English for type 2) of speech signals are selected from the well-known SQAM files [30] as test speech signals, which are shown in **Table 1**. Time consumptions are measured not only on a PC with 3.40 GHz CPU and 4.00 GB main memory but also on a smartphone with 1.70 GHz CPU and 3.00 GB main memory (The brands of the

smartphone is “honor”, which model number is H60-L01).

Time consumptions of encrypting a 5-second Chinese speech file and a 4.5-second English speech file by the proposed scheme on the PC are 30.2466 seconds and 25.3075 seconds respectively. Decrypting of the Chinese speech and the English speech cipher-texts on the PC takes 26.7990 seconds and 24.2481 seconds respectively. Time consumptions of the proposed encryption scheme are compared with the Paillier cryptosystem’s. The trend of time consumptions of encryption is shown in **Fig. 2**, from which we can see that the proposed encryption scheme has less time consumption than Paillier cryptosystem, and the time consumptions on the smartphone are almost the same as that on the PC. The time consumptions of proposed decryption scheme are compared with the Paillier cryptosystem’s. The trend of time consumptions of decryption is shown in **Fig. 3**, from which we can see that the proposed decryption scheme has less time consumption than Paillier cryptosystem, and the time consumptions on the smartphone are almost the same as that on the PC. Therefore, it is practically feasible to achieve privacy protection of digital speeches for storing in the public cloud.

Table 1. Parameters of speech type 1/speech type 2 used for the time comparison simulations

Class	Number of samples	Length of speeches (s)
Chi_1/Eng_1	32049/31984	2
Chi_2/Eng_2	47834/48073	3
Chi_3/Eng_3	64016/63951	4
Chi_4/Eng_4	80082/80165	5
Chi_5/Eng_5	95921/96166	6
Chi_6/Eng_6	111990/112095	7
Chi_7/Eng_7	128312/127900	8
Chi_8/Eng_8	143818/144254	9
Chi_9/Eng_9	160092/160866	10
Chi_10/Eng_10	175848/175946	11
Chi_11/Eng_11	191890/191779	12
Chi_12/Eng_12	208017/208173	13
Chi_13/Eng_13	224007/224164	14

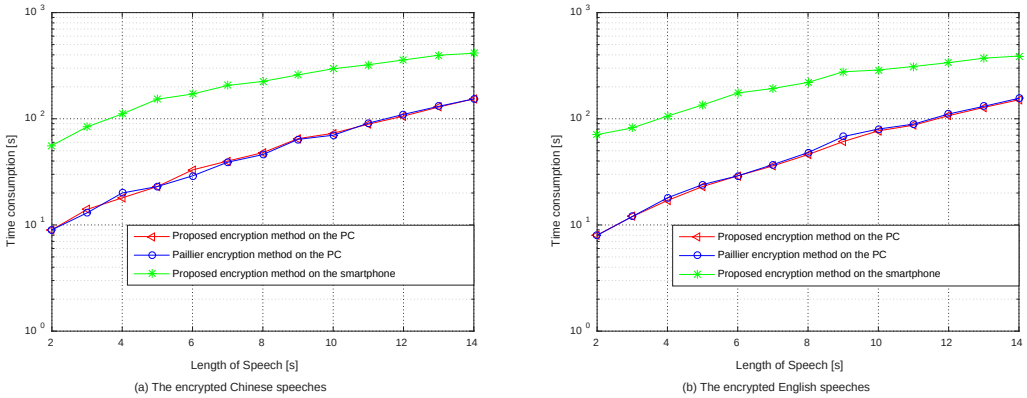


Fig. 2. The time consumption of encryption of Paillier cryptosystem and the proposed method. (a) The encrypted Chinese speeches, (b) The encrypted English speeches

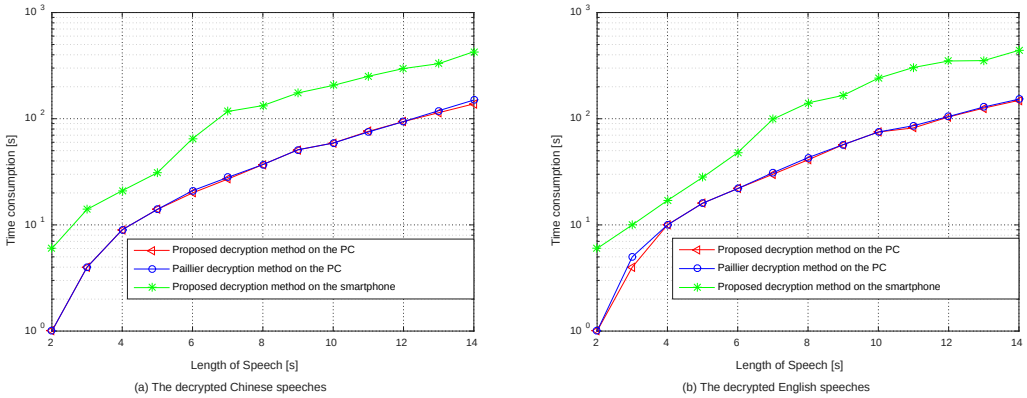


Fig. 3. The time consumption of decryption of Paillier cryptosystem and the proposed method. (a) The decrypted Chinese speeches, (b) The decrypted English speeches

