International Journal of Security and Its Applications
Vol. 10, No. 1 (2016), pp.103-118
http://dx.doi.org/10.14257/ijsia.2016.10.1.11

Security Analysis of Speech Perceptual Hashing Authentication Algorithm

Zhang Qiu-yu, Ren Zhan-wei, Xing Peng-fei, Huang Yi-bo and Yu Shuang

School of Computer and Communication, Lanzhou University of Technology, Lanzhou, 730050, China
zhangqylz@163.com, 1092813613@qq.com

Abstract

Speech perceptual hashing authentication algorithm is an efficient method for content integrity authentication and identity authentication. But the algorithm becomes transparent under the principle of Kerckhoffs which makes the algorithm unsafe. In this paper, the algorithm is encrypted under the principle of Kerckhoffs to protect the security of the algorithm. Then the specific link that won't affect the performance of the algorithm is identified through the experiments. Next, this paper analyzes the security of the encrypted algorithm based on the concept of Shannon unicity distance. The unicity distance is figured out finally based on the experiments. That is to say, the algorithm loses its security even though the algorithm has been encrypted. Under this circumstances, the most important thing is to continue keeping the safety of the algorithm. Therefore, this paper proposed an efficient random secret key method to guarantee the safety of the algorithm after the unicity distance is figured out.

Keywords: *Speech authentication, Speech perceptual hashing, Unicity distance, Secret key, Security analysis*

1. Introduction

With the rapid development of information technology and Internet technology, the speech remote communication is becoming more and more convenient. But people can use multimedia software to edit and modify the digital products. An attacker can easily tamper and eavesdrop the speech information with the opening of communication channel of wireless and network. At the same time the source of the information is difficult to certification. So, it's necessary to carry out security authentication method for the speech communication [1-2].

The speech perceptual hashing authentication algorithm can protect speech messages by verifying content integrity and authenticity of speech information. It has robustness to content preserving operations and strict distinction to malicious tampering. At the same time, it has good security for communication. The algorithm can better realize the content integrity authentication of speech and broadband audio signal. Since put forward, this technology has been widely used because of its good performance [3-7], but most of them don't consider the security of the algorithm itself. For example, M. Nouri *et al.* [3] proposed a speech perceptual hashing algorithm based on secret key, so the algorithm has certain security and collision resistance. Y. Huang *et al.* [4] proposed a speech perceptual hashing algorithm based on modified Linear Prediction Analysis. This algorithm has high efficiency for authentication, but it has weak robustness. H. Wang *et al.* [5] proposed a speech remote identity authentication algorithm based on perceptual hashing through extracting fingerprint biometric, the fingerprint information that embedded into the algorithm would not affect the normal use of the speech and the experiment results show that it has good collision resistance, good robustness and security for remote

ISSN: 1738-9976 IJSIA

authentication. Q.Y. Zhang *et al*. [6] proposed an efficient robust speech perceptual hashing authentication algorithm based on wavelet packet decomposition. This algorithm has good robustness and distinction. N. Chen *et al*. [7] proposed a perceptual hashing algorithm based on discrete-wavelet-transform and non-negative matrix factorization. This algorithm has good robustness, but it has high FAR and weak distinction, could not resist the effects of low pass filter and noise.

At present, there are a few literatures about security of the speech perceptual hashing algorithm and many algorithms ensure their security by increasing their complexity. For example, Z. Liu *et al*. [8] proposed a modified perceptual hashing authentication algorithm based on fuzzy commitment scheme for the weak security of the algorithm. The improved perceptual hashing authentication algorithm enhances the security of each part and increases the complexity of the algorithm. It makes the improved algorithm having a certain security and improvement on the performance compared with before. But the algorithm ignores the premise which is called Kerckhoffs principle. Under the principle of Kerckhoffs the algorithm becomes transparent and an attacker can easily tamper or forge the speech information and not be found. That is why the perceptual hashing authentication algorithm needs new requirements to protect their security [9]. However, most of the existing encryption methods are proposed for specific algorithm and these methods don't consider the generality of quantitative analysis of security. It even doesn't give an explanation about the performance of the encrypted algorithm.

L. Liu *et al*. [10] use a typical robust video hashing algorithm as an example to mathematically model the feature extraction process. The author uses the Shannon unicity distance to measure the level of security of the algorithm considering of the different types of attacks and then quantitatively analyzes the security of the algorithm. The authors reveal a detailed introduction by the way of theory analysis and obtain the calculation formula of the unicity distance. But the authors don't give the specific experiment process to verify the correctness of the analysis. D. Hu *et al*. [11] proposed a secure architecture based on robust perceptual hashing for the weakness that traditional perceptual hashing algorithm included. The security protocols that the author designed based on the architecture further ensure that an attacker can not obtain the plaintext/hash value pair or secret key/hash value pair at the same time. It further ensures the security of the algorithm. But the architecture ignores the attacker's attack during the transmission process. J. Zhou *et al*. [12] proposed a key estimation method. The authors use the embedded virtual watermarking serving as secret key to protect the security of the algorithm. But the security analysis is based on specific perceptual hashing algorithm and the authors don't give a modified algorithm after the key is estimated. Y. Mao *et al*. [13] proposed a security analysis for the image perceptual hashing algorithm according to the principle of Shannon unicity distance and finally the unicity distance is worked out. However, the evaluation method has a high dependence on specific algorithm and different algorithm has different amount of calculation. O. Koval *et al*. [9,14] proposed to use Shannon equivocation to analyze the security of perceptual hashing. The difference is that in [9] the authors only present the upper bound of information leakage and in [14] the authors give a more detailed description. But both the conclusions are proposed based on theory analysis and the general experiments are not revealed. M. Long *et al*. [15] proposed the security analysis for the collisions and drawbacks of the algorithm of the chaotic one-way hash function. It takes corresponding prevention countermeasures, improvement measures and safety design principles to research. But its analysis is not comprehensive and the algorithm's revise is designed according to specific attack, which don't have universality. Y.H. Jiao *et al*. [16] proposed two indexes of security analysis according to the research on perceptual hashing. One is entropy rate and the other is the ability of weak collision resistance. But the authors don't present explanations for the specific algorithm. Y.W Liu *et al*. [17] analyzed the security of perceptual hashing from four aspects. The four aspects are unipolarity, scrambling, diffusion and transmission security. Although the authors

analyze the concepts of the four aspects, the authors don't present specific analysis method and analysis index. From the above analyses we can see that most of the security analyses are proposed based on specific perceptual hashing algorithm which is related with secret key. This is because under the principle of Kerckhoffs the algorithm should uses the key to protect its security. But the above analyses don't reveal the relationship between the algorithm and the key, and don't present specific improvement measures to protect the security of perceptual hashing algorithm after the process of security analysis.

