Journal of applied science and engineering innovation Vol.1 No.3 2014
ISSN (Print): 2331-9062
ISSN (Online): 2331-9070

A Voice Biometrics-based Web Authentication Scheme

Yu Luan[1, a],HongzuoLi[1,b] and Yafei Wang[2,c]

[1]School of Electronics and Information Engineering, Changchun University of Science and Technology, Changchun 130022, China

[2]English Department, North China University of Technology, Beijing100041, China

[a]171181393@qq.com, [b]lihongzuo@sohu.com, [c] faye791108@163.com

Abstract. This paper proposes a novel voice biometric-based web authentication framework with a secure mechanism. The contribution of this paper is to introduce voice biometrics verification with a high security as a new web authentication way to replace the traditional way by password. On the voice biometrics scenario, security requirements of the web authentication should be considered in detail. And reasons that employ voice biometrics on computer and mobile phone for authentication rather not other biometrics have been explained. To demonstrate the effectiveness of the proposed framework, an implementation with Single Sign On for web authentication is shown.

Keywords: Voice biometrics, biometrics, web authentication, SSO.

1. Introduction

During the past decade, a large number of governments have paid attention to biometrics. For example, American government has become a main sponsor by providing some biometric solutions to public security since the event of "9.11". Its FBI has invested 1 billion dollars to construct a comprehensive nationwide multi-biometric samples database including human DNA, fingerprint, face, etc, in order to implement human identity detection for immigration of America. American immigration has granted passport of citizens with personal face image to ensure a secure way of administration of the human identities since 2005. Germany government supported the development of the biometric security market, thus its market grew rapidly from 12 millionEuros, 2004 to 377 million Euros, 2009. In addition, Australian immigration views biometrics as a sub-component of its Smartgate immigration identitiesadministration system. So far this kind of biometric secure mechanism is only used for government, high secret department or immigration. User's web account especially e-banking account has not been protected by this means.

To improve the rank of security for web users' accounts, this paper proposes a novel voice biometric-based web authentication scheme with a secure mechanism. After the paper analyzes the design and security requirements of the scheme, the framework of the proposed scheme has been shown. And then a voice biometric authentication module in the scheme is proposed in order to provide a higher secure way to protect user's credential information. Also, the reasons why the scheme use voice biometrics rather not other biometrics, e.g, face, fingerprint, finger vein etc, are explained. To emphasis the availability of proposed methods, its SSO implementation is given.

2. Related works

Biometrics. Biometrics is a human identity detecting technique by authentication or identification algorithms with human being's biometric samples collected by sensors, e.g., human's face image [1], palmprint [2], finger vein [3], voice [4], iris [5], fingerprint [13] etc.

The collective sensors include digital camera, webcam, infrared CCD camera, digital tablet, microphone, etc. Biometric authentication is a higher secure solution than password in web authentication. Its features are listed as follows.

1. Biometrics is of a human being's natural characteristic as a unique identity identifier to replace with traditional password for authentication. So user does not need remember it, but shows his or her biometrics each authentication time.

2. Compared to the password, biometric features are hard to be duplicated, distributed, forged and destroyed.

3. Due to the nature of biometrics, it does not involve with the case that writes down the pin number or password on the paper. Therefore, biometric authentication ensures that it has a higher security than ever.

4. Due to the uniqueness of biometrics, it is impossible for multiple persons to share the same account.

Speaker recognition. Speaker recognition, also called voice biometrics [6], is a branch of biometrics. Regardless of anatomy, physiology, and acoustics, it is well known that no person has the same voice in the world [7]. Thus, human's voice can be utilized as a biometrics to detect human's identities. In 1962, Kersta et al. [8] proposed a voiceprint paper on Nature, and they firstly proposed a term named voiceprint identification. From that, the voiceprint identification was accepted by mass and media, e.g., newspaper, soap, and film. In fact, Campell et al. [7] stated that this term misleads mass that they believeacousticsignal can be obtained to a graphic representation like fingerprint from some extraction methods, e.g, spectrum analysis. Normally, speaker recognition can be classified two types: one is Text-dependent speaker recognition that speaker needs to say the specific text during both training and testing sessions; the other one is text-independent speaker recognition that its training and testing sessions is without the above restriction. The text-dependent speaker recognition uses template-based matching algorithm to detect human identities, e.g., Dynamic Time Warping (DTW) [10], Hidden Markov Model (HMM) [9]. The text-independent speaker recognition is friendlier and more widely used than the text-dependent speaker recognition. The users can gain a well user experience from the text dependent system. In this paper, we mainly discuss the text-independent speaker recognition.

