2020 IEEE/ACM Symposium on Edge Computing (SEC) | 978-1-7281-5943-0/20/$31.00 ©2020 IEEE | DOI: 10.1109/SEC50012.2020.00049

# TAES: Two-factor Authentication with End-to-End Security against VoIP Phishing

Dai Hou*, Hao Han*, Ed Novak†
*College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, China
†The Department of Computer Science, Franklin and Marshall College, USA
Email: houdai@nuaa.edu.cn, hhan@nuaa.edu.cn, ed.novak@fandm.edu

***Abstract*—In the current state of communication technology, the abuse of VoIP has led to the emergence of telecommunications fraud. We urgently need an end-to-end identity authentication mechanism to verify the identity of the caller. This paper proposes an end-to-end, dual identity authentication mechanism to solve the problem of telecommunications fraud. Our first technique is to use the Hermes algorithm of data transmission technology on an unknown voice channel to transmit the certificate, thereby authenticating the caller's phone number. Our second technique uses voice-print recognition technology and a Gaussian mixture model (a general background probabilistic model) to establish a model of the speaker to verify the caller's voice to ensure the speaker's identity. Our solution is implemented on the Android platform, and simultaneously tests and evaluates transmission efficiency and speaker recognition. Experiments conducted on Android phones show that the error rate of the voice channel transmission signature certificate is within 3.247 %, and the certificate signature verification mechanism is feasible. The accuracy of the voice-print recognition is 72%, making it effective as a reference for identity authentication.**

***Index Terms*—Identity authentication, Voiceprint recognition, voice-print, voice, MFCC, GMM-UBM**

## I. Introduction

Voice over Internet Protocol (VoIP), commonly known as VoIP, IP telephony and so on, is a technology allowing live, interactive audio conversations between two or more parties. In recent years, due to the development and popularization of VoIP technology, various telecommunications fraud incidents have occurred with frequency all over the world. This has brought huge losses to people of all countries. China is one of the worst-hit areas of telecommunications fraud. At present, for most telecommunications fraud cases in China, the technical means used by the suspects are realized based on VoIP technology. By modifying and forging the calling number, the suspects claim to be relevant administrative personnel, relevant customer service personnel, or simply pretend to be a person who knows the victim to carry out the fraud.

A VoIP server is convenient to set up, with very low cost. Moreover, it is often set up overseas to better evade detection. This brings difficulties for investigation and evidence collection. The traditional way to deal with fraudulent calls is to analyze the number of the attacker/caller, combine the prefix, code length and standardization with complaint data to obtain the fraudulent phone number, and output it to a public blacklist database [1]. However, because VoIP technology can arbitrarily modify the calling number, or even fake the number of normal users, the effectiveness of this traditional method is very limited.

On the Web, SSL/TLS is used to protect data integrity and provide authentication, but modern telephony infrastructures do not perform any authentication, especially for other telephony network access information. The third generation (3G) and the fourth generation (4G) cellular networks do realize the mutual authentication between the user and telecommunication providers. But, these mechanisms are designed to ease the operator's task of accurate billing. Helping users identify other users is not a goal, and therefore 3G and 4G cannot provide authentication *between users*. The existing telecommunications networks lack any robust security authentication mechanism, which has led to common exploitation: VoIP fraud.

There is a method of short message service (SMS) verification [2], but this method cannot be applied to terminals without short message service functions such as landlines. In addition, the domestic operators also activate the SMS reminder function for overseas calls to remind users to prevent telecom fraud, but this method is only for overseas calls, and the remedial method, done in hindsight, is of little effectiveness in prevention.

In recent years, some research put forward the ”end-to-end call security authentication protocol”, but the agreement can only solve the problem in an equipment-centric way. It cannot be used for safe user authentication. For example: when the user's mobile phone is stolen or lost, or the user's previous mobile phone number is used by a new user. In these cases the suspect/attacker can directly use the number on the original user identity. Therefore, we consider voice-print, as a valid biometric. Voice-print is gradually becoming widely used in the field of identity authentication. Over a talk channel, the other person's voice can also be used as a powerful authentication feature. We imagine that voice-print authentication technology can be applied to achieve identity authentication of the voice channel as a supplement to the device authentication. The current state of end-to-end authentication mechanisms, such as Reaves's and Blue's work AuthentiCall[4], has the limitation that audio digests cannot detect altered audio less than one second in length.