Based on above, this paper presents the quantitative security analysis for the speech perceptual hashing algorithm in [6] based on the Shannon unicity distance. Firstly, the algorithm in [6] is encrypted to guarantee the security of the algorithm under the principle of Kerckhoffs. Secondly, the encryption scheme is confirmed that won't affect the performance of the algorithm through the experiment simulation. Thirdly, the security of encrypted algorithm is analyzed based on the Shannon unicity distance and the unicity distance is calculated. Finally, this paper presents a modified encryption scheme according to the unicity distance to continue protecting the security of the algorithm. The experiment results illustrate that the encrypted speech perceptual hashing algorithm still satisfies the performance requirements of hash function, such as real-time property, robustness and safety under the mobile computing environment.

2. Problem Statements and Preliminaries

The speech perceptual hashing algorithm mainly consists of three parts—the random selection of perceptual feature, the subsequent processing of feature value and the generation of perceptual hash value. The security of speech perceptual hashing algorithm is mainly reflected on the encryption process to these three parts. And the most important part in these three parts is the feature extraction stage [9,11,12,14,16,17]. That's because the robustness and security of speech perceptual hashing algorithm are mainly reflected on this part. But under mobile computing environment, the speech perceptual hashing authentication algorithm is subject to certain restrictions, such as the requirement of real-time, the requirement of computing ability and the storage capacity of the mobile terminals, and so on. Therefore, it is not realistic to encrypt the perceptual hashing algorithm in the feature extraction stage based on the present situation of mobile terminals. Besides, the encryption process that conducted in feature extraction stage should aim at specific algorithm and consider specific encryption process. This point greatly increases the complexity of the algorithm and the time consumption. So, most algorithms only use simple randomization process to protect their security in feature extraction stage. Based on the above reasons, this paper doesn't consider encryption process in feature extraction stage under mobile computing environment.

The subsequent processes of feature value of speech perceptual hashing algorithm generally include quantization, compression and other operations. In this process, the feature value of perceptual hashing can be further randomized, such as the confusion and diffusion operations of the feature value. However, in order to remain the well robust feature value that received at the feature extraction stage, the randomization operation in this stage is limited. For example, if the feature value is encrypted based on the cryptography, the little change or even one bit change of the feature value all would lead to a lot of change of the perceptual hash value that received at last. Because the inherent structure of the feature value will be changed after various operations on feature value. At last, the perceptual hash value that received after content preserving operations will be also changed. So, the most important point in speech perceptual hashing is how to keep stability of the inherent structure of feature value. Therefore, in order to guarantee the stability of inherent structure of feature value, this paper mainly focuses on the generation stage of the final perceptual hashing value. This is because the majority of perceptual hash values are expressed as binary form and the encrypted binary sequence based on certain

rule satisfies the requirements of robustness and distinction. It also will be found in the following experiments that the encryption process on the feature value can lead to great change in the end and seriously influence the purpose of the experiments.

This paper uses $h(\bullet)$ to represent speech perceptual hashing function, $\boldsymbol{I}$ represents the input speech segment, $\boldsymbol{K}$ represents the secret key, the output speech perceptual hash vector is represented by $V=h_k(I)$. When repeatedly use the same key to encrypt n perceptual hash values, there can be received n speech hash value pairs, represented by (I_1, V_1), (I_2, V_2), …, (I_n, V_n). If an attacker can intercept the speech hash value pairs in the process of attacking, the attacker would estimate the uncertainty of the secret key according to the conditional entropy $H\left(K\left|\sum_{j=1}^{n}(I_j|V_j)\right.\right)$.

According to the knowledge of information theory, the uncertainty of $\boldsymbol{K}$ will reduces gradually with the increasing of n. The $\boldsymbol{K}$ will be estimated completely when n reaches to a certain limit, that is to say the uncertainty of $\boldsymbol{K}$ becomes zero. So, this paper mainly focuses on the reducing of uncertainty of secret $\boldsymbol{K}$ according to the increasing of speech hash value pairs. The experiment ideas of this paper are explained as following: first, the perceptual hash value sequence should be encrypted by a randomly selected key. Then, the perceptual hash value and the encrypted perceptual hash value are sent into a key estimation algorithm and then an estimated key can be received. The accuracy of the estimated key will increase with the increasing of the observed speech hash value pair. We wonder that the estimated key more and more close to the actual key until the estimated key and the actual key are exactly the same.

The encrypted perceptual hash value that comes from the original perceptual hash value and the encrypted perceptual hash value that comes from the content preserving operations of the original perceptual hash value should be similar due to the robustness constraints of constructing the speech perceptual hashing algorithm. This suggests that the secret key can be estimated approximately when observing one speech and its corresponding encrypted perceptual hash value. Then the estimated key can be used to encrypt another perceptual hash value and the encrypted perceptual hash value can be received. The distance d_K of the estimated key and the actual key can be reflected by the distance d_H of the two encrypted perceptual hash values, conversely too. These two parameters reflect the basic of the key estimation algorithm in this paper.

3. Secret Key-based Speech Perceptual Hashing Authentication Algorithm

This paper analyzed the security of efficient speech perceptual hashing authentication algorithm based on the algorithm proposed in [6]. Then the security analysis indexes of speech perceptual hashing algorithm can be obtained. This paper simply introduced the flowchart of the algorithm that taken as the experiment object in [6]. Then the specific encryption link that will not affect the performance of algorithm was found out through the experiment simulation. Considering that the requirements of real-time and the requirements of resource limitations of mobile terminals are quite high under the mobile computing environment, the encryption process that used in this paper should not be too complicated.