In fig.1, an utterance sample extracted from microphone is referred as wave signal file with time (x axis) and sample altitude (y axis) as shown as follows.

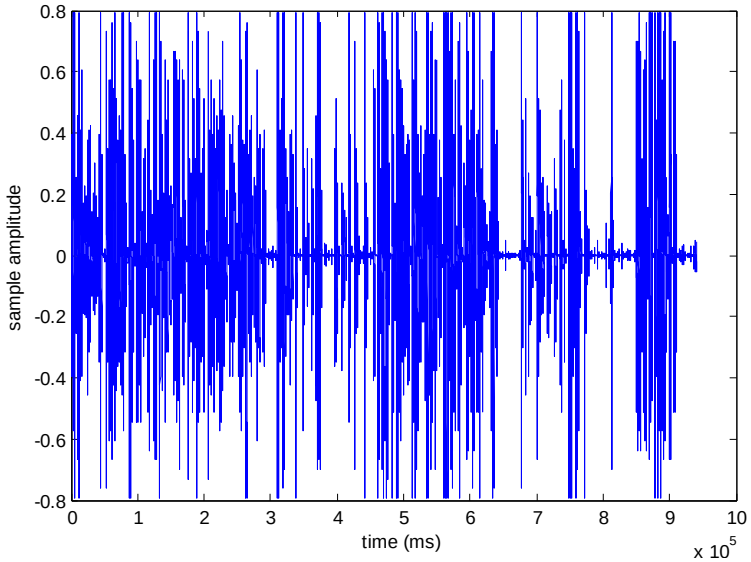

Fig. 1 A sample of utterance waveform

In fig.1, the wave file is extracted from NIST corpus [11]. The duration of the sample lasts about 900 s. This one can be used as a training sample for speaker recognition. Due to the variation of the sample altitude with time change, wave file is not directly used in the speaker recognition system, but spectrum feature, quality of signal, time duration, power of the samples, fundamental frequency etc. These can be used as features to make a verification or identification.

3. The proposed scheme

Security requirements of the proposed system. An ideal web authentication system should be of properties as follows.

A. Considering on the aspect of user

1. Reliabilityof Service Provider (SP).

In the processing of the communication between user and servers, the system should ensure SP with a valid identity all the time, in case impostor or attacker seizes sensitive data of the users.

2. Privacy of user's biometric information.

To protect user's biometrics from stealing by SP or attacker, the proposed system should be forbidden to obtain user's utterance data by SP or attacker.

B. Considering on the aspect of SP

3. Reliability of user.

SP should make sure the valid identity of whocommunicate with, rather than hacker or attacker. In some case, the web authentication mechanism requires user is anonymous to SP. Thus SP has to validate user's identity.

4. Privilege limitation.

SP often provides services with different scope, and it should have the ability to identify different user's scopes.

5. Lifecycle.

With regard to certain service, e.g., email logging on, if user needs to show his or her biometrics multiple times for signing up in a short time, this will reduce user experience. In order to solve this problem, SP should maintain user's valid identity a period of time, i.e., lifecycle, when user has signed up with his or her biometrics.

C. Considering on the aspect of the interaction between user and SP

6. Availability of transaction.

SP can notignore a request by any valid user.

7. Simplicity of process.

The process of the interaction between User and SP should be as simple as possible, avoiding a series of unnecessary operations.

8. Robustness.

Web security protocol should be compatible with terminals with different OS, e.g windows, linux for computer, and iOS, Android, WinET for cell phone. Meanwhile, the system should have robustness to error tolerance on web.

9. High efficiency.

The basic operations should be finished without latency.

10. Low cost.

The prices of the transmit devices, and related equipments should be cheap.

The proposed model. Given an Authentication Server (AS) that makes a web authentication for user c, this server uses feature vectors from user's utterance with encrypted information to authenticate, in order to ensure the availability and privacy of requests form user. If authentication is accepted, the same server or the Privilege Sever (PS) will give a grant to offer a privilege by Application Server (AppS). Thus AppS can provide the service that user required, as shown in Fig.2.