In summary, this paper presents the following contributions:

- We consider the complex problem of preventing telecommunication fraud with the goal of ensuring that the calling number was not forged and that the speaker is the owner

978-1-7281-5943-0/20/$31.00 ©2020 IEEE
DOI 10.1109/SEC50012.2020.00049

of the number (i.e., not an insider who can access a business number or a fraudster using a stolen phone and number).
- We propose a two-factor authentication scheme to defend against VoIP phishing and fraud. Coupled with a modern "data over voice" channel, we add the voice-print as a factor into the certificate, which can verify the true identity of the speaker.
- We implement TAES on Android smartphones, which encompasses these two techniques. We also evaluate the system in real world scenarios. The experimental results confirm the performance of our approach in authenticating both the calling number and the caller simultaneously.

## II. Related Work

### A. Terminal Equipment Certification

At present, there are many encryption mechanism for end-to-end authentication mechanisms, such as Silent Phone [5]. Silent Phone is an Android APP, providing end-to-end encryption of the voice channel. It uses the general number to make and answer phone calls, but if the other party is also equipped with the program, calls can be upgraded to encrypted calls. Actually, the method adopted by these systems still depends on the data networ, which has a high demand for data bandwidth, and cannot be extended to the global telephone network, which is heterogeneous.

Tu and his team [6] describe how to modify the core telephony signaling system, SS7, to support the authentication of the caller, but this protocol is not end-to-end. At the same time, it requires both sides to call from the same SS7 session network. Above all, this method still needs to modify the core network entity of each network, which is complicated and difficult to implement.

Hossen Mustafa et al. [2] proposed a short message service (SMS) to detect a spoofed caller ID. A challenge is sent to the caller ID through a short message, which is detected by the caller side and automatically responds to the caller to prove the caller ID authenticity. However, this detection mechanism cannot be applied to terminals such as landlines that do not have SMS service.

Literature [7] presents an approach called the "AuthLoop" system, which can provide authentication on the voice channel. The system is mainly divided into two parts. Firstly, a modulator and demodulation scheme and a support link layer protocol are designed to solve the problem of reliable transmission over the voice channel. Then, a security model and protocol are designed to verify the number of the caller and realize end-to-end authentication. Similar to AuthLoop, literature [4] proposes AuthentiCall authentication system, set up a third-party server, and designed a registration, authentication and integrity call three agreement, through the third party server, end-to-end authentication can be completed before the call is answered, which takes less time than the AuthLoop, and the call is encrypted to ensure the content integrity of the call.

### B. Voice-print Recognition

Voice print recognition is also one of the common methods to prevent telecommunications fraud. Similar to fingerprints, different people's voices have their own unique characteristics. By analyzing the sound characteristics, people are identified. This process is called voice-print recognition.

Voice-print recognition is mainly divided into two parts. First, the feature extraction process and second, the pattern matching process. For feature parameters, the following two parameters are usually adopted to represent the features of a voice-print: linear prediction cepstrum coefficient LPCC[8] and MFCC[9]. The linear prediction cepstrum coefficient has a good effect on eliminating some excitation information generated in the speech process. However, MFCC[9], the most widely used feature parameter with good simultaneous effect, is more in line with the auditory characteristics of human ears. And, it performs better in recognition performance and noise robustness. For pattern matching, common techniques mainly include: Dynamic Time Warping (DTW) and Hidden Morkov Models (HMM) [11].

In 2000, the Gaussian Mixture Model - Universal Background Model (GMM-UBM[13]) was put forward on the basis of the Gaussian Mixture Model GMM[14]. It has made an important contribution to the practice of speaker recognition. On top of the traditional GMM-UBM method, JRA[15] and I-vector Model [16] have been put forward in combination with factor analysis, which can extract characteristic information related to the speaker. The characteristics related to the channel are removed, the channel influence is well overcome, and the system performance is improved [17].

## III. System Design

### A. The General Design of TAES

TAES is mainly divided into two parts. The overall structure of the system is shown in Fig. 1.