3.1. The Efficient Speech Perceptual Hashing Authentication Algorithm Based on Wavelet Packet Decomposition

In [6], the author presented a speech perceptual hashing authentication algorithm based on wavelet packet decomposition for speech content authentication. Firstly, a wavelet packet coefficient matrix of wavelet reconstruction is generated from an original speech signal which experienced high frequency pre-processing followed by wavelet packet

decomposition. Secondly, the wavelet packet coefficient matrix is partitioned into equal-sized sub-matrixes, and each sub-matrix is converted into new sub-matrix which has good decorrelation and energy compaction by two-dimensional discrete cosine transform (DCT). Finally, the feature parameter matrix is obtained by QR Decomposition (QRD) using a Givens Rotation (GR) on the new sub-matrix. The experiment results illustrate that the proposed algorithm is very robust in content preserving operations and it has very low hash bit rate. Besides, it can meet the requirements of real-time speech authentication with high certification efficiency. The flow chart of the perceptual hashing algorithm proposed in [6] is shown in Figure 1.

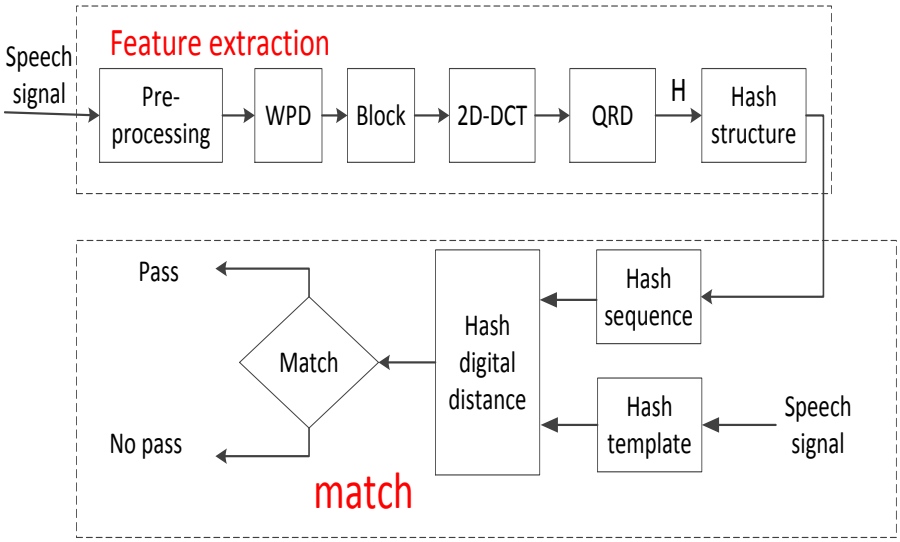


Figure 1. The Flow Chart of Speech Perceptual Hashing Authentication Algorithm in [6]

3.2. The Choice of Encryption Scheme

The purpose of encryption to the algorithm is to guarantee the safety of the algorithm, because the algorithm becomes transparent under the principle of Kerckhoffs. That is to say every attacker can obtain the specific algorithm. This makes the algorithm no longer safe. Now, the effective way to guarantee the security of the algorithm is the encryption process to the algorithm.

From the introduction of Section 2 we know that speech perceptual hashing algorithm can be divided into three parts. The encryption process to the algorithm is mainly reflected on the encryption of these three parts. As shown in Figure 2, the red dotted box of the figure represents the link that can be encrypted. Due to the real-time requirement of speech perceptual hashing algorithm under the mobile computing environment, the encryption to the perceptual feature selection stage is no longer applicable, because the time consumption and the resource consumption of the mobile terminals will be greatly increased. So, at this stage the simple randomization operations are often applied without using encryption process. Therefore, this paper mainly focuses on the encryption process to the perceptual feature value and the perceptual hash value, because the perceptual feature value and the perceptual hash value occupy less storage resources. Besides, it meets the requirement of mobile computing environment because the amount of calculation that needed to encryption are small. It's important to note that no matter choose which link to encrypt, the performance of encrypted perceptual hashing algorithm can't be lower than the performance that before encryption.

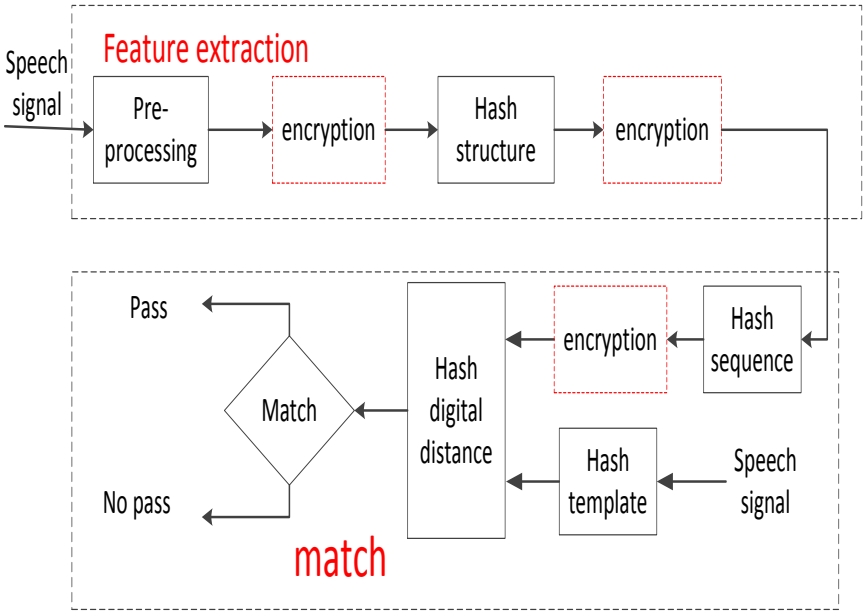


Figure 2. The Encryption Scheme

In this section, the perceptual feature value is encrypted first. Then the performance will be compared before and after encryption. Finally, the analysis of the performance after encryption is simply introduced. After above processes to the perceptual feature value, we do the same thing to the perceptual hash value. At the same time, in order to make the encryption process as simple as possible and have lower computational complexity, this paper uses the random number encryption scheme to encrypt.