To accomplish user's services safely, PS has to ensure the identities of user c and AP as genuine. Simultaneously, considering the privacy of user utterance feature, all the authentication steps, e.g., user registration, matching, identification, should be conducted in AS. For security mechanism, we require as follows.

1. The terminals that send requests, e.g., computer or cell phone, should be equipped with microphone, in order to acquire user's utterance. Of course, cell phone has been equipped with mic by default.

2. Each AppS should have a pair of key k_{AS-pri} and k_{AS-pub} of users to ensure privacy of the offered services.

3. AS also should have a pair of key k_{AS-pri} and k_{AS-pub} of users to make an authentication for them.

4. PS should be trusted by AppS, and provide a user's privilege.

According to Fig.2, the steps in detail will be described as follows. Denotes $E(m,k)$ and $D(m,k)$ as the encrypted and decrypted functions with key k for a message m.

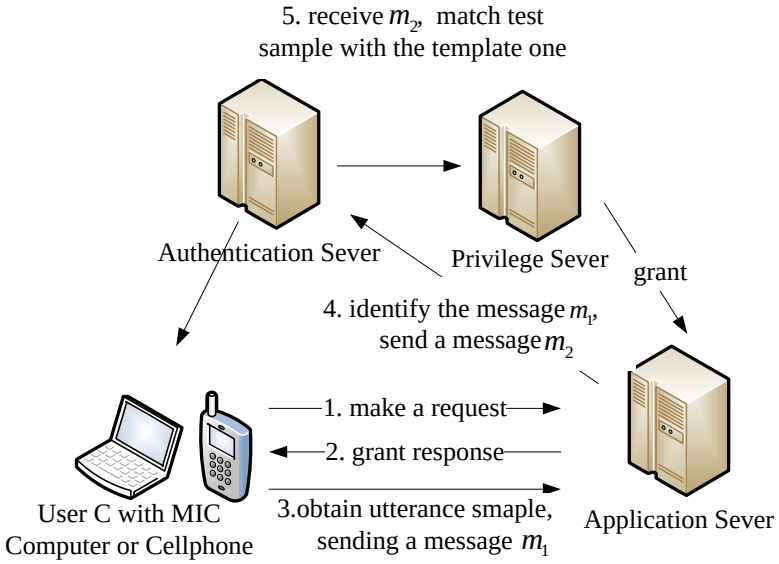

Fig.2 Flow of the proposed system with a web authentication

A. User's behavior

1. A user c on terminal send a request to AS, intending to obtain certain service.

2. AppS receives uses's message to wait the notification by AS.

3. User's utterance will be acquired by MIC on the terminal, then sends a message m_1, this message includes utterance features I_c encrypted by AS, and user credential. Notice that I_c can not be encrypted by AppS.

$$m_1 = E(E(I_c + service\_info, k_{AS\_pub}) + service\_info^*, k_{AS\_pub}) \tag{1}$$

This needs twice service information $service\_info$ and $service\_info^*$.

B. Behavior of AppS

AppS waits for a request message m_1 from user c. When the message comes, it initially ensures the availability of user c, identifying partial of the message via its private key, and then provides a service that user required.

AppS check the correctness of service information,

$$D(m_1, k_{AppS\_pri}) = E(I_c + service\_info, k_{AS\_pub}) + service\_info^* \tag{2}$$

Then send a new message m_2 which includes digital signature in m_1 to AS, so that AS can confirm that ApS makes available to transaction,

$$m_2 = E(Sign(E(I_c + service\_info, k_{AS\_pub}) + service\_info^*, ApS), k_{AS\_pub}). \tag{3}$$

C. Behavior of AS

1. When AS receives the message m_2, then uses it own key and the key of AppS k_{AppS_pub} interpret m_2,

$$\begin{aligned} m_2 &\rightarrow D(m_2, k_{AS\_pri}) = Sign(E(I_c + service\_info, k_{AS\_pub}) + service\_info^*, AppS) \\ &\rightarrow E(I_c + service\_info, k_{AS\_pub}) + service\_info^*, \end{aligned} \tag{4}$$

where it decrypts by public key of AppS.

$$E(I_c + service\_info, k_{AS\_pub}) \rightarrow I_c + service\_info \tag{5}$$

where it decrypts by private key of AppS.