In the Section 5.1 and Section 5.2, the algorithm analysis shows that the proposed scheme outperforms the speech homomorphic cryptosystem based on Paillier cryptosystem [25] in terms of data expansion and computational complexity. Meanwhile, it is because of the low computational complexity and data expansion that the proposed method can be ran on the smartphone, thus it is precisely satisfy the application trend of public cloud computing.

6. Experimental Results

To evaluate the proposed speech homomorphic encryption scheme, four different type speeches (speech type 1 and 3 are Chinese speeches, speech type 2 and 4 are English speeches) are selected for testing. The number of sampling points of Chinese speeches is 39760 and 41874, and that of English speeches is 102400 and 440832. In all the experiments, the speech signals are adapted to 16-bit quantified monaural speech files. They were sampled at 44.1 kHz in the WAVE format. Each frame has 160 samples, and two frames are taken in a speech matrix unit. The sizes of n and h are set as 20 and 18, respectively. All experiments are performed on a PC with 3.40 GHZ CPU and 4.00 GB main memory. Meanwhile, some related experiments are conducted on a smartphone with 1.70 GHZ CPU and 3.00 GB main memory (The brands of the smartphone is “honor”, which model number is H60-L01).

6.1 Speech encryption and decryption

We demonstrate the performance of our proposed speech homomorphic encryption cryptosystem in **Fig. 4** with speech type 1 and speech type 2 on the PC and in **Fig. 5** with speech type 3 and speech type 4 on the smartphone, respectively. From **Fig. 4** and **Fig. 5**, we can see that the waveforms of the encrypted speeches are uniform appearance, and the waveforms of decrypted speeches are visually no differences with the original speech signal. They prove that the proposed speech homomorphic encryption scheme has a good encryption and decryption performance. Therefore, it is practically feasible to achieve privacy protection of digital speeches in the public cloud not only on the PC but also on the smartphone.

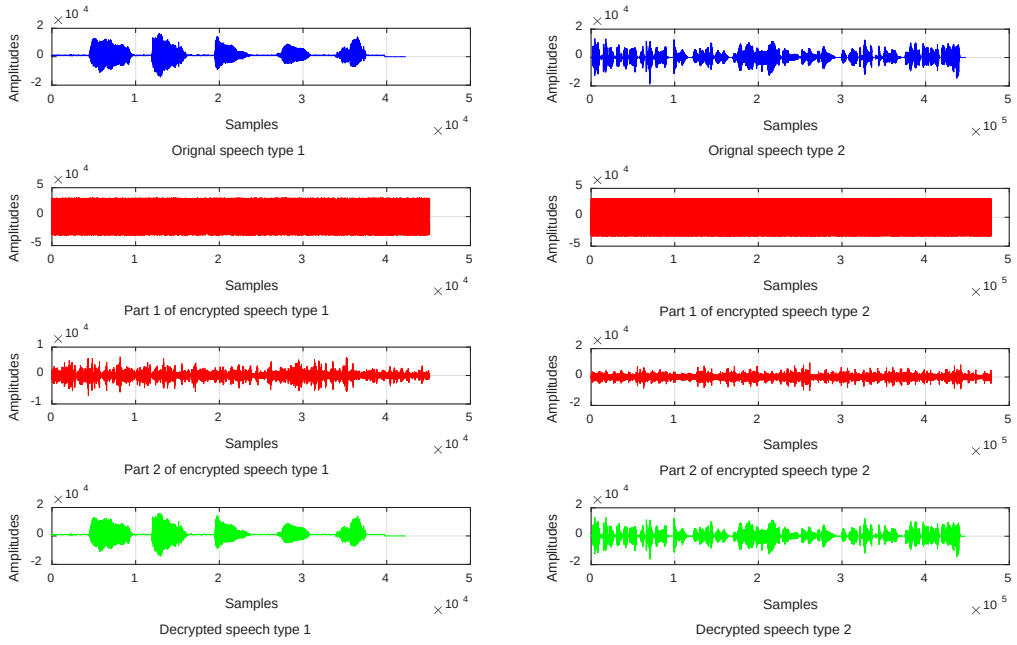


Fig. 4. Homomorphic encryption and decryption of speech type 1 and speech type2 on the PC

