International Journal of Security and Its Applications
Vol. 10, No. 1 (2016), pp.229-240
http://dx.doi.org/10.14257/ijsia.2016.10.1.21

Security Authentication Method of Speech Perceptual Hashing Based on Fuzzy Commitment Scheme

Zhang Qiu-yu, Ren Zhan-wei, Huang Yi-bo, Yu Shuang and Hu Wen-jin

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

Abstract

Existing speech perceptual hashing authentication algorithms generally encrypt the perceptual hash value to protect the security of the algorithms. But under the principle of Kerckhoffs, the algorithm becomes transparent and fragile. And it can be seen that the secret key may be estimated when the number of times that reuses the secret key reache to a limit through the safety analysis of Shannon unicity distance. To solve these problems, we present a novel security authentication scheme of speech perceptual hashing based on the fuzzy commitment scheme (FCS). Firstly, in sender, the randomly selected BCH code word and the extracted perceptual hash value are sent to the Cmt() function to calculate and the final secure perceptual hash value (h(c),δ) can be obtained. Secondly, in receiver, the perceptual hash value is calculated again from the received speech information and then the calculated perceptual hash value with the received information δ (commitment) are sent to the Decmt() function to calculate to get the h(c'). Finally, the h(c) and h(c') are matched. The experimental results show that the proposed scheme effectively avoids the probability that an attacker obtains the plaintext/ciphertext pairs without affecting the performance of original algorithm, and at the same time it ensures the security of perceptual hashing algorithm.

Keywords: *Multimedia security, Speech perceptual hashing, Fuzzy commitment scheme (FCS), Security*

1. Introduction

The perceptual hashing technology which satisfies the requirements of perceptual robustness and security is a kind of one-way mapping from multimedia data set to perceptual abstract set. It provides secure and reliable technique supports for the content identification, retrieval and authentication of multimedia information [1-4].

The speech perceptual hashing authentication algorithm becomes transparent under the principle of Kerckhoffs, it becomes easy for an attacker to tamper and delete the speech information without knowing by the receiver. So, it is necessary to analyze the security of speech perceptual hashing authentication technology [5]. At present, the security analysis of perceptual hashing mainly includes the following two types.

The first type: Directly encrypt the perceptual hashing algorithm. This method is mainly to encrypt one specific link of perceptual hashing algorithm and the security of the algorithm mainly depends on the unknowing of secret key [6]. This kind of method is the simplest and most effective way to protect the security of perceptual hashing algorithm under the principle of Kerckhoffs. But this kind of direct encryption method is not absolutely safe, because an attacker can analyze the relationship of original information and the encrypted perceptual hash value according to the intercepted information and then the specific value of secret key can be calculated without error. Liu *et al*. [7] proposed a method to protect the security of perceptual hashing algorithm based on fuzzy

ISSN: 1738-9976 IJSIA

commitment scheme. The realization of security of this method is not depending on the direct encryption to the perceptual hash value, but its function is same to the direct encryption to the perceptual hash value. This method avoids the probability that an attacker can make the algorithm unsafe through the analysis of the intercepted information. But the proposed algorithm is not based on speech and the requirement of real-time is not high.

The second type: Comprehensive analysis of perceptual hashing algorithm. This kind of analysis including:

1) The analysis based on diffusion and confusion. For example, Coskun and Memon [8] analyzed the diffusion and confusion properties of perceptual hashing and pointed out that the perceptual hashing algorithm should meet the diffusion and confusion properties at the same time to guarantee its security, or the perceptual information can be tampered by the attacker without being found. But it doesn't propose effective improvement measures to guarantee the security of the algorithm according to the above situations.

2) The safety analysis method based on information theory. This method analyzes the safety of the perceptual hashing based on entropy principle in information theory. For example, in both [5] and [9] the authors analyzed the security of perceptual hashing based on the information theory. But these two methods are only about the theoretical analysis, not specific experiment method is given to verify its correctness. Mao and Wu [10] analyzed the security of perceptual hashing according to the concept of unicity distance in information theory. It points out that if the times that reusing the same secret key to encrypt reaches to a limit the secret key may be estimated without error. But the proposed analysis method is targeted at a specific encryption method and doesn't have versatility.

3) The safety analysis methods based on related concepts. Hu *et al*. [11] proposed a kind of robust perceptual hashing security architecture against disadvantages that traditional perceptual hashing algorithm suffered. The security protocol designed according to the architecture above further guaranteed that an attacker can't get the plaintext/perceptual hash value pairs or secret key/perceptual hash value pairs at the same time. But this architecture ignores the attacker's attack on algorithm in the process of transmission. Zhou and Au [12] proposed a method to estimate the secret key. The proposed method uses the embedded virtual watermark as equivalent keys to ensure the safety of the algorithm, but its security analysis is proposed based on specific perceptual hashing algorithm and has no improvements about security after the secret key is estimated.