3.2.1. The Encryption of the Perceptual Feature Value: The algorithm is slightly modified in the generation stage of perceptual feature value. The encryption process and the perceptual hashing algorithm are separated, that is to say the perceptual hashing algorithm and the encryption process are divided into two independent processes. The feature value is directly sent to the encryption process part to encrypt as soon as the feature value is calculated through the perceptual hashing algorithm. When the encryption process is finished, the encrypted feature value is directly returned to the perceptual hashing algorithm without any middle link. This operation further guarantees the security of the algorithm and prevents the leakage of information. The encryption process is shown in Figure 3.

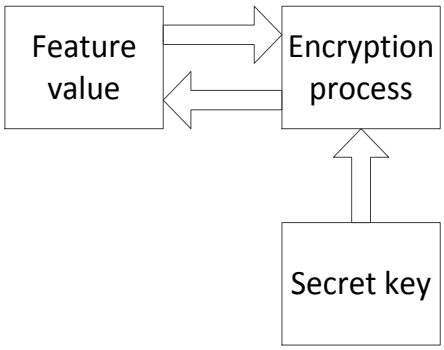


Figure 3. The Encryption Process

In order to adapt to the various requirements of the mobile computing environment, the encryption algorithm should be simple and fast in the encryption link. Therefore, this paper uses the random number generator to generate a random number encryption matrix to encrypt the feature value.

Suppose that the received perceptual feature value after feature extraction is a sequence with dimension $m\times1$, denoted by $\boldsymbol{A}_{m\times1}$. So, the random number matrix with dimension $m\times m$ is generated from the random number generator in the link of encryption, denoted by $\boldsymbol{S}_{m\times m}$. This matrix is generated totally random and doesn't subject to any rule. A new feature value whose size is same to the original feature value will be received after the matrix $\boldsymbol{S}$ and the sequence $\boldsymbol{A}$ to multiply, denoted by $\boldsymbol{D}_{m\times1}$. The encryption process is shown as follows

$$D_{m\times1} = S_{m\times m} A_{m\times1} \tag{1}$$

Among the equation, $\boldsymbol{D}$ is the new feature value of encrypted. The sequence $\boldsymbol{D}$ is directly sent back to the perceptual hashing algorithm to continue the subsequent processing of feature value. Finally, the perceptual hash value can be obtained. It can be seen from the comparison between this perceptual hash value and the original perceptual hash value that these two perceptual hash values are obviously different. That is to say the final perceptual hash value is actually changed after encryption.

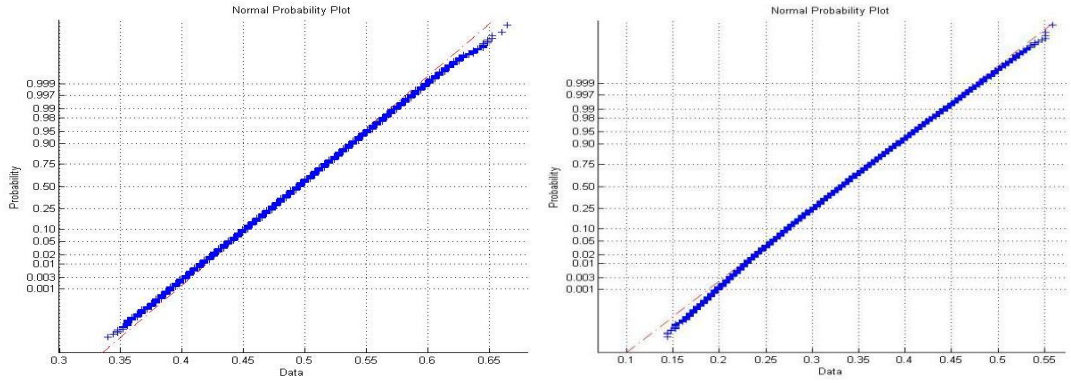
The performance of speech perceptual hashing algorithm of before and after encryption is compared through the experiments. The performance indexes used in here are the basic performance indexes of perceptual hashing, including Bit Error Rate (BER), False Accept Rate (FAR) and False Reject Rate (FRR) [6].

BER: Can reflect the percentage of error bits in the total number of bits. It is widely used in the research area of perceptual hashing as evaluation criterion of robustness.

FAR: Refers to the percentage that the speeches of different perceptual content are considered as the speeches of same perceptual content and mistakenly accepted by the system.

FRR: Refers to the percentage that the speeches of same perceptual content are considered as the speeches of different perceptual content and mistakenly rejected by the system.

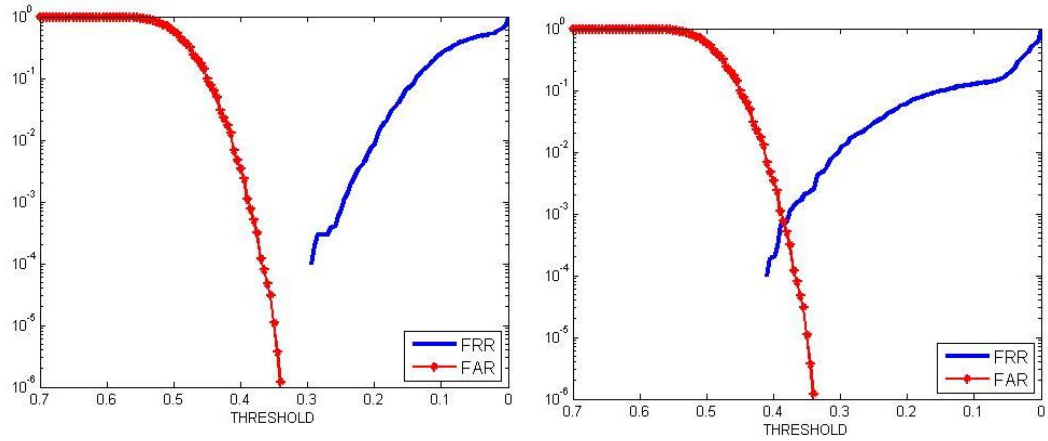
The FAR-FRR curve reflects the relationship of robustness and distinction of perceptual hashing algorithm. We say that the algorithm has good robustness and distinction when the curve has no intersections. That is to say the algorithm has good robustness to the content preserving operations and has good distinction to malicious tampering. The BER comparison of before and after encryption of the speech signal is shown in Figure 4, the performance comparison of FAR-FRR of before and after encryption is shown in Figure 5. The experiment process is completely same to the process in [6].