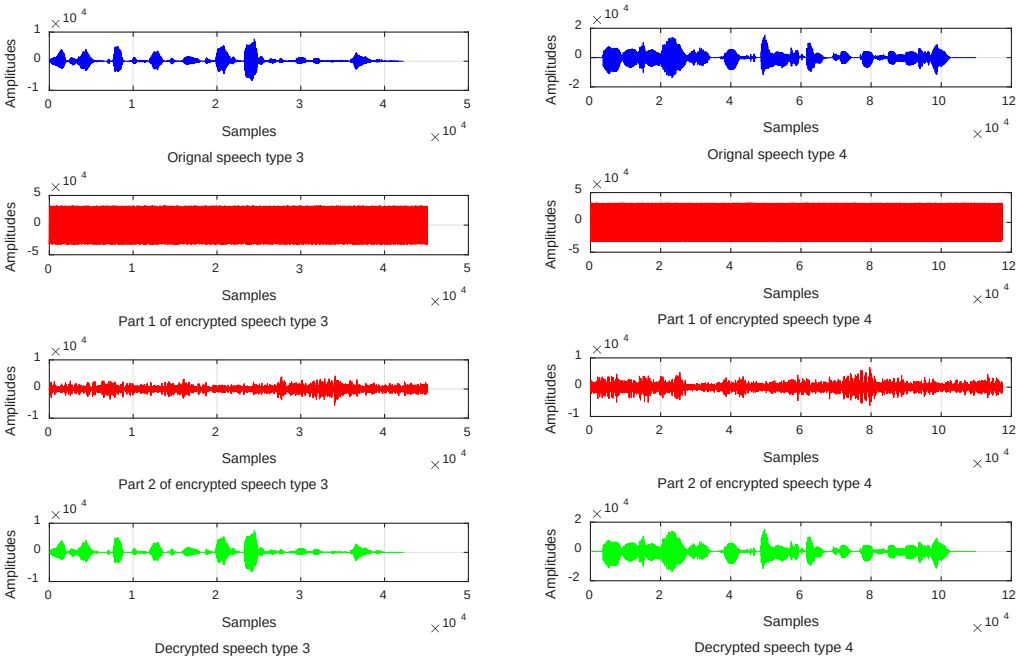


Fig. 5. Homomorphic encryption and decryption of speech type 3 and speech type 4 on the smartphone

6.2 Residual intelligibility

Speech spectrogram of cipher speech plays an important role in the evaluation of speech encryption. The spectrograms of the original and encrypted for speech type 1 and speech type 2 are shown in **Fig. 6**, and the spectrograms of the original and encrypted for speech type 3 and speech type 4 are shown in **Fig. 7**.

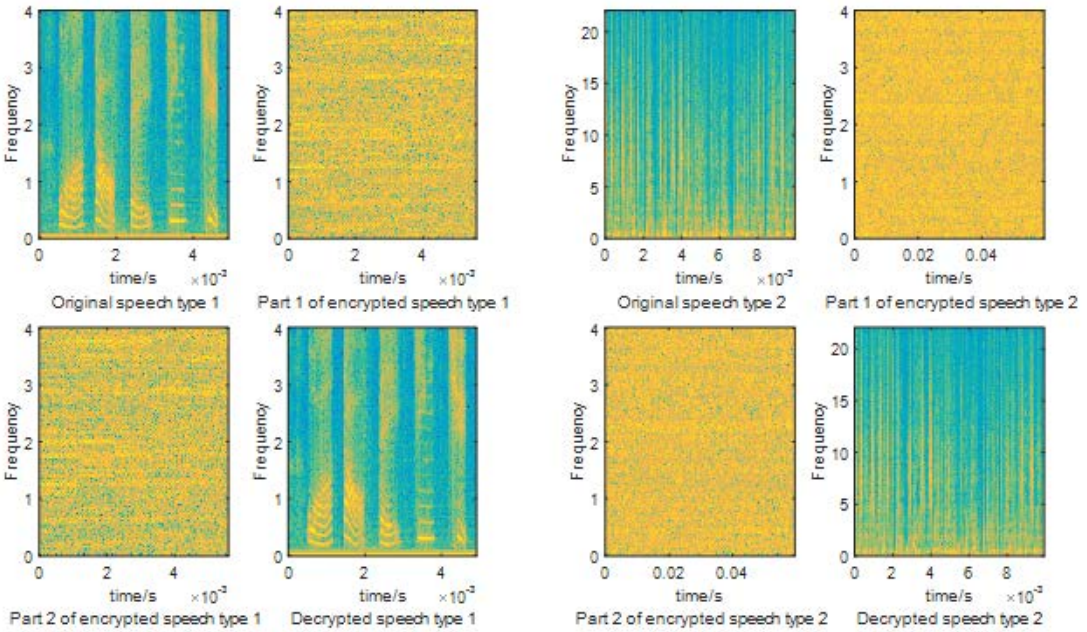


Fig. 6. Spectrograms of the original speeches and the encrypted speeches for speech type 1 and speech type 2 on the PC