The fuzzy commitment scheme has been widely used in the field of biometric authentication and it shows good performance in guaranteeing the security of the authentication information. For example, Kelkboom *et al*. [13] firstly described a kind of decodability attack based on cross-matching in biometric systems, then the authors proved that applying a random bit-permutation process can secure the fuzzy commitment scheme from cross-matching based on the decodability attack. Kelkboom *et al*. [14] analyzed the analytical relationship between classification performance and the key size based on the fuzzy commitment scheme, then got the relationship between the best classification performance and the maximum key size from the theoretical and experimental analysis. The security of the algorithm can be effectively guaranteed by adjusting the relationship above. Hidano *et al*. [15] proposed a biometric cryptosystem based on fuzzy commitment scheme. Thang Hoang *et al*. [16] proposed a kind of biometric authentication scheme by detecting the gait information. The above fuzzy commitment algorithms are structured based on classic coding and classical password scheme and these algorithms are received by using the classic algorithm transform and processing on the biometric template. But these algorithms don't give corresponding analysis and evaluation about the security on post-quantum cryptography age. Cao and Song [17] proposed a kind of quantum fuzzy commitment scheme based on entanglement auxiliary code aiming at the insecurity of

traditional fuzzy commitment scheme and proved its security through analysis. But the authors don't give specific experimental scheme to verify its safety.

In conclusion, either the first or the second perceptual hashing security analysis method has its own theoretical foundation and basis, no any kind of analysis method is perfect. But trying to fuse the first and the second analysis method is possible. So, this paper proposes a kind of speech perceptual hashing safety authentication method based on the fuzzy commitment scheme according to the feature of speech perceptual hashing and the basic theory of fuzzy commitment scheme on [18] and analyzed its security. This method is not to directly encrypt the perceptual hashing algorithm, but its function is equivalent to encryption. This article tries to analyze its theory in order to achieve the fusion of encryption and theory analysis and better guarantee the safety of speech perceptual hashing authentication algorithm.

The rest of this paper is organized as follows. Section 2 describes the basic theory of fuzzy commitment scheme. Section 3 describes the main process of authentication. The detailed process of FCS-based authentication and the performance analyses are shown in Section 4. Finally, we conclude our paper in Section 5.

2.The Related Theory of Fuzzy Commitment Scheme

The fuzzy commitment scheme (FCS) is a kind of cryptographic primitives proposed by Juels and Wattenberg [19]. It is used for the safety authentication of biological systems. The fuzzy commitment scheme is a new method for safety authentication by combining cryptographic hash function and error correcting code. It mainly consists of enrolment process and verification process. As shown in Figure 1.

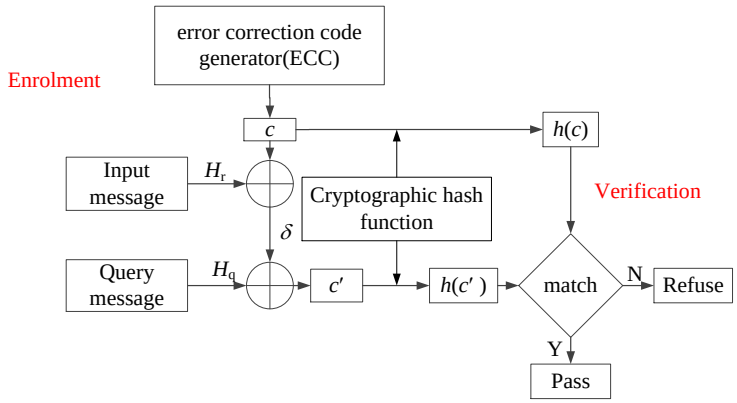


Figure 1. The Components of FCS

1) Enrolment process

(1) In enrolment process stage, the information that the user entered is summary information H_r. The information is binary bit string that the user extracted from the original information. In the field of biometric authentication the bit string is called biological template, in the field of perceptual hashing the bit string is called perceptual hash value.

(2) The error correction code generator (ECC) randomly generates an error correction code word c after the input information is entered. The error correction code word can be expressed as (n, k, t) form which the n represents the length of error correction code word, k represents the digits of information bit, t represents the mistake bit that the error correction code can corrected. The code length of error correction code word should be in accordance with the binary bit string that the user entered. A new binary bit string δ which called a commitment can be obtained after the XOR operation is operated on error

correction code word c and input information H_r. In the field of biometric authentication, the commitment δ is stored in the database, in the field of perceptual hashing, the commitment δ is sent to the receiver with the original information.

(3) The $h(c)$ can be generated by using the cryptographic hash function to encrypt the error correction code word c that randomly generated by the error correction code generator. In the field of biometric authentication, $h(c)$ is stored in the database, in the field of perceptual hashing, $h(c)$ is sent to the receiver with the original information for real time authentication.