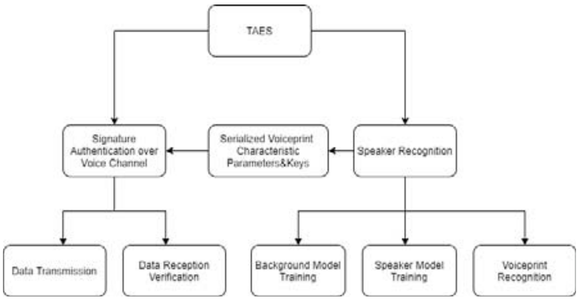
TAES
Signature Authentication over Voice Channel
Serialized Voiceprint Characteristic Parameters&Keys
Speaker Recognition
Data Transmission
Data Reception Verification
Background Model Training
Speaker Model Training
Voiceprint Recognition

Fig. 1. General System Design

TAES is aimed at caller identity authentication. The first authentication mechanism is voice channel signature authentication, which verifies the authenticity of the caller's number to make sure that the caller is the owner of the number. The second authentication mechanism is speaker identification, which verifies the caller's identity. We serialize the characteristic parameters of the speaker's voice-print and add them to a digital certificate as a factor. At the same time, we extract the key from the speaker's voice-print characteristic parameters to encrypt the certificate.

### B. Transmission over Voice Channel

In order to authenticate the identity during the call, we need to transmit data through the voice channel, which needs to encode and modulate the data. The process of the voice channel data transmission is shown in Figure. 2.

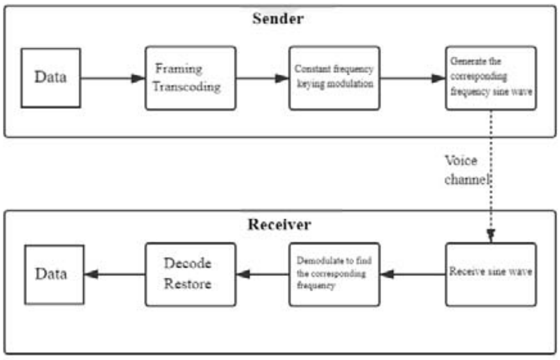
Sender
Data
Framing Transcoding
Constant frequency keying modulation
Generate the corresponding frequency sine wave
Voice channel
Receiver
Data
Decode Restore
Demodulate to find the corresponding frequency
Receive sine wave

Fig. 2. Voice channel transmission system

Among the transmitted data is the certificate that can prove the identity of the owner of the number. The transmission starts after the call is connected, and the received certificate is verified at the receiving end. If the verification is successful, the number is real; otherwise, it is a fake/forged number modified by the attacker using VoIP and other technology vulnerabilities.

### C. Voice-print Recognition of Speaker

Speaker recognition is mainly a supplement to the first authentication mechanism. In this paper, the traditional GMM-UBM method is adopted.

First, we carry out some preprocessing of the input voice data for the convenience of feature extraction. Training the model is mainly divided into 1) general background model training and 2) target user model training. Above all, the background model needs to be trained well. On the basis of this background model, the speaker model of the target user can be obtained by combining the target speaker data.

The process of recognition is to test whether the features of the speech fragment match the speaker model. This score will judge whether the speech fragment belongs to the speaker according. It must surpass a previously set threshold value.

## IV. SIGNATURE AUTHENTICATION OVER VOICE CHANNEL

### A. Algorithm Design of Hermes

In this section, we illustrate the Hermes algorithm design. The caller data is transcoded by the transcoding module first, and then a binary data stream is generated, and the binary data is modulated into sound. It is sent through the telephone call to the called party. The receiver (called party) performs the reverse process to restore data, below we detail each module.

**The limit of channel** Our research is based on end-to-end communication, and in real life voice channels are composed of different networks, so we assume that any data transmitted is sent over unknown voice channels. Due to this the factors, such as Channel frequency range, Automatic Gain Control (AGC), Voice Activity Detection (VAD) and the total transmission distance, vary greatly. Therefore, the traditional audio modulation algorithm is greatly limited on voice networks, and it is difficult to apply it to the voice channel for data transmission.

Due to the above limitations, we choose The Hermes algorithm for transmitting data over the voice channel. This modulation algorithm is based on frequency shift keying FSK modulation and has a good transmission effect on unknown voice channel.