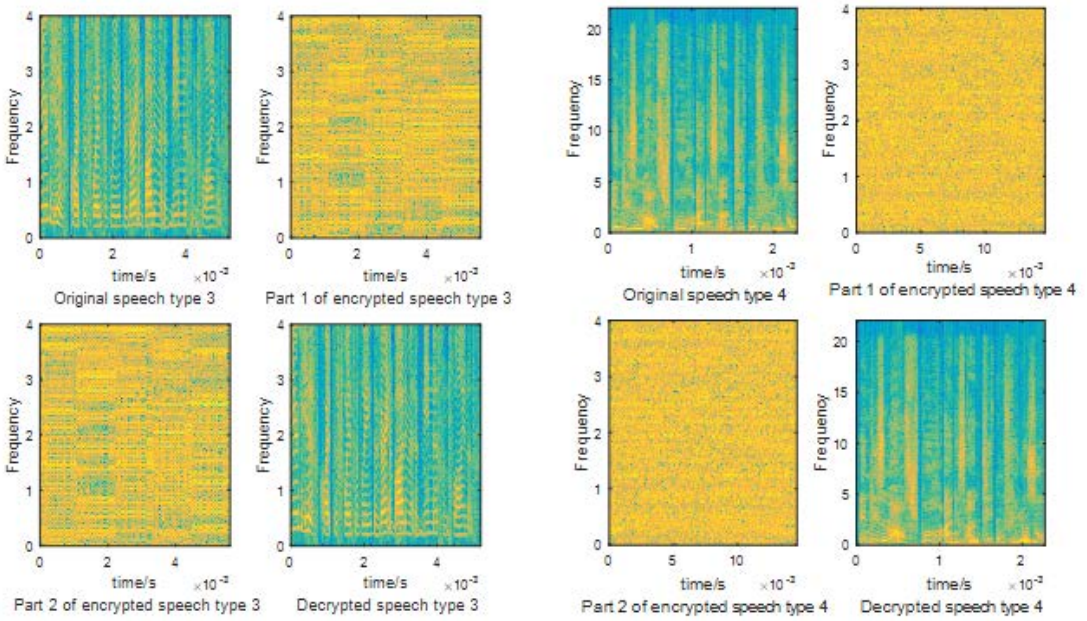


Fig. 7. Spectrograms of the original speeches and the encrypted speeches for speech type 3 and speech type 4 on the smartphone

It can be seen that the intonation of original speeches is removed, and the spectrograms of encrypted speeches are very similar to white noise. The significative speech spectrograms are completely changed after encrypting by the proposed scheme, which indicates that no residual intelligibility of encrypted speeches can be used. Therefore, it is practically feasible to achieve privacy protection of digital speeches in the public cloud not only on the PC but also on the smartphone.