(a) The BER of Different Speech Before Encryption

(b) The BER of Different Speech After Encryption

Figure 4. The BER Comparison of Speech Signal before and after Encryption



(a) The FAR-FRR Curve Before Encryption

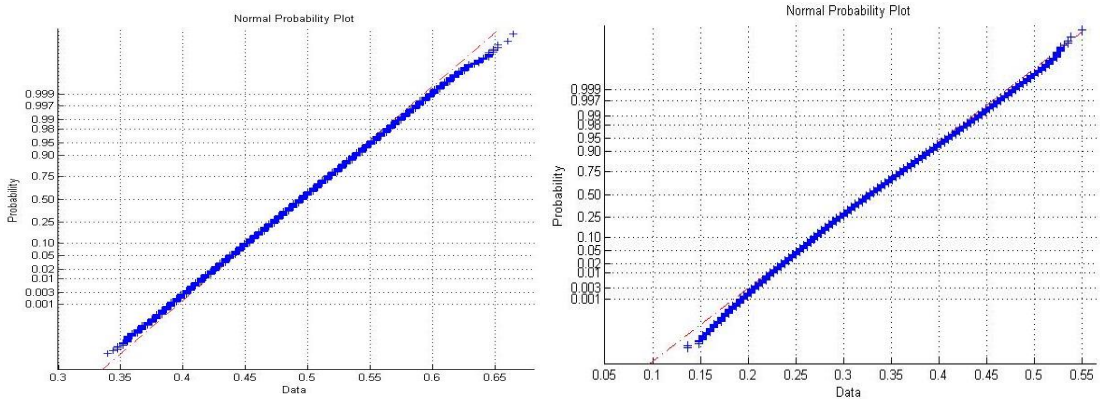
(b) The FAR-FRR Curve After Encryption

Figure 5. The Performance Comparison of FAR-FRR before and after Encryption

It can be seen from the performance comparisons of before and after encryption of Figure 4 and Figure 5 that the process of using random number matrix to encrypt perceptual feature value seriously influence the performance of the perceptual hashing algorithm.

The random number matrix we used above that generated from the random number generator doesn't subject to any rule. At this moment, we use random number generator to generate a random matrix that conforms to standard normal distribution with mean 0 and variance 1. Then this matrix is used to encrypt the perceptual feature value just as above. The performance indexes are observed before and after encryption.

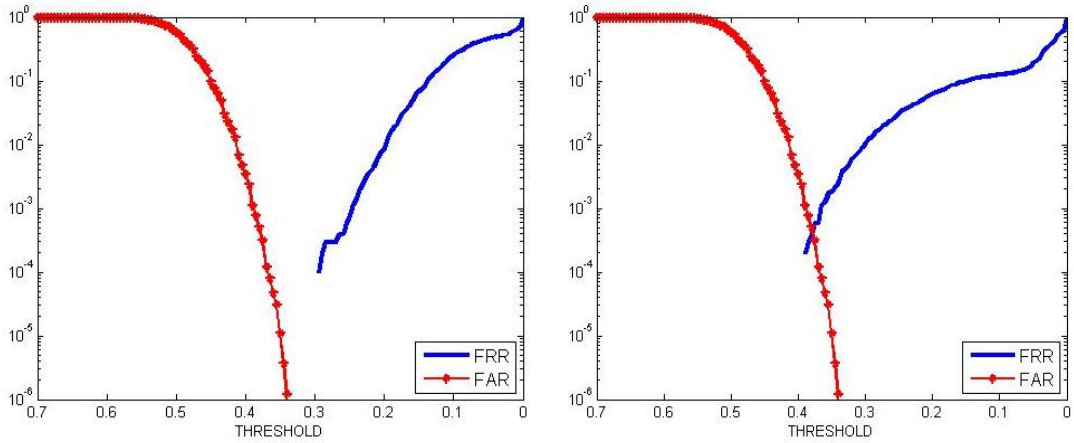
It can be seen from the performance comparisons of Figure 6 and Figure 7 of before and after encryption that the performance has certain improvement after using the random number matrix that has certain regularity to encrypt the perceptual feature value. However, the performance still doesn't satisfy the requirements of perceptual hashing comparing with the original algorithm.



a) The BER of Different Speech Before Encryption

(b) The BER of Different Speech After Encryption

Figure 6. The BER Comparison of Speech Signal before and after Encryption



(a) The FAR-FRR Curve Before Encryption

(b) The FAR-FRR Curve After Encryption

Figure 7. The Performance Comparison of FAR-FRR before and after Encryption

Based on above discussions we can conclude that the performance of perceptual hashing will be seriously influenced when using random number to encrypt the perceptual feature value. Just as section 2 has talked about that it is not suitable for speech perceptual hashing authentication algorithm that uses random number to encrypt perceptual feature value. Because the inner structure of the perceptual feature value will be changed when using random number to encrypt perceptual feature value. This will lead to great change on the final perceptual hash value. For example, there are two speeches α and β, β is the speech that comes from content preserving operations of α. These two speeches are sent to the perceptual hashing algorithm that includes encryption link. Ideally, these two perceptual hash values that finally generated from the algorithm should be very similar, but the reality is that these two perceptual hash value sequences are very different. Because the encryption process of the perceptual feature value changes the inner structure of the feature value and makes the final perceptual hash value changed a lot.