**Modulation and Transcoding** The modulation algorithm of Hermes [19] is relatively simple. Set an initial baseline frequency $f = fbase$ and a fixed $\delta$, and then read in the binary data stream. If the current bit is 0, subtract $\delta$ from $f$, i.e., $\delta = f - \delta$ . If the current bit is 1, add $\delta$, that is, $\delta = f + \delta$. The specific algorithm framework is shown in algorithm 1. The advantage of such modulation is that all jumps are bounded by $\delta$, which can prevent large jumps.

However, some special circumstances may occur. For example, when data meets a long string of 0's or 1's, the current frequency will exceed the limit of allowed range. In addition, if there is a long string of 0's, sound modulation will work at low frequency all the time, reducing the number of sinusoidal signals that can be sent per unit time, and thus reducing throughput.

In order to solve this problem, we also need to transcode the data before modulation. The transcoding algorithm is very simple. Convert '1' to '10', and convert '0' to '01.' In this way, although the amount of data may be doubled, the problem of exceeding the frequency range in the case of many 0's or 1's in succession is solved, and the modulated output voice has a fundamental frequency fixed at the base frequency, which makes the signal sound more like speech.

**Algorithm 1** Converts binary data into an audio signal and sends it over a voice channel

```
Input: binary data stream S
Output: base frequency base, delta frequency $\delta$
Initialize $f$ = $f$base
for each bit b in the input stream S  do
    if b=0 then
        $f = f – \delta$
    else
        $f = f + \delta$
    end if
    Generate a sinusoid of frequency $f$
end for
```

**Demodulation and Decoding** At the receiving end, demodulation requires the sound signal to be converted back to a binary data stream. This demodulation algorithm is also very simple. There are two variables, $f$cur and $f$prev. $f$cur represents the current received sine wave frequency, and $f$prev represents the previous sine wave frequency. $f$prev is initialized to the $f$base set at the sending end, and then judged when each frequency signal is received. When $f$cur is less than or equal to $f$prev, a 0 is decoded. If $f$cur is greater than $f$prev, a 1 is decoded. As shown in algorithm 2.

After demodulation, the task of decoding is to restore the 2n bit stream to the n bit stream of original data. However, there may be many errors and omissions. Assuming that there are no errors, the task of decoding is simply to convert ‘10’ to ‘1’, and ‘01’ to ‘0.’

However, in practice transmission errors will certainly occur. Specific errors can be divided into three types: bit-flip, bit-insert, and bit-delete. Assuming that the alignment is normal, if the data stream after demodulation is taken two digits at a time, the case with no error should be 10 or 01. When there is any error, there will be a situation of 00 or 11, which we resolve in our discussed of the classification.

We define the function ErrorMetric(i), where i is defined as the current location of the error. ErrorMetric(i) is a forward lookup function used to determine whether the alignment point is at location i. Assuming the large forward lookup digit M, this function reads from i to i+M in 2-bit blocks and returns the number of blocks that cannot be transcoded correctly.

**Algorithm 2** Converts the received sound signal back to the binary data stream

```
Input: sound signal
Output: base frequency base
Initialize pre = base
for each sinusoid in the input sound signal do
    Let cur = frequency of current sinusoid
    Let pre = frequency of previous sinusoid
    if cur $<=$ pre then
        Output 0
    else
        Output 1
    end if
end for
```

In the event of incorrect decoding, the ErrorMetric(i) function is called to judge the current bit i and the next bit i+1. If the current bit and the number of subsequent error blocks is less than or equal to i+1, then there is no problem with alignment and the error code is due to a bit-flip. If not, an alignment error is present and i is pushed back one bit. It means to be determined When X is in a particular position. Traditional error correction coding can be used for error correction.

For example, when the input is 011010100110, the first digit $m_0$ is 0, i=0 is substituted into the function ErrorMetric(i) to determine how many error blocks there are from the first digit. In this input data string, 01,10,10,1001,10 are all data blocks with correct format, so the number of error blocks is 0 and the return value is 0. When the second $m_1$=1, i=1 is substituted into the function, the data block should then be divided into 11,01,01,01,00,11, 0. The last 0 is not calculated, the number of error blocks is three, and the return value is 3, so the probability of $m_0$ being the correct alignment bit is greater than $m_1$. The system then moves on assuming that $m_0$ is the aligned bit.