6.3 Statistical analysis

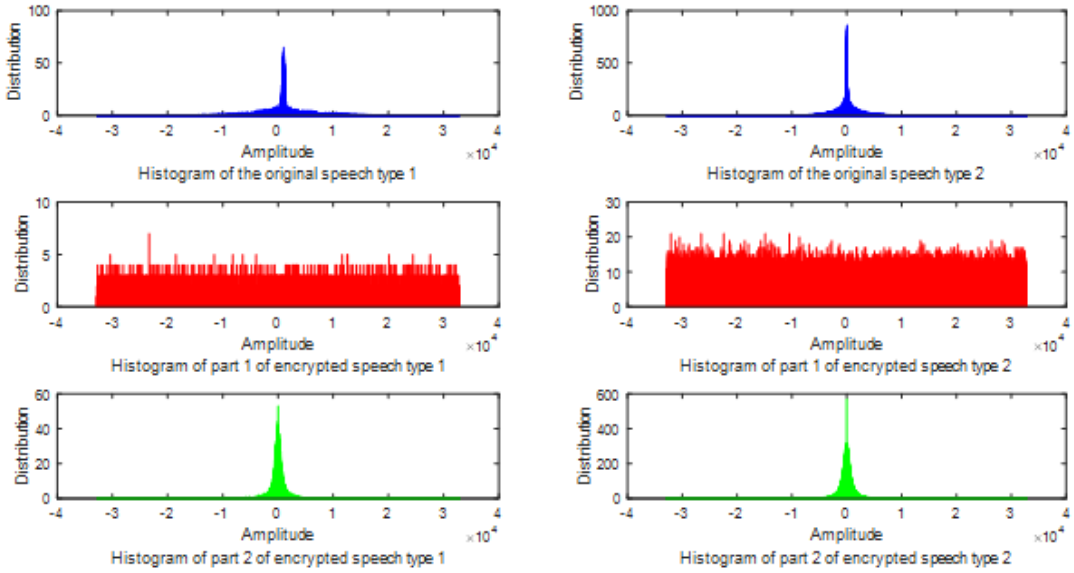


Fig. 8. Histograms of the original speeches and the encrypted speeches for speech type 1 and speech type 2 on the PC

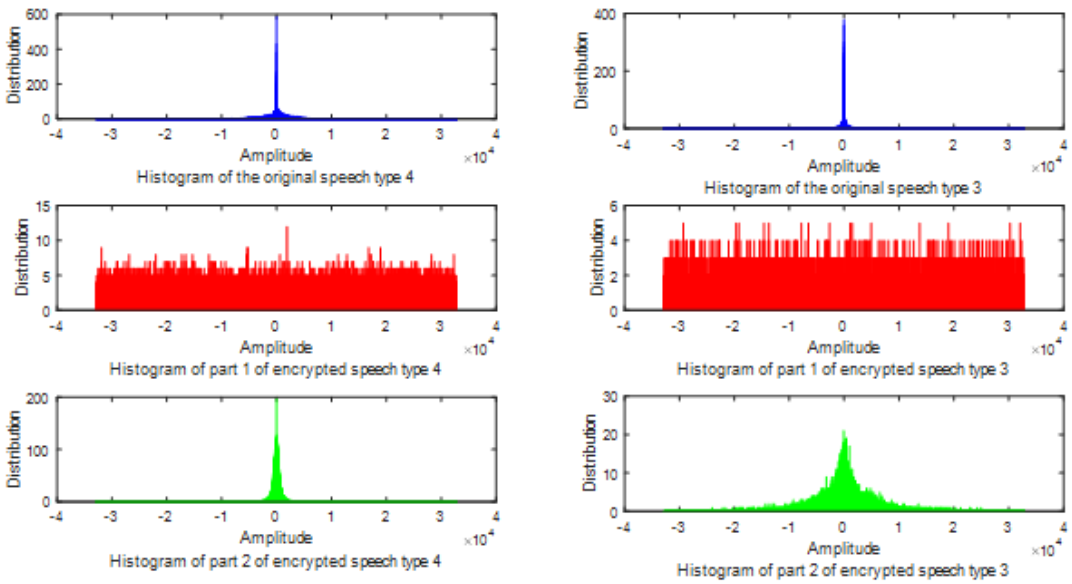


Fig. 9. Histograms of the original speeches and the encrypted speeches for speech type 3 and speech type 4 on the smartphone

Attackers can attack the encrypted speeches through revealing the histograms, which do not change or change little. In addition to spectrogram analysis, histogram analysis is also used to evaluate the performance of speech encryption here. **Fig. 8** shows the histograms of encrypted speech for speech type 1 and speech type 2 on the PC, and **Fig. 9** shows the histograms of encrypted speech for speech type 3 and speech type 4 on the smartphone. It can be seen that the histogram of the encrypted speeches by the proposed speech homomorphic encryption scheme is nearly uniformly distribution, which makes the statistical attacks difficult. The attackers cannot reveal the original histograms from the encrypted speeches. Therefore, it is practically feasible to achieve privacy protection of digital speeches in the public cloud not only on the PC but also on the smartphone.

6.4 Correlation of two adjacent samples

In order to further assess the encryption quality of our proposed speech homomorphic encryption algorithm, the correlation between two adjacent samples in the original speech and encrypted speech is used. We randomly select 15000 pairs of adjacent samples from the original speech and encrypted speech. The correlation coefficients are calculated by the Eq. (26) and Eq. (27). The correlation coefficients of four types selected speeches are calculated and tabulated in **Table 2**.

$$\operatorname{cov}(x,y)=\frac{1}{N}\cdot\sum_{i=1}^{N}\left(x_i-E(x)\right)\cdot\left(y_i-E(y)\right), \tag{26}$$

and

$$r(x,y)=\frac{\operatorname{cov}(x,y)}{\sqrt{D(x)}\cdot\sqrt{D(y)}} \tag{27}$$

where

$$E(x)=\frac{1}{N}\cdot\sum_{i=1}^{N}x_i\ ,\ E(y)=\frac{1}{N}\cdot\sum_{i=1}^{N}y_i\ , \tag{28}$$

and

$$D(x)=\frac{1}{N}\cdot\sum_{i=1}^{N}\left(x_i-E(x)\right)^2\ ,\ D(y)=\frac{1}{N}\cdot\sum_{i=1}^{N}\left(y_i-E(y)\right)^2 \tag{29}$$

$\operatorname{cov}(x,y)$ is the covariance between x and y, $E(x)$ and $E(y)$ are the means of x and y, $D(x)$ and $D(y)$ are the variances of x and y, and N is the number of speech samples involved in the calculations. Besides, x and y are the two adjacent sample values in the speech.