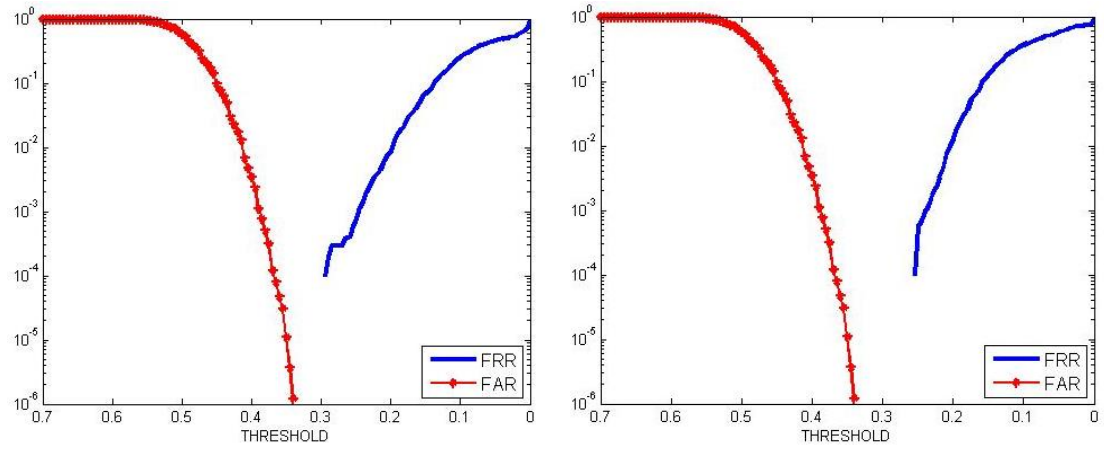
Therefore, the random number encryption of perceptual feature value will seriously influence the performance of perceptual hashing algorithm. As a result, the encryption of perceptual feature value doesn't satisfy the requirements of this paper.

3.2.2. The Encryption of Perceptual Hash Value: According to the discussion of section 3.2.1, the encryption scheme of perceptual feature value based on the random number matrix can't satisfy the requirements of perceptual hashing algorithm. So, this section discusses the encryption scheme of perceptual hash value after hash sequence construction. The encryption process in this stage is similar to the process of perceptual feature value stage. The calculated perceptual hash value is directly sent to the encryption process part and encrypted by the random number matrix that generated from the random number generator. It's important to note that the function of random number matrix generated by the random number generator is different from the random number matrix that proposed in the encryption scheme of perceptual feature value. In this stage, the function of random number matrix is equivalent to confusion operations and this matrix is called random number confusion matrix. The final perceptual hash value that generated from the perceptual hashing algorithm is binary sequence with dimension $m\times1$, denoted by $\boldsymbol{Z}_{m\times1}$. The generated random number confusion matrix should be with dimension $m\times m$, denoted by $\boldsymbol{X}_{m\times m}$. Multiply matrix $\boldsymbol{X}$ with sequence $\boldsymbol{Z}$ and the encrypted perceptual hash value with dimension $m\times1$ can be received, denoted by $\boldsymbol{C}_{m\times1}$. The generation process is shown as follows:

$$C_{m\times1} = X_{m\times m} Z_{m\times1} \tag{2}$$

The $\boldsymbol{C}$ in equation (2) is still a binary sequence, because the random number matrix generated from the random number generator is generated according to specific requirements. The function of the matrix is to confusion the perceptual hash value, but it doesn't change the value of the perceptual hash sequence. The final encrypted perceptual hash value and the speech information are sent to the receiver after encryption process.

It can be seen that the performance of encrypted speech perceptual hashing algorithm doesn't reduce comparing with before encryption through the experimental proof. This is because the perceptual hash value is calculated from perceptual hashing algorithm and inherited the robustness and distinction of perceptual hashing. That is to say the perceptual hash value that go through content preserving operations has very high robustness and the perceptual hash value that go through malicious tampering has very high distinction. The performance comparisons of FAR-FRR before and after encryption are shown in Figure 8.



(a) The FAR-FRR Curve Before Encryption

(b) The FAR-FRR Curve After Encryption

Figure 8. The Performance Comparison of FAR-FRR before and after Encryption

It can be concluded from the above discussions that the encryption on perceptual feature value will seriously affect the performance of perceptual hashing algorithm. However, the encryption process of perceptual hash value won't affect the performance of perceptual hashing algorithm. This point is also discussed in [16]. Y.H. Jiao *et al*. [16] pointed out that when the relationship between secret key and input information is small, the performance of resisting weak collision is strong. This won't affect the performance of perceptual hashing and obviously improves its security. Y.W. Liu *et al*. [17] also pointed out that the security mainly depended on secret key and the specific optimization of attacks on local content. Therefore, this paper chooses the encryption scheme of perceptual hash value. The encryption process that applied in this paper is just simple random number encryption scheme and its purpose is to guarantee the good working under mobile computing environments. In reality, any kinds of encryption schemes that won't affect the performance of perceptual hashing algorithm can be used to guarantee the security of the algorithm.

4. Proposed Scheme

The perceptual hash value is called plaintext before encryption, denoted by $\boldsymbol{I}$. The perceptual hash value is called ciphertext after encryption, denoted by $\boldsymbol{V}$. The purpose of the security analysis of hash algorithm is to calculate the Shannon unicity distance and then the secret key used in encryption process can be estimated based on the Shannon unicity distance. This paper firstly estimates the secret key, then this key is used to calculate the encrypted perceptual hash value, finally the mathematical distance of the estimated perceptual hash value and the actual perceptual hash value is used to reflect the distance of estimated secret key and the actual secret key.