2) Verification Process

(1) In the verification process stage, the information that the user entered is called query information H_q. Its processing method is same with the stage of enrolment, the received query information is binary bit string too.

(2) After the query information entered, the operated information c' can be received after the XOR operation is operated on the query information H_q and the commitment δ that generated in the enrolment stage. In theory, the c' should be same with c, while the H_r and H_q may be just similar information on perceptual content, may lead to c' different from c. But in the fuzzy commitment scheme, c is an error correction code word, itself has the ability to correct the error code element. So, as long as the wrong digits of error correction code are less than t, the error correction code can correct the error code word to correct code word.

(3) The $h(c')$ can be generated by using the cryptographic hash function that same with the enrolment stage to encrypt the calculated c'.

(4) Finally, the authentication information $h(c)$ and $h(c')$ that the receiver received will be compared to judge that whether the authentication information has been maliciously tampered.

3. The Authentication Process of Speech Perceptual Hashing based on FCS

This paper applies the fuzzy commitment scheme to the speech perceptual hashing authentication system to protect the security of perceptual hashing algorithm based on discussions above. The FCS consists of *Cmt*() function and *Decmt*() function. The definitions are as follows.

Definition 1:

$$Cmt(H_r, c) = (h(c), H_r \oplus c) = (h(c), \delta) \quad (1)$$

In (1), H_r is perceptual hash value of input information, c is a randomly selected (n, k, t) error correction code word, h is the cryptographic hash function, δ is the calculated value of XOR operation on H_r and c, it is the commitment talked above.

Definition 2:

$$Decmt(H_q, \delta) = h(Dec(H_q \oplus \delta)) = h(c') \quad (2)$$

In (2), H_q is the perceptual hash value of query information, δ is the calculated value of Definition 1, *Dec*() represents error correction code decoding function, c' is the decoding value.

Based on definitions above, the flow chart of proposed speech perceptual hashing authentication process based on fuzzy commitment scheme in this paper is shown in Figure 2.

As can be seen from Figure 2, speech perceptual hashing authentication process based on FCS is mainly divided into the following three processes.

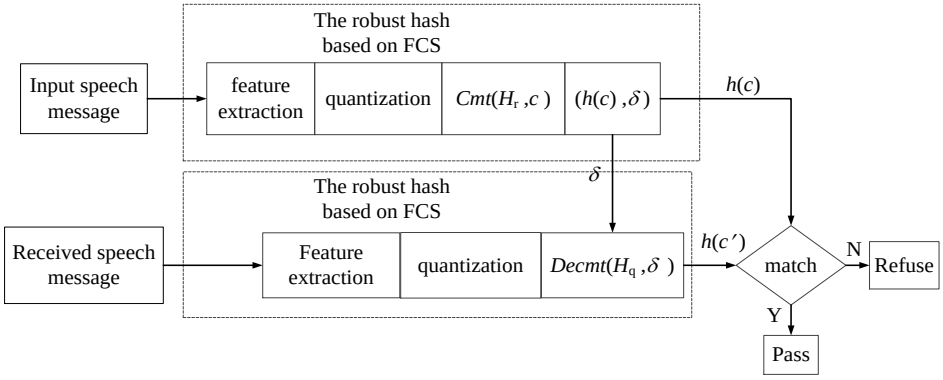


Figure 2. Speech Perceptual Hashing Authentication Process Based on FCS

1) Assume the input speech information is I_r, the perceptual hash value that extracted from input information is H_r. At this moment, H_r is not directly sent to the receiver, nor directly encrypted before sent to receiver, but sent to the *Cmt*() function operation. From the definition of *Cmt*() function we can know that δ is generated through XOR operation on perceptual hash value H_r and the randomly generated error correction code word c and then the $h(c)$ can be generated using the cryptographic hash function to encrypt c. As a result, the actual transmission information in channel is $(h(c),\delta)$. The $(h(c),\delta)$ that go through *Cmt*() function operation is called secure perceptual hash value, because an attacker is unable to get the plaintext/perceptual hash value pair, that is (I_r, H_r).

2) The speech information that received in the receiver is called I_q, using the feature extraction process that same to the sender to extract the speech perceptual hash value which called H_q. The final result $h(c')$ can be generated through the *Decmt*() function operation on H_q that calculated in the receiver and δthat the sender sends to the receiver.

3) During the matching stage, the data that used to match is no longer the original perceptual hash value, nor the encrypted perceptual hash value, but $h(c)$ and $h(c')$ for comparison. This is because in receiver the error correction code decoding operation on the calculated error correction code word c' can effectively correct the limited error during the transmission process, so the content preserved speech information can be authenticated maximumly.