From **Table 2**, we can see that the encrypted speeches have small values of correlation coefficients, which demonstrate that the proposed speech homomorphic encryption scheme has low correlation and good encryption quality.

Table 2. Correlation coefficients of two adjacent quantitative values

Types of speech	Original speech	Encrypted speech
Male speech (Chinese)	0.9984	0.0746
Female speech (Chinese)	0.8220	0.1910
Male speech (English)	0.7813	-0.5314
Female speech (English)	0.9488	0.0027

6.5 Quality of decrypted speech

To evaluate perceptual quality of decrypted speech signals, there are three objective criteria: Signal to-Noise Ratio (*SNR*), segmental Signal-to-Noise Ratio (*SegSNR*) and the correlation coefficient. *SNR* is a measure that compares the level of a desired signal to the level of background noise; it can be applied to the speech signals. The *SegSNR* is determined by short segments of the signal, and is a good estimator for speech signal quality. The correlation coefficient is a number that quantifies a type of correlation and dependence, meaning statistical relationships between the original speech and the decrypted speech in fundamental statistics. In this paper, we implement these three objective metrics on 100 speeches in SQAM database [30]. The results are listed in **Table 3**, from which we can see that the decrypted speeches have good quality.

The definition of *SNR* is given as Eq. (30):

$$SNR = 10 \cdot \lg \left(\frac{\sum_{i=1}^{L} a(i)^2}{\sum_{i=1}^{L} \left(a(i) - a'(i) \right)^2} \right) \tag{30}$$

where $a(i)$ is the original speech signal, and $a'(i)$ is the decrypted speech signal.

Segmental Signal-to-Noise Ratio is defined as follows:

$$SegSNR = \frac{10}{W} \cdot \sum_{w=0}^{W-1} \lg \sum_{i=K \cdot w}^{K \cdot w + K - 1} \left(\frac{a(i)}{\left(a(i) - a'(i) \right)} \right)^2 \tag{31}$$

where W is the number of segments in the speech signal, K is the length of each segment, $a(i)$ and $a'(i)$ are the original speech signal and the decrypted speech signal respectively.

The correlation coefficients defined as Eq. (32).

$$R_{aa'} = \frac{\operatorname{cov}\left(a(i), a'(i) \right)}{\sqrt{D\left(a(i) \right)} \cdot \sqrt{D\left(a'(i) \right)}} \tag{32}$$

where $a(i)$ is the original speech signal, $a'(i)$ is the decrypted speech signal, and $R_{aa'}$ is the correlation coefficient.

Table 3. Metrics values for the decrypted speech quality

Quality metrics	Value
SNR (dB)	Inf
segSNR (dB)	Inf
$R_{aa'}$	1.0000

7. Conclusion

The privacy preservation and signal processing over encrypted domain of personal speech have become more and more important in the cloud computing. In this paper, one practical and low-complexity addition homomorphic symmetrical encryption scheme for speech signal is proposed. The proposed scheme firstly preprocesses the original speech as matrix and then encrypts the preprocessed speech signal by the proposed homomorphic encryption scheme. The encrypted speech is compressed in order to be stored in the cloud with less storage space used. The algorithm analysis and experimental results show that the proposed scheme is an

addition homomorphic encryption with good diffusibility and perfect randomness of speech cipher-texts. The proposed scheme also wipes off some speech characteristics of original speech, and it is robust to statistical analysis attacks. Compared with the popular Paillier cryptosystem, the proposed scheme has less data expansion and lower computational time complexity. In addition, the proposed scheme is fit for operation on the smartphone.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC) under the grant Nos. U1536110, 61402219.

References

[1] Y. Mao, J. Y. Wang and B. Sheng, "Skyfiles: Efficient and secure cloud-assisted file management for mobile devices," in *Proc. of IEEE Int. Conference on Communications*, pp. 4202-4207, June 10-14, 2014. Article (CrossRef Link).

[2] J. Y. Wang, T. Wang, Z. Y. Yang, Y. Mao, N. Y. Mi and B. Sheng, "SEINA: A stealthy and effective internal attack in Hadoop systems," in *Proc. of Int. Conference on Computing, Networking and Communications*, pp. 525-530, January 26-29, 2017. Article (CrossRef Link).