### B. Design of Signature Verification System

Authentication between entities on the Internet usually depends on the use of an encryption mechanism. So far, the SSL/TLS protocol is the most widely used for Web, E-mail, instant messaging and other applications. Although it has some vulnerabilities at the present time, it provides an base idea to solve the problem of authentication over the phone.

Our design is based on the assumption that the adversary can initiate a call from any calling device and can modify the incoming call number at will, regardless of whether the call number is real or not, and that the adversary can play audio over the calling channel and can interact with the target. These are the kinds of behaviors that most attackers can perform.

The system usage scenario we designed includes three types of participants, callers (Prover), callees (Verifier), and certificate authorities (CA). The main steps of the system are as follows:

*a) Register:* When each user registers for a mobile phone number, they need to provide personal information to prove their identity, such as number, name, ID number and so on. After the CA verifies that the information is correct, it uses it’s own private key to sign the applicant file to generate a certificate C. The certificate includes number information, Voice-print information and signature information. In real life, the CA should be assumed by the telecom operator and CA provides the signature. In this paper, the experiment is simplified and the self-signature is adopted.

*b) Transmisson:* The caller P dials the caller V’s phone, and after the call is connected, the caller P uses the Hermes algorithm to transmit the transcode-modulated certificate information to the caller V through voice information through the audio communication channel. The caller V’s terminal converts the transmitted voice signal into the original data through demodulation decoding and returns to the certificate information C.

*c) Verification:* After receiving the information C from caller P’s certificate, the caller V uses the CA certificate of public key (used in the experiment the sender since the signature and the key of the public key) to decrypt the certificate C, remove the check code certificate. The check code for the certificate is calculated. If the comparison results are the same, it is known that the certificate has not been tampered with. The number information contained in the certificate is then extracted. If it matches the apparent caller number, then it is known that the caller P’s number has not been tampered with.

This kind of signature verification method borrows from the SSL/TLS protocol, and we apply it to the call. However, due to the limitation of the data transmission bandwidth of the voice channel, the size of the certificate needs to be designed to be as small as possible in the case of security. In addition, in order to speed up the verification, it is better to cache the public key information of the CA certificate in advance at the caller side.

## V. Voiceprint recognition and authentication

### A. Introduction of Voiceprint Recognition

Voiceprint recognition, also known as speaker recognition, is the process of identifying a speaker from a given segment of speech. Speaker recognition can be divided into two tasks: speaker recognition and speaker confirmation. The former is to identify the best match with the specified speech from a group of known speech samples. The latter verifies, from a voice sample, whether someone is who they claim to be.

### B. Voiceprint Feature Extraction

In the field of speaker recognition, MFCC is a very common feature parameter, which is often used in speech processing and recognition applications. It is especially common in applications where noise cannot be avoided. The principle is to map the linear spectrum to the Meir spectrum and then convert it to the cepstrum. The conversion formula of ordinary frequency and Mel frequency is as follows. In the end the cepstrum analysis is performed to obtain the MFCC.

$$f_{mel} = 2595 * \log_{10}\left(1 + \frac{f}{700}\right) \tag{1}$$

The extraction process of MFCC characteristic parameters includes Pre-emphasis, Frame blocking, Hamming window, Fast Fourier Transform (FFT), Triangular Bandpass Filters, Discrete cosine transform(DCT), Log energy and Delta cepstrum. The speaker'S GMM is trained with the obtained MFCC parameter, so as to obtain the speaker's exclusive GMM voiceprint model.

### C. Voiceprint Recognition Based on GMM/UBM

Because the research is aimed at voice-print recognition in a phone call, which has the characteristics of text-independent, open-set training, numbers of users and so on. According to the above characteristics, the GMM/UBM model is more appropriate, which is a more traditional and mature method.

### D. Extract key from GMM

We take the average vector of the Gaussian model in GMM and segment its range. The segmentation depends on the user's different voice-print characteristics. The principle is that the user's voice-print characteristic parameters appear in each segment with a stable probability. We then map the average vector of GMM to the corresponding bit. Take a 16 segmentation for example, 16 segments corresponds to 4bits. Each of the average vectors has 20 components, so each GMM can generate a key of 80 bits in length.Then we can use the key that has voice-print characteristics to encrypt the certificate.