4. The Description of FCS-Based Security Authentication Scheme of Speech Perceptual Hashing

The speech perceptual hashing security authentication scheme based on FCS can be realized by simply modifying the algorithm in [18] and the feasibility of process in Section 3 can be proved through experiment. The flow chart of proposed algorithm in [18] is shown in Figure 3.

There are some points needed to pay attention when upgrading the original perceptual hashing algorithm to FCS-based perceptual hashing algorithm.

1) The selection of error correction code is random, but the type of code is not randomly selected. This paper selects the BCH code as the error correction code word. A BCH code word can be expressed as the form of (n, k, t), where n represents the length of code word, k represents the digits of information bit, t represents the digits of error elements that the BCH code can corrected. In order to ensure the algorithm has security, the value of n, k, t is not fixed, but changed timely according to different algorithms.

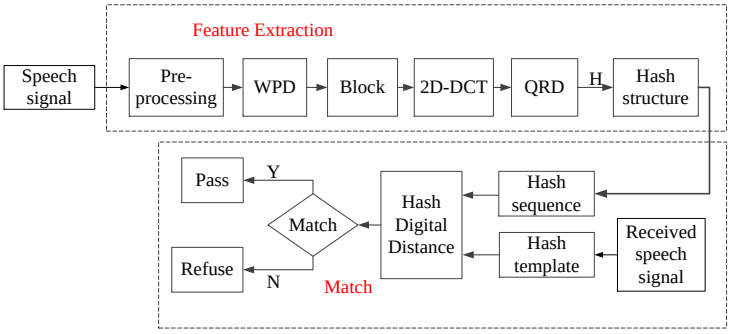


Figure 3. The Flow Chart of Speech Perceptual Hashing Authentication in [18]

2) The choice of error correction code encryption algorithm, namely the choice of *h*. The purpose of this study is to guarantee the security of speech perceptual hashing algorithm in mobile computing environment, so the requirements of real-time and resource limitations are high. This directly limits the selection of the encryption algorithm. This article selects the most simple cryptographic hash function encryption algorithm to encrypt the information. This is because the summary information that has been encrypted by the cryptographic hash function has one-way property, it is very difficulty for an attacker to figure out the original information through the summary information.

3) The most important point on the premise of guaranteeing the safety of perceptual hashing algorithm and that often been overlooked is the performance can't be affected after the original perceptual hashing algorithm is upgraded to the more secure hash algorithm.

4.1. The FCS-Based Speech Perceptual Ashing Algorithm

The flow chart of speech perceptual hashing security authentication algorithm based on FCS is shown in Figure 4.

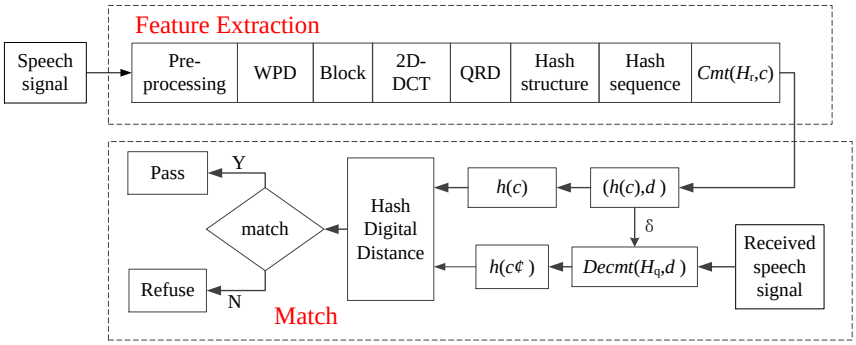


Figure 4. The Flow Chart of Speech Perceptual Hashing Security Authentication Algorithm Based on FCS

In combination with FCS, the specific procedure of improved speech perceptual hashing algorithm is as follows.

Step 1: Pre-processing

The signal I' is obtained after the inputing speech signal I undergone the pre-emphasis, making the high frequency part of useful signal get promoted for the follow-up feature extraction. Set the length of speech signal I is M.

Step 2: Wavelet Packet Decomposition(WPD)

The wavelet packet coefficients matrix is obtained by 3 layers global WPD on signal I', as follows.

$$W_{M\times n}=\begin{bmatrix} s_1^{u_0} & s_1^{u_1} & \cdots & s_1^{u_n} \\ s_2^{u_0} & s_2^{u_1} & \cdots & s_2^{u_n} \\ \vdots & \vdots & \ddots & \vdots \\ s_M^{u_0} & s_M^{u_1} & \cdots & s_M^{u_n} \end{bmatrix} \tag{3}$$

where $S=\{S_j|\ j=1,\ 2,\cdots,\ M\ \}$, M is the length of speech signal I, and n is data width of WPD. In consideration of the information capacity finiteness of a single speech segment, WPD layer-number in [18] is set to 3, and $n=2^3$.