D. Privilege Sever

1. PS receives authentication result and $service\_info$.

2. As to the request from user, PS creates a toke as a grant of the service that customer required, including the grant time and duration. Then PS sends them to SP to confirm the authentication result. In fact, this token needs to be encrypted. For simplifying the proposed model, we assume that it is reliable that the communication between PS and AppS.

Security mechanism of the proposed system. According to the requirements of section 3.1, the satisfied security mechanisms in the proposed system are shown as follows.

1. AS has arbitrated the transaction among user and servers, which satisfies with the property 6. AS verifies user's voice biometric authentication and checks digital signature of AppS, so that it can offer reliability of the two parties, which satisfies with the property 1 and 3.

2. User's credential, i.e., biometric information has been protected, which can be only used for matching algorithm. This avoids AS administrator stealing user's voice biometrics. This point satisfies with the property 2.

3. PS sends user's scope and lifecycle time to SP, in order to ensure the security of the token, which satisfies with the property 4 and 5.

4. User is easy to sign up without any password or key, so that user's public key official authentication can be avoided, which satisfied with the property 7.

5. The keys of AppS and AS have been encapsulated in the administration software, so users are unnecessary to remember a complicated key.

6. The mechanism of the message confirmation makes sure that user and AppS would be notified the transaction result, so that it has a strong robustness to the web system errors, which satisfies with the property 8.

7. The process of user authentication should be implemented in a short of time, approximately less than 1 second. And the transmitted data should be small. This satisfies with the property 9.

8. The cost of online voice biometric exactor, i.e., microphone can maintain within a low price, which satisfied with the property 10.

Authentication module. In the step 4 of the fig.2, the process of biometric authentication should be described as follows.

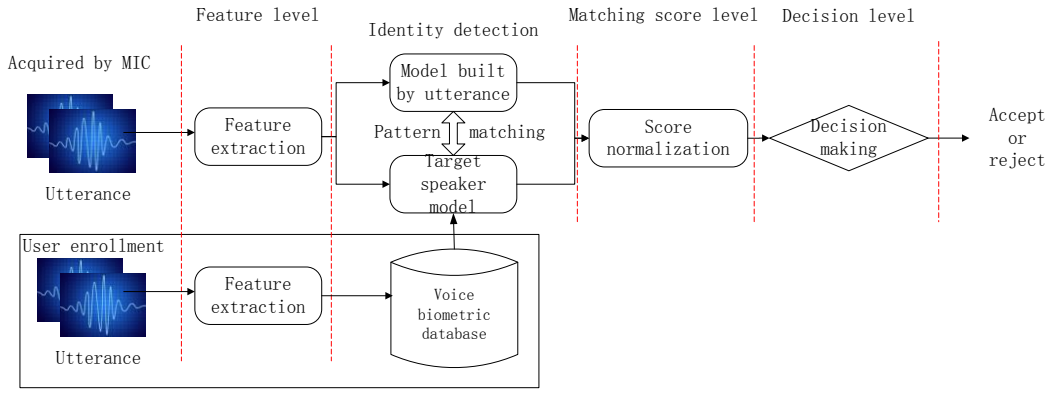

Fig.3 The module of speaker verification

From the above figure, an utterance is acquired by MIC on computer or mobile phone, and then the system proposes feature exaction by certain methods in the feature level, e.g., MFCC extraction. After that, the model built by testing utterance is matched with the target speaker model claimed. The outcomes of the pattern matching are matching scores. The target speaker model is trained by voice samples from the voice biometric database which is in the procedure of user's enrollment. In the matching score level, score normalization will be conducted, in order to make all the scores within the range of [0, 1]. In the decision level, the system can compute the final result, i.e., accept or reject the utterance by the decision making procedure.