### E. Simulate the Signature Authentication Process

To simulate the certificate verification process on a mobile device, a few simplifications are made: The message to be sent is the phone number, the speaker's voice-print characteristic parameters are also serialized and added to the certificate. We use the SHA256 as the digest. A pair of RSA keys are built into the APP, representing the public key and private key of the CA. Use private key to sign the certificate before sending, and sign the certificate on behalf of CA; After receiving the certificate, the CA public key is used to verify the certificate. After decrypting the certificate, the value of SHA256 is first obtained for the number to determine whether it has been tampered with, and then compare the phone number as well as the voiceprint. If all the information is consistent, we can identify the speaker.

## VI. Evaluation

This section presents our experimental evaluation of TAES.

### A. Data Transmission

First, we test different fundamental frequency and delta frequency error rates, in order to find the most appropriate frequency. we found that in real experiments the daily noise is usually below 1000Hz, so the fundamental frequency delta frequency range should be restricted within [1000Hz, 3000 Hz], the experimental setup every time step length value of fundamental frequency is 100 Hz, the delta frequency step size of 50, each experiment transmission 32 bit random string, namely four strings. In each experiment we transmit 50 times. Figure 3 gives the bit error rate under different frequencies.

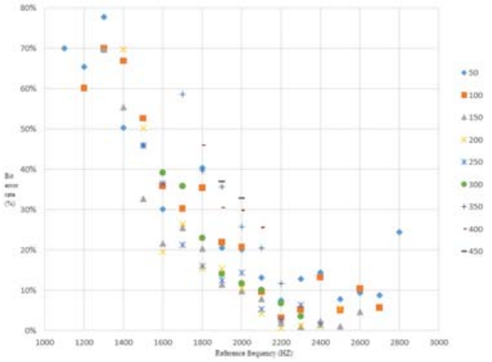

Fig. 3. Bit error rates at different base and delta frequencies

Through experiments, it can be found that when the reference frequency works at [2200Hz, 2500Hz] and delta frequency works best at [100Hz, 250Hz]. At this intersection, the transmission efficiency is the highest and the bit error rate can be kept within 3%. The best combination effect is [2300Hz,150Hz] with bit error rate of only 0.09%. Of course, due to the limitations of our experimental conditions, the bit error rate could be high. If the call permissions can be directly obtained, clearer call data flow can be obtained, and the bit error rate should be reduced to a lower level.

### *B. Speaker Recognition*

The voice data of 10 speakers were collected through mobile phone calls. For convenience, the voice fragments were divided into about 3s each. A total of 50 voice sample fragments of about 150s were used to train the speaker model with 60s of voice, and then the remaining fragments were spliced into voice fragments above 10s for testing. The speaker model was tested one at a time with all the remaining speaker fragments, including the speaker's own speech and those of other speakers as impostors. Each speaker was tested 9 times for correct speech and 81 times for false speech, each speaker was tested 90 times, for a total of 900 times.

The FRR, FAR and DET curves of different judgment thresholds are set by the test system, as shown in Figure 4. It can be seen that BER is 28%. That is, when the FRR equals to FAR, the best recognition rate of the system is 72%, and the threshold set by the system is 15 at this time.

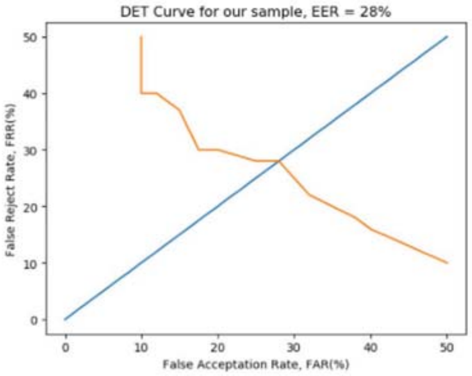

Fig. 4. DET graph of Speaker recognition system

## VII. CONCLUSION

This paper proposes TAES, which can authenticate caller phone number authenticity and caller identity on VOIP calls. TAES is implemented and tested on the Android platform.

The Hermes algorithm is adopted, which can transmit data through the voice channel during the call, making it possible to verify the authenticity of the number. Then a signature certificate authentication system is designed, which can verify the caller's certificate, so as to guarantee the authenticity of the number.