[3] A. Sinha, "Cloud-based mobile device security and policy enforcement," US 9609460. B2 [P], March 28, 2017.

[4] Y. J. Ren, J. Shen, J. Wang, J. Han and S. Y. Lee, "Mutual verifiable provable data auditing in public cloud storage," *Journal of Internet Technology*, vol. 16, no. 2, pp. 317-323, March, 2015. Article (CrossRef Link).

[5] T. H. Ma, J. J. Zhou, M. L. Tang, Y. Tian, A. Al-Dhelaan, M. Al-Rodhaan and S. Y. Lee, "Social network and tag sources based augmenting collaborative recommender system," *IEICE transactions on Information and Systems*, vol. E98-D, no. 4, pp. 902-910, April, 2015. Article (CrossRef Link).

[6] Y. Z. Ren, J. Yang, J. W. Wang and L. N. Wang, "AMR steganalysis based on second-order difference of pitch delay," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 6, pp. 1345-1357, June, 2017. Article (CrossRef Link).

[7] Y. Z. Ren, T. T. Cai, M. Tang and L. N. Wang, "AMR steganalysis based on the probability of same pulse position," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 9, pp. 1801-1811, September, 2015. Article (CrossRef Link).

[8] S. M. Eldin, S. A. Khamis, A. A. I. M. Hassanin and M. A. Alsharqawy, "New audio encryption package for TV cloud computing," *International Journal of Speech Technology*, vol. 18, no. 1, pp. 131-142, March, 2015. Article (CrossRef Link).

[9] A. Mostafa, N. F. Soliman, M. Abdalluh and F. E. A. EI-samie, "Speech encryption using two dimensional chaotic maps," in *Proc. of 11th Int. Computer Engineering Conference*, pp. 235-240, December 29-30, 2015. Article (CrossRef Link).

[10] H. X. Wang, L. N. Zhou, W. Zhang and S. Liu, "Watermarking-based perceptual hashing search over encrypted speech," in *Proc. of 12th Int. Workshop on Digital forensics and Watermarking*, pp. 423-434, October 1-4, 2013. Article (CrossRef Link).

[11] H. Hermassi, M. Hamdi, R. Rhouma and S. M. Belghith, "A joint encryption-compression codec for speech signals using the ITU-T G.711 standard and chaotic map," *Multimedia Tools and Applications*, vol. 76, no. 1, pp. 1177-1200, January, 2017. Article (CrossRef Link).

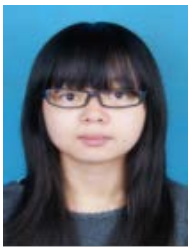
[12] M. Hamdi, R. Rhouma and S. Belghith, "An appropriate system for securing real-time voice communication based on ADPCM coding and chaotic maps," *Multimedia Tools and Applications*, vol. 76, no. 5, pp. 7105-7128, March, 2017. Article (CrossRef Link).