Step 3: Block

The matrix $W_{M\times n}$ is split into N equal and non-overlapping block around rows to generate sub-matrix group $\boldsymbol{W}=\{\ W_i^{m\times n}\ |i=1,2,\cdots,N,\ m=M/N\ \}$, the process of which is as follows.

$$W=\begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_N \end{bmatrix},\quad W_i=\begin{bmatrix} s_{(i-1)\times N+1}^{u_0} & s_{(i-1)\times N+1}^{u_1} & \cdots & s_{(i-1)\times N+1}^{u_n} \\ s_{(i-1)\times N+2}^{u_0} & s_{(i-1)\times N+2}^{u_1} & \cdots & s_{(i-1)\times N+2}^{u_n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{i\times N}^{u_0} & s_{i\times N}^{u_1} & .. & s_{i\times N}^{u_n} \end{bmatrix} \tag{4}$$

where N is the total number of block matrix, in accordance with frame length. This algorithm set m=256, corresponding to the speech signal frame length is 32 ms, so the speech signal is short-time stability.

Step 4: 2D-DCT

Each sub-matrix W_i is transformed by a two-dimensional DCT, and we can obtain the new sub-matrix group $\boldsymbol{D}=\{\ D_i^{m\times n}\ |i=1,\ 2,\ \cdots,\ N,\ m=M/N\ \}$, the process is as follows.

$$D=\begin{bmatrix} D_1 \\ D_2 \\ \vdots \\ D_N \end{bmatrix},\qquad D_i=\begin{bmatrix} \theta_{1,1}^i & \theta_{1,2}^i & \cdots & \theta_{1,n}^i \\ \theta_{2,1}^i & \theta_{2,2}^i & \cdots & \theta_{2,n}^i \\ \cdots & \cdots & \ddots & \cdots \\ \theta_{m,1}^i & \theta_{m,2}^i & \cdots & \theta_{m,n}^i \end{bmatrix} \tag{5}$$

Step 5: Quadrature Rectangle Decomposition(QRD)

Each new sub-matrix D_i is transformed by QRD using Givens Rotation to generate the matrix group $\boldsymbol{R}=\{R_i^{m\times n}|i=1,\ 2,\ \cdots,\ N,\ m=M/N\ \}$, which is computed as follows.

$$R=\begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_N \end{bmatrix},\qquad R_i=Q_i^{-1}D_i \tag{6}$$

where Q_i^{-1} is the inverse of matrix Q_i, is a real $m\times m$ orthogonal matrix, and the standard deviation of each matrix R_i is computed to obtain the feature parameter sequence $H(N,1)$ as follows.

$$H=\begin{bmatrix} std_1 \\ std_2 \\ \vdots \\ std_N \end{bmatrix}, \qquad std_i=\sqrt{\frac{1}{m\times n}\sum_{j_1=1}^{m}\sum_{j_2=1}^{n}(R_i(j_1,j_2)-\mu_i)^2} \quad j_1, j_2 \in N^+$$

(7)

where $\mu_i=\frac{1}{m\times n}\sum_{j_1=1}^{m}\sum_{j_2=1}^{n}R_i(j_1,j_2)$.

Step 6: Hash structure

The final perceptual hash value is expressed as $\boldsymbol{ph}=\{ph(i)|i=1, 2, \cdots, N\}$, it is decided by the symbol of hash sequence $\boldsymbol{H}=\{H(i)|i=1, 2, \cdots, N\}$, the hash structure method as follows.

$$ph(i)=\begin{cases} 1 & if \quad H(i)>H(i-1) \\ 0 & else \end{cases} \tag{8}$$

Step 7: Secure Perceptual Hash Value

The secure perceptual hash value δ and the encrypted information h(c) used for authentication can be generated after the *Cmt*() operation is operated on error correction code word c and perceptual hash value, where the perceptual hash value is generated above and the error correction code word c is randomly generated using the error correction code generator.