We've also designed a voice-print recognition system that authenticates the caller's voice and determines whether the current speaker is who they claim to be. After studying the current commonly used voice print recognition model methods, we decide to adopt the GMM-UBM model which is easy to migrate to mobile on the Android platform.

According to our experimental tests, the bit error rate of the 1024-bit signature certificate transmitted by the voice channel can be kept within 3.9%. Moreover, the certificate verification algorithm can verify the caller number. As for speaker recognition, our tests show a recognition rate of 72%. We are eager and optimistic about the work to be done to improve TAES in the future.

## REFERENCES

[1] Zhigang Wang, Jinguang Qu. Research on Telecommunitions Fraud Governance Technology based on Big Data [J]. Telecom Engineering Technology and Standardization 2017,30(04):86-89.

[2] Mustafa H, Xu W, Sadeghi A R, et al. End-to-end detection of caller ID spoofing attacks[J]. IEE E Transactions on Dependable and Secure Computing, 2016, 15(3): 423-436.

[3] Hua Chai .Overview of telecom Fraud Prevention Technology Patents [J]. Science and Technology Innovation,2019(31):69-70.

[4] Reaves B, Blue L, Abdullah H, et al. Authenticall: Efficient identity and content authentication f or phone calls[C]//26th USENIX Security Symposium (USENIX Security 17). 2017: 575-592.

[5] Silent Phone https://play.google.com/store/apps/details?id=com.silentcircle.silentphone

[6] Tu H, Doupé A, Zhao Z, et al. Sok: Everyone hates robocalls: A survey of techniques against te lephone spam[C]//2016 IEEE Symposium on Security and Privacy (SP). IEEE, 2016: 320-338.

[7] Reaves B, Blue L, Traynor P. Authloop: End-to-end cryptographic authentication for telephony ov er voice channels[C]//25th USENIX Security Symposium (USENIX Security 16). 2016: 963-9 78.

[8] Atal B S, Hanauer S L. Speech analysis and synthesis by linear prediction of the speech wave[J]. The journal of the acoustical society of America, 1971, 50(2B): 637-655.

[9] Vergin R, O'Shaughnessy D, Farhat A. Generalized mel frequency cepstral coefficients for large-v ocabulary speaker-independent continuous-speech recognition[J]. IEEE Transactions on speech and audio processing, 1999, 7(5): 525-532.

[10] Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition[J]. IEEE transactions on acoustics, speech, and signal processing, 1978, 26(1): 43-49

[11] Wang L, Kitaoka N, Nakagawa S. Robust distant speaker recognition based on position-dependent CMN by combining speaker-specific GMM with speaker-adapted HMM[J]. Speech communicatio n, 2007, 49(6): 501-513

[12] Fang Zheng, Lantian Li,Hui Zhang, Escale Luzi. Research on information security,2016,2(01):44-57.

[13] Reynolds D A, Quatieri T F, Dunn R B. Speaker verification using adapted Gaussian mixture mo dels[J]. Digital signal processing, 2000, 10(1-3): 19-41

[14] Reynolds D A, Rose R C. Robust text-independent speaker identification using Gaussian mixture speaker models[J]. IEEE transactions on speech and audio processing, 1995, 3(1): 72-83

[15] Dehak N, Dumouchel P, Kenny P. Modeling prosodic features with joint factor analysis for speak er verification[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2007, 15(7): 20 95-2103

[16] Dehak N, Kenny P J, Dehak R, et al. Front-end factor analysis for speaker verification[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2010, 19(4): 788-798

[17] Qunwei Hu. Study on Channel and Time Mismatch Compensation in Speaker confirmation [D]. University of Science and Technology of China,2016.

[18] Dhananjay A, Sharma A, Paik M, et al. Hermes: data transmission over unknown voice channels [C]//Proceedings of the sixteenth annual international conference on Mobile computing and networ king. 2010: 113-124

[19] Longfei Liu. Design and Implementation of Voice Channel Data Transmission Algorithm [D]. Xidian University,2018.

[20] Zhiyi Zhang. Research and Application of voice-print Recognition Technology [D]. Nanjing University of Aeronautics and Astronautics,2016.

[21] Nakagawa S, Wang L, Ohtsuka S. Speaker identification and verification by combining MFCC and phase information[J]. IEEE transactions on audio, speech, and language processing, 2011, 20(4): 1085-1095.