The steps of security analysis of encrypted speech perceptual hashing authentication algorithm are shown as follows:

Step 1: We say that an attacker gets a plaintext/ciphertext pair when the attacker intercepts a set of speech signal and perceptual hash value. It can be shown as follows:

$$(I_1, V_1) \tag{3}$$

Assume that the received plaintext and ciphertext are sequence with dimension $m\times 1$, the secret key is a matrix $\boldsymbol{K}$ with dimension $m\times m$. So, an encryption equation can be received and shown as follows:

$$V_1 = KI_1 \tag{4}$$

It also can be shown as follows:

$$V_{m\times 1} = K_{m\times m} I_{m\times 1} \tag{5}$$

According to the theory of linear algebra there is

$$V_1 I_1^{-1} = K \tag{6}$$

The $\boldsymbol{K}$ can't be calculated in this stage because there is no inverse matrix of $\boldsymbol{I_1}$. But at this moment the least square method can be used to approximately estimate the secret key. The secret key $\boldsymbol{K'}$ can be approximately estimated according to the concept of the least square method. The received secret key $\boldsymbol{K'}$ can be used to encrypt the plaintext $\boldsymbol{I_1}$ and then the ciphertext $\boldsymbol{V_1'}$ is obtained. The mathematical distance d_{H} of $\boldsymbol{V_1}$ and $\boldsymbol{V_1'}$ is calculated according to the equation

$$d_H = \frac{V_1 \oplus V_1'}{m} \tag{7}$$

Then the d_H is used to approximately reflect the distance d_K of real secret $\boldsymbol{K}$ and estimated secret key $\boldsymbol{K'}$.

Step 2: When the attacker intercepts a pair of plaintext and ciphertext again, the attacker can obtain another encryption equation

$$V_{m\times 2} = K_{m\times m} I_{m\times 2} \tag{8}$$

At this moment the specific value of secret $\boldsymbol{K}$ still can't be calculated. Based on above, here the least square method still can be used to approximately calculate the secret key $\boldsymbol{K'}$. The estimated secret key $\boldsymbol{K'}$ is used to encrypt the plaintext $\boldsymbol{I_2}$ and then the ciphertext $\boldsymbol{V_2'}$ can be obtained. The mathematical distance d_H of $\boldsymbol{V_2}$ and $\boldsymbol{V_2'}$ is calculated to approximately reflect the distance d_K of real secret key $\boldsymbol{K}$ and estimated secret key $\boldsymbol{K'}$.

Step 3: Assume that the number of plaintext-ciphertext pairs that the attacker intercepted are n, denoted by

$$(I_1, V_1), (I_2, V_2), \cdots, (I_n, V_n) \tag{9}$$

There exist two situations:

The first situation: the number of plaintext/ciphertext pairs that the attacker intercepted less than m, that is to say $n<m$. It can be expressed by linear equation

$$V_{m\times n} = K_{m\times m} I_{m\times n} \tag{10}$$

When $n<m$, there are no solutions of the equation. At this moment, the concept of the least square method still can be used to calculate the mathematical distance d_H of $\boldsymbol{V}_n$ and $\boldsymbol{V}_n'$. The calculated d_H can be applied to approximately reflect the distance d_K of real secret key $\boldsymbol{K}$ and the estimated secret key $\boldsymbol{K'}$.

The second situation: the number of plaintext/ciphertext pairs that the attacker intercepted is equal or greater than m, that is to say $n\geq m$. At this moment we make $n=m$ and obtain the equation

$$K_{m\times m} = V_{m\times m} I^{-1}_{m\times m} \tag{11}$$

Now, the specific value of $\boldsymbol{K}$ can be calculated.

The mathematical distance of estimated secret key and real secret key is shown in Figure 9. The horizontal axis shows the number of plaintext/ciphertext pairs that the attacker intercepted, the vertical axis shows the normalization mathematical distance of estimated secret key and real secret key, that is to say d_K. This paper assumes that m is equal to 256.

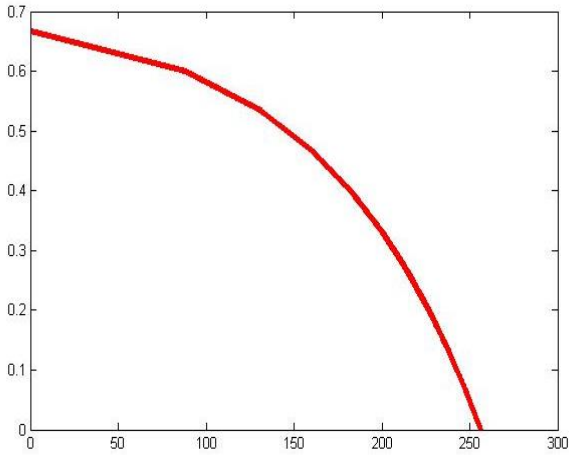


Figure 9. The Mathematical Distance of Estimated Secret Key and Real Secret Key

It can be seen from the Figure 9 that with the increasing of the number of plaintext/ciphertext pairs that the attacker intercepted the mathematical distance of estimated secret key and real secret key is correspondingly decreasing, until m=256. That is to say the normalization mathematical distance of estimated secret key and real secret key reduced to zero when the number of plaintext/ciphertext pairs that the attacker intercepted reaches to 256 or more. The estimated secret key at this moment is the real secret key. We say that the minimum number of m that needed to obtain the plaintext/ciphertext pairs is the unicity distance of this encryption scheme according to the concept of Shannon unicity distance. We say that the algorithm loses its security when an attacker calculates the unicity distance.

From above discussions we know that the secret key can be estimated by an attacker without error when using the same secret key to encrypt reaches a certain limit in encryption algorithm. At this moment we say that the algorithm loses its security. So, how to continue keeping the security of the algorithm at this moment? There are two points needed to pay attention: the first point is the value of m. It can be seen from the above discussions that with the increasing of m the number of plaintext/ciphertext pairs that needed to observe by the attacker should also be increasing. This obviously increases the difficulty of the attacker to estimate the key. But if the value of m is too large it will accordingly increases the amount of calculation. This is not suitable for mobile computing environment. Therefore, the selection of m should comprehensively consider the actual situation of the amount of calculation and the application environment. The second point is the using of secret key. This paper uses the same secret key, that is to say repeatedly using the same key to encrypt the algorithm. So, the secret key can be estimated without error when the attacker intercepts enough plaintext/ciphertext pairs. Therefore, in order to protect the security of the algorithm, the random key method can be used to encrypt in the stage of encryption process. That is to say the random number generator generates a new random key when the number of repeatedly using the same key to encrypt reaches a certain limit. Then the new random key will be applied to encrypt. The aim is to avoid repeatedly using the same key too many times so that the key can be figured out by the attacker.