Step 8: Matching and Detection

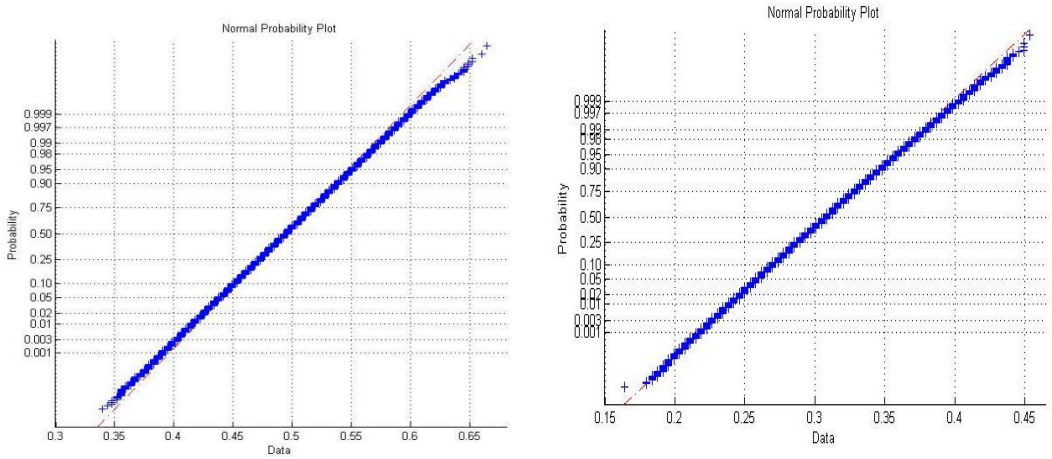
In receiver, the perceptual hash value should be recalculated from the received speech information. Then the authentication information $h(c')$ can be obtained via the *Decmt*() operation on the calculated perceptual hash value and the safety perceptual hash value that the receiver received. The Hamming distance of $h(c)$ and $h(c')$ is calculated to judge whether it can pass the matching. The detailed calculation process can be found in [18].

In this improvement method we can see that the algorithm of the original scheme doesn't be modified, there are just several unimportant links are applied. In sender, δ and $h(c)$ are generated independently, they are separated with the algorithm. In the receiver, the generation of c' and $h(c')$ is also simple operations and don't affect the main algorithm proposed in [18]. So, these operations proposed above don't occupy too many resources.

4.2. The Performance Analysis and Comparison of Algorithm

In [18], the authors use the Bit Error Rate (BER) normal distribution curve and the False Accept Rate - False Reject Rate (FAR-FRR) curve to measure the performance of speech perceptual hashing authentication algorithm. This paper continues to use these two performance indexes to test if the algorithm performance is affected after the original speech perceptual hashing algorithm is upgraded to the FCS-based algorithm. The BER comparison of speech signal before and after encryption is shown in Figure 5.

The experiment environment and the speech signal data sample size are same to Ref. [18], so that the experiment result can be compared with Ref. [18].



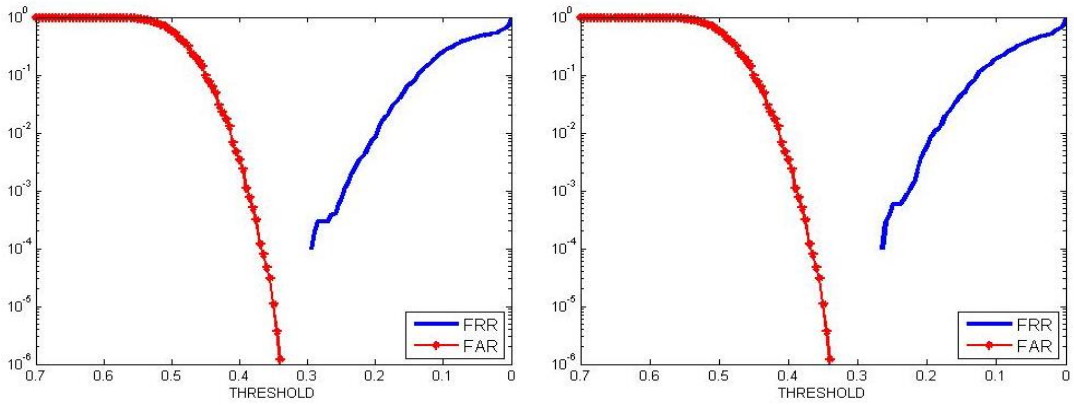
(a) The BER of Different Speech Signal before Encryption

(b) The BER of Different Speech Signal after Encryption

Figure 5. The BER Comparison of Speech Signal before and after Encryption

It can be seen that the horizontal axis of BER normal probability plot of different speech becomes different before and after encryption from the comparison results of Figure 5. This is because the BER of before encryption is obtained by comparing the original perceptual hash value, while the BER of after encryption is obtained by comparing the safety perceptual hash value. Both comparative data is no longer the same, and thus lead to the difference of horizontal axis. But this doesn't affect the performance of perceptual hashing algorithm. Because it can be seen from the encrypted BER normal probability plot that the BER of different speech after encryption still obey the normal distribution, and its effect is better than before encryption. This is because the FCS-based perceptual hashing algorithm used the error correction code, making the transmission errors can be corrected in receiver. This makes the normal probability plot better that before encryption.

The FAR-FRR curve expresses the relationship between robustness and distinction of perceptual hashing algorithm. We said that the algorithm has good robustness and distinction when the curve has no intersection. That is to say the algorithm has good robustness for content preserving operations and has good distinction for malicious tampering operations. The performance comparison of FAR-FRR before and after encryption is shown in Figure 6.