[13] H. J. Liu, A. Kadir and Y. L. Li, "Audio encryption scheme by confusion and diffusion based on multi-scroll chaotic system and one-time keys," *Optik*, vol. 127, no. 19, pp. 7431-7438, October, 2016. Article (CrossRef Link).
[14] S. Fahmeeda and A. Tabassum, "Audio data security and feature extraction over cloud," *International Journal of Computer Applications*, vol. 168, no. 10, pp. 33-37, June, 2017. Article (CrossRef Link).
[15] M. A. Yakubu, P. K. Atrey and N. C. Maddage, "Secure audio reverberation over cloud," in *Proc. of 10th Annual Symposium on Information Assurance*, pp. 39-43, June 2-3, 2015.
[16] S. J. Xiang and X. R. Luo, "Reversible data hiding in homomorphic encrypted domain by mirroring ciphertext group," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 11, pp. 3099-3110, November, 2018. Article (CrossRef Link).
[17] S. J. Xiang and X. R. Luo, "Efficient reversible data hiding in encrypted image with public key cryptosystem," *EURASIP Journal on Advances in Signal Processing*, vol. 2017, no. 1, pp. 2017(59), December, 2017. Article (CrossRef Link).
[18] X. P. Zhang, J. Long, Z. C. Wang and H. Cheng, "Lossless and reversible data hiding in encrypted images with public key cryptography," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 9, pp. 1622-1631, September, 2016. Article (CrossRef Link).
[19] P. Yang, X. L. Gui, J. An, F. Tian and J. C. Wang, "An encrypted image editing scheme based on homomorphic encryption," in *Proc. of Int. Conference on Computer Communications*, pp. 109-110, April 26-May 1, 2015. Article (CrossRef Link).
[20] Y. Zhang, L. Zhuo, Y. F. Peng and J. Zhang, "A secure image retrieval method based on homomorphic encryption for cloud computing," in *Proc. of 19th Int. Conference on Digital Signal Processing*, pp. 269-274, August 20-23, 2014. Article (CrossRef Link).
[21] Y. Y. Li, J. T. Zhou, Y. M. Li and O. C. Au, "Reducing the ciphertext expansion in image homomorphic encryption via linear interpolation technique," in *Proc. of Int. Conference on Signal and Information Processing*, pp. 800-804, December 14-16, 2015. Article (CrossRef Link).
[22] R. C. Hendriks, Z. Erkin and T. Gerkmann, "Privacy-preserving distributed speech enhancement for wireless sensor networks by processing in the encrypted domain," in *Proc. of Int. Conference on Acoustics, Speech and Signal Processing*, pp. 7005-7009, May 26-31, 2013. Article (CrossRef Link).
[23] M. A. Yakubu, N. C. Maddage and P. K. Atrey, "Encryption domain cloud-based speech noise reduction with comb filter," in *Proc. of Int. Conference on Multimedia & Expo Workshops*, pp. 1-6, July 11-15, 2016. Article (CrossRef Link).
[24] Z. Erkin, A. Piva, S. Katzenbeisser, R. L. Lagendijk, J. Shokrollahi, G. Neven and M. Barni, "Protection and retrieval of encrypted multimedia content: When cryptography meets signal processing," *EURASIP Journal on Information Security*, vol. 2007, no. 17, pp. 1-20, December, 2007. Article (CrossRef Link).
[25] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Proc. of Advances in Cryptology-EUROCRYPT'99*, pp. 223-238, May 2-6, 1999. Article (CrossRef Link).
[26] T. Bianchi, A. Piva and M. Barni, "Composite signal representation for fast and storage-efficient processing of encrypted signals," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 1, pp. 180-187, March, 2010. Article (CrossRef Link).
[27] C. P. Gupta and I. Sharma, "A fully homomorphic encryption scheme with symmetric keys with application to private data processing in clouds," in *Proc. of 4th Int. Conference on the Network of the Future*, pp. 1-4, October 23-25, 2013. Article (CrossRef Link).
[28] A. Kipnis and E. Hibshoosh, "Efficient methods for practical fully homomorphic symmetric key encryption, randomization and verification," *IACR Cryptology ePrint Archive*, 2012. Article (CrossRef Link).
[29] S. Goldwasser and S. Micali, "Probabilistic encryption," *Journal of Computer System Sciences*, vol. 28, no. 2, pp. 270-299, April, 1984. Article (CrossRef Link).
[30] EBU, "Sqam-sound quality assessment material,". Article (CrossRef Link).

Canghong Shi received the B. S. degree from Hebei Normal University, Shijiazhuang, in 2009, the M.S. degrees from Chengdu University of Information Technology, Chengdu, in 2014, respectively. He is currently pursuing the Ph.D. degree from Southwest Jiaotong University, Chengdu. His current research focuses on multimedia information security, digital audio signal forensics.

Hongxia Wang received the B.S. degree from Hebei Normal University, Shijiazhuang, in 1996, and the M.S. and Ph.D. degrees from University of Electronic Science and Technology of China, Chengdu, in 1999 and 2002, respectively. She engaged in postdoctoral research work in Shanghai Jiaotong University from 2002 to 2004, and worked in School of Information Science and Technology, Southwest Jiaotong University from 2004 to 2018. Currently she is a professor with College of Cybersecurity, Sichuan University, Chengdu. Her research interests include multimedia information security, digital forensics, information hiding and digital watermarking. She has published 100 peer research papers and wined 10 authorized patents.

Yi Hu is a Professor of Computer Science at Northern Kentucky University, USA. He has a Ph.D. degree (2006) in Computer Science from the University of Arkansas. His research concentrates on Information Assurance, Database Systems, Data Security, Data Mining, and Trust Management in Cyberspace. He is also a CISSP and CEH.



Qing Qian received the B. S., the M. S. and Ph.D. degree from Southwest Jiaotong University, Chengdu, in 2009, 2012 and 2018, respectively. She is currently pursuing the Ph.D. degree. Her current research focuses on digital watermarking and audio signal processing.

Hong Zhao received his B.S. and Ph.D. degrees in information security from the Southwest Jiaotong University, Chengdu, China, in 2007 and 2013, respectively. From 2010 to 2012, he was a visiting scholar at the University of Michigan-Dearborn. He is currently a research fellow with the Department of Electrical and Electronic Engineering, Southern University of Science and Technology. His current research interests include steganalysis, audio forensics, wireless communication security, etc.