5. Conclusions

This paper proposed a security analysis method for the speech perceptual hashing algorithm. The experimental results show that the encrypted speech perceptual hashing algorithm is still robust to content preserving operations and the encryption process won't

affect the original performance of speech perceptual hashing algorithm. In order to quantitatively analyze the security of speech perceptual hashing algorithm, this paper analyzed the security of encrypted speech perceptual hashing algorithm based on the Shannon unicity distance. The unicity distance is calculated and obtained an important conclusion: the number of times that reusing the same key to encrypt can't exceed the unicity distance, or the secret key can be estimated without error. That is to say it is necessary to efficiently avoid reusing the same secret key too many times. Based on this, this paper proposed a solution that use random key to encrypt. This action efficiently improved the security of the algorithm. The key point of this paper is to determine the specific link that won't affect the performance of the algorithm after encryption. In addition, this paper presents a method of how to continue keeping the security of algorithm after the unicity distance is calculated.

In this paper, the research is established on the basis of the experimental simulation. Therefore, the research point of next step is to combine the theoretical analysis of security and the experiments of perceptual hashing.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 61363078), the Natural Science Foundation of Gansu Province of China (No. 1212RJZA006, No. 1310RJYA004). The authors would like to thank the anonymous reviewers for their helpful comments and suggestions.

References

[1] N. Côté and J. Berger, Speech Communication, Edited Sebastian Möller and Alexander Raake, Springer International Publishing **(2014)**, Vol.12, pp.165-177.
[2] Adibi, "A low overhead scaled equalized harmonic-based voice authentication system", Telematics and Informatics, vol.31, no.1, **(2014)**, pp.137-152.
[3] M. Nouri, Z. Zeinolabedini, B. Abdolmaleki and N. Farhangian. Analysis of a novel audio hashfunction based upon stationary wavelet transform. Proceedings of the 6th IEEE International Conference on Application of Information and Communication Technologies (AICT), **(2012)** October 17-19; Tbilisi, Georgia, pp.1-6.
[4] Y. Huang, Q. Zhang and Z. Yuan, "Perceptual Speech Hashing Authentication Algorithm Based on Linear Prediction Analysis", TELKOMNIKA Indonesian Journal of Electrical Engineering, vol.12, no.4, **(2014)**, pp.3214-3223.
[5] H. Wang, L. Zhou, W. Zhang and S. Liu, Watermarking-Based Perceptual Hashing Search Over Encrypted Speech, Edited Yun Qing Shi, Hyoung-Joong Kim, Fernando Pérez-González, Springer Berlin Heidelberg **(2014)**, Vol.30, pp.423-434.
[6] Q.Y. Zhang, P.F. Xing, Y.B. Huang, R.H. Dong and Z.P. Yang, "An Efficient Speech Perceptual Hashing Authentication Algorithm Based on Wavelet Packet Decomposition", Journal of Information Hiding and Multimedia Signal Processing, vol.6, no.2, **(2015)**, pp.311-322.
[7] N. Chen, W.G. Wan and H.D. Xiao, "Robust audio hashing based on discrete-wavelet-transform and non-negative matrix factorisation", Communications, IET, vol.4, no.14, **(2010)**, pp.1722-1731.
[8] Z. Liu, Q. Li and X.M. Niu, "Improve the security of image robust hash using fuzzy commitment scheme", Neural Computing and Applications, vol.23, no.1, **(2013)**, pp.67-72.
[9] . Koval, S. Voloshynovskiy, F. Beekhof and T. Pun. Security analysis of robust perceptual hashing. Proceedings of the SPIE 6819, Security, Forensics, Steganography, and Watermarking of Multimedia Contents X. International Society for Optics and Photonics, **(2008)** February 26; pp.681906-681906-10.
[10] L. Liu and Y.Q. Wang, "Security Analysis of Video Hashing", Journal of Southwest Jiaotong University, vol.47, no.4, **(2012)**, pp.675-679.
[11] D.H. Hu, B. Su, S.L. Zheng and Z. Zhang. Secure Architecture and Protocols for Robust Perceptual Hashing. Proceedings of the 9th IEEE International Conference on Computational Intelligence and Security (CIS2013), **(2013)** December 14-15; Leshan, China, pp.550-554.
[12] J. Zhou and O.C. Au. Security evaluation of a perceptual image hashing scheme based on virtual watermark detection. Proceedings of the IEEE International Conference on Multimedia and Expo (ICME2011), **(2011)** July 11-15; Barcelona, Spain, pp.1-6.
[13] Y. Mao and M. Wu, "Unicity distance of robust image hashing", Information Forensics and Security, IEEE Transactions on, vol.2, no.3, **(2007)**, pp.462-467.

[14] O. Koval, S. Voloshynovskiy, P. Bas and F. Cayre. On security threats for robust perceptual hashing. Proceedings SPIE 7254, Media Forensics and Security, 72540H, **(2009)** February 04; International Society for Optics and Photonics, pp.72540H-72540H-13.
[15] M. Long and H. Wang, "Collision Analysis and Improvement of a Parallel Hash Function based on Chaotic Maps with Changeable Parameters", International Journal of Digital Crime and Forensics (IJDCF), vol.5, no.2, **(2013)**, pp.23-34.
[16] Y.H. Jiao, M.Y. Li, Q. Li and X.M. Niu. Key-dependent compressed domain audio hashing. Proceedings of the 8th IEEE International Conference on Intelligent Systems Design and Applications(ISDA'08), **(2008)** November 26-28; Kaohsiung, Taiwan, vol.3, pp.29-32.
[17] Y.W. Liu, "Research on Speech Perceptual Hashing Authentication Algorithm and Security Analysis Based on Compressed Domain", MS Thesis, Lanzhou University of Technology, Lanzhou, China, **(2014)**.