(a) FAR-FRR Curve before Encryption

(b) FAR-FRR Curve after Encryption

Figure 6. The Performance Comparison of FAR-FRR before and after Encryption

It can be seen from the comparisons of Figure 5 and Figure 6 that the performance of speech perceptual hashing algorithm doesn't suffer any affect after the original speech perceptual hashing algorithm is upgraded to FCS-based algorithm. This is because when the correlation of final calculated safety perceptual hash value and the input information is small, the ability of week collision resistance is strong. At this moment not only the performance of perceptual hash won't be affected, and its security is also got a lot of ascension. This shows that it is feasible in this paper that the fuzzy commitment scheme is applied to the speech perceptual hashing authentication.

4.3. The Security Analysis of the Algorithm

Firstly, we analyze some attack operations that the attacker makes in speech perceptual hashing authentication system. This paper shows the possible attacks that the attacker may make in speech perceptual hashing authentication system in Figure 7.

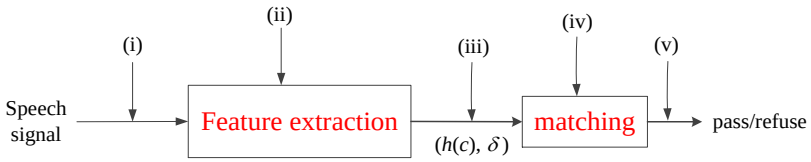


Figure 7. The Diagram of Attacks That the Speech Perceptual Hashing Authentication System May Suffer

It can be concluded from Figure 7, that the attacks the speech perceptual hashing authentication system may suffer from are mainly divided into two categories: The first category is attacks occurred in the channel, such as (i), (iii) and (v); The second category appears to be in the process chain, for example (ii) and (iv).

Under normal circumstances, the attacks occurred in channel mainly rely on direct encryption processing or similar encryption processing solution, such as digital watermarking, biometric fingerprint, etc. Attacks occurred in processing link are mainly through the access control or chip design to ensure the safety.

Based on the proposed FCS-based speech perceptual hashing authentication scheme, this paper analyzed several common attacks and indices that affect the safety, discussed how the proposed authentication scheme to guarantee the safety of speech perceptual hashing in detail.

1) The promotion of low diffusion performance. In [8] pointed out that the security of perceptual hashing relies on the implementation of confusion and diffusion performance at the same time. In original perceptual hashing system the confusion can be realized via encryption process, but the diffusion properties are not guaranteed. This allows an attacker might intercept a group of (I_i, H_i) in the process of transmission, after the analysis on (I_i, H_i) the algorithm may suffer the collision attack from the attacker so that the speech information can't be normally authenticated. But in the proposed scheme of this paper, the transmittal data is no longer the perceptual hash value H_i, but δ_i. At this moment the attacker only can intercept a group of (I_i, δ_i), where $\delta_i = H_i \oplus c_i$. Because c is randomly selected for generating code word of secure perceptual hash value, so δ can be considered as an encrypted ciphertext that using c to encrypt H. So, even if the original perceptual hashing system has a very low diffusion performance, the original information I and δ almost have no contact, thus the proposed scheme of this paper effectively overcomes the drawbacks of low diffusion.

2) Avoided the estimate of secret key. In [10] pointed out that according to the concept of Shannon unicity distance, if the times that reusing the same secret key to encrypt the plaintext reach to a certain limit then the secret key may be estimated without error. The proposed scheme of this paper effectively avoids the happening of this situation. It can be

seen easily from Figure 7. What transmitted in channel is speech information and secure perceptual hash value ($h(c)$, δ), an attacker want to obtain original perceptual hash value and error correction code word is very difficult in calculation via analyzing the secure perceptual hash value. In addition, in receiver, what used for matching is $h(c)$ and $h(c')$, no longer to match the perceptual hash value. So, an attacker can alter the information transmitted in the channel into a meaningless information making the authentication failed, but it is very difficult to alter the information into an illegal but meaningful information in calculation.

3) Avoided the vulnerability of the authentication information. This is because the authentication information transmitted in channel is cryptographic perceptual hash summary $h(c)$. The one-way property of cryptographic hash function guarantees that an attacker can't figure out the original information only through the perceptual summary, this ensures the security of authentication information to some extent.

4) Overcome the shortcomings of direct encryption scheme. Because in the matching stage of receiver the plaintext summary H_r of perceptual hash is no longer needed, meanwhile the plaintext summary of perceptual hash won't be seen in any processing link and transmitting procedure of speech authentication system. Therefore, the plaintext exposure is effectively avoided.

The above 4 points are the main aspects of the security of the proposed scheme of this paper. The scheme of this paper contributes a certain security improvement for other links that this paper doesn't mention. Such as the illegal forgery of speech information in sender and the illegal modification of match result output in receiver, and so on.

To sum up, the FCS-based speech perceptual hashing authentication scheme proposed in this paper has obvious improvement compared with before. The security analysis about FCS can refer to Ref. [20], this paper doesn't talk about it any more. The security analysis of this paper is proposed for the lack of security of existing speech perceptual hashing algorithm. Through the experiment we can see that the application of fuzzy commitment scheme has good performance in speech perceptual hashing authentication and its application in speech authentication will not affect the performance of algorithm itself.

5. Conclusions

Aiming at the speech perceptual hashing authentication algorithm lacking of security in present situation, this paper proposes a speech perceptual hashing security authentication scheme based on FCS. Through the analyses of this paper, we can conclude that as long as the system parameter is selected properly, the FCS-based perceptual hashing authentication scheme can effectively guarantee the safety of the hash algorithm without affecting the original performance of robustness and distinction. Meanwhile, it overcomes some important safety problems, such as the low diffusion performance, the estimate of secret key, and so on. In addition, the vulnerability of authentication information, the direct encryption scheme and other security problems that met in the speech perceptual hashing authentication process can also be overcome and improved the flexibility of the authentication process.

Further research will focus on the optimization of FCS-based speech perceptual hashing authentication scheme, and ultimately hoping to find a general safety analysis method.

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] M. Nouri, N. Farhangian and Z. Zeinolabedini, "Conceptual authentication speech hashing base upon hypotrochoid graph", Proceedings of the 6th IEEE International Conference on Symposium Telecommunications (IST), Tehran, Iran, **(2012)**, pp.1136-1141.
[2] N. Chen and H.D. Xiao, "Perceptual audio hashing algorithm based on Zernike moment and maximum-likelihood watermark detection", Digital Signal Processing, vol.23, no.4, **(2013)**, pp. 1216-1227.
[3] Q.Y. Zhang, Y.W. Liu, Y.B. Huang, P.F. Xing and Z.P. Yang, "Perceptual hashing algorithm for speech content identification based on spectrum entropy in compressed domain", International Journal on Smart Sensing and Intelligent Systems, vol.7, no.4, **(2014)**, pp.283-300.
[4] Y.B. Huang, Q.Y. Zhang and Z.T. 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] O. 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, **(2008),** pp. 681906-681906-10.
[6] 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), Kaohsiung, Taiwan, vol.3, **(2008),** pp.29-32.
[7] Q. 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.
[8] B. Coskun and N. Memon, "Confusion/diffusion capabilities of some robust hash functions", Information Sciences and Systems, 2006 40th Annual Conference on. IEEE, Princeton, NJ, **(2006),** pp.1188-1193.
[9] 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),** pp. 72540H-72540H-13.
[10] 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.
[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), Leshan, China, **(2013),** 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), Barcelona, Spain, **(2011),** pp.1-6.
[13]] E.J.C. Kelkboom, J. Breebaart, T.A.M. Kevenaar, I. Buhan and R.N.J. Veldhuis, "Preventing the decodability attack based cross-matching in a fuzzy commitment scheme", Information Forensics and Security, IEEE Transactions on, vol.6, no.1, **(2011)**, pp.107-121.
[14] E.J.C. Kelkboom, J. Breebaart, I. Buhan and T.A.M. Kevenaar, "Maximum key size and classification performance of fuzzy commitment for gaussian modeled biometric sources", Information Forensics and Security, IEEE Transactions on, vol.7, no.4, **(2012)**, pp.1225-1241.
[15] S. Hidano, T. Ohki and K. Takahashi, "Evaluation of Security for Biometric Guessing Attacks in Biometric Cryptosystem using Fuzzy Commitment Scheme", Proceedings of the International Conference of the Biometrics Compendium, IEEE, Darmstadt, Germany, **(2012),** pp.1-6.
[16] Thang Hoang, Deokjai Choi and Thuc Nguyen, "Gait authentication on mobile phone using biometric cryptosystem and fuzzy commitment scheme", International Journal of Information Security, Published online: 30 January **(2015)**.
[17] D. Cao and Y.L. Song, "The quantum fuzzy commitment and biometric authentication based on entanglement auxiliary code", Acta Electronica Sinica, vol.40, no.7, **(2012)**, pp.1492-1496.
[18] 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.
[19] A. Juels and M. Wattenberg, "A fuzzy commitment scheme", Proceedings of the 6th ACM conference on Computer and communications security. ACM, **(1999),** pp.28-36.
[20] Y. Dodis, L. Reyzin and A. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data", Advances in cryptology-Eurocrypt 2004, Springer Berlin Heidelberg, Heidelberg, Berlin, **(2004),** pp.523-540.