In the procedure of the pattern matching, GMM-UBM model [12] should be utilized to detect speaker's identity. Preliminarily, Gaussian Mixture Model (GMM) [14] should be introduced. GMM is a model that estimates probability density of each speaker's utterance, i.e., probability density who speaks. GMM is summed by M single Gaussian models with their own weights,

$$p(x \mid \lambda) = \sum_{i}^{M} w_i p_i(x), \tag{6}$$

where the testing sample x is a feature vector with d dimensionality, and w_i is a weight of each Gaussian component and $\sum_{i=1}^{M} w_i = 1$. $p_i(x), i = 1 \cdots M$ denotes as a probability density of the single Gaussian model. The means μ of the Gaussian component are vectors with dimensionality of $d \times 1$. The dimensionality of Covariance matrix Cov_i is $d \times d$. For each single Gaussian model can be expressed as

$$p_i(x) = \frac{1}{(2\pi)^{d/2} \mid Cov_i \mid^{1/2}} \exp\{-\frac{1}{2}(x-\mu_i)' \sum_{i}^{-1} (x-\mu_i)\} \tag{7}$$

From the equation (6) and (7), a Gaussian model is composed of the parameters mean vector, covariance, and weight. Thus the parameter model is

$$\lambda = \{w_i, \mu_i, Cov_i\}, i = 1, \cdots, M \tag{8}$$

A speaker model can be expressed by a GMM model. The training procedure of a GMM model is to estimate the best parameters λ for each Gaussian component. Expectation Maximum (EM) [15] is usually employed for the training procedure.

Now GMM-UBM model will be introduced. For a speaker S, its GMM, i.e., the equation (6) can be modified as

$$p(x \mid \theta) = \sum_{i=1}^{n} w_i g(x; \mu_i, C \text{ov}_i) \tag{9}$$

where Cov_i is a diagonal matrix. In the testing procedure, system would identify whether the feature vector of the testing utterance x belongs to the speaker S or not. Its likelihood ratio model is

$$p(x \mid m) = \frac{p(x \mid \theta_{hyp})}{p(x \mid \theta_{\overline{hyp}})} \geq \tau \tag{10}$$

where θ_{hyp} is a hypothesis that the utterance x belongs to this speaker, and $\theta_{\overline{hyp}}$ is the one that the utterance belongs to impostor. $p(x \mid m)$ denotes as likelihood ratio of GMM, i.e., matching score. τ is a threshold. Ideally, $\theta_{\overline{hyp}}$ should have a wide range of impostor speaker's utterances. However, it is hard to implement in practice. To solve this problem, Universal Background Model (UBM) [16] can be replaced to imposter model. This UBM is a generalized GMM background model, usually trained by a large number of utterance samples from different speakers. In contrast, the model corresponding to θ_{hyp} is a GMM model trained by the target speaker's utterance. Thus the system of speaker recognition detects the identity of testing utterance by likelihood ratio

$$\log p(x \mid m) = \log \frac{p(x \mid \theta_{hyp})}{p(x \mid \theta_{\overline{hyp}})} = \log p(x \mid \theta_{hyp}) - \log p(x \mid \theta_{\overline{hyp}}) \tag{11}$$

Discussion. The reason why we choose voice biometrics rather not other biometrics, e.g., face, fingerprint, finger vein, handwriting, etc, is that voice biometrics is most feasible biometrics especially on mobile communication. Recently, mobile network services are so popular that mobile e-commerce, mobile location service, etc emerges. Normally, a feature can be viewed as a voice biometrics with the following properties:

This feature should be of intra-class and inter-class discriminate;
The feature should have anti-noisiness, robust to the signal distort.
It exists frequently and naturally in the utterance signal.
It should be extracted conveniently from the utterance.
It is hard to imitate.
It is invariance to the change of human being's physical condition (health or illness), and time.

The voice biometrics can be easily extracted with MIC from not only computers, but also mobile phones. In contrast, face biometrics has to be considered the factors of the complex background, occlusion and noises subverted if this kind of biometrics is utilized. For fingerprint, despite the fingerprint has been widely used on computers, it is merely incorporated within the mobile phone. The requirement of finger vein extraction is so strict that its implementation can be hardly done so far. And handwriting must need a special tablet on computers and some mobile phones can not support this functionality of handwriting or without the interface of handwriting.

4. A real case of voice biometrics-based web authentication system

Single Sign On. Nowadays SSO [17] is a popular solution which user only logs in once, he or she is able to access all the application or modules that they trust each other in different websites. The deployment of SSO decreases network services burdens, and shortens the time cost that administrator adds, deletes, and modifies user credential information. Thus the system offers a better administrative solution of accounts information. The user experience turns to be more friendly and higher efficiency than ever. In this section, we will propose a voice biometric-based SSO scheme based on our idea as below.

Voice biometrics-based SSO system. User registration.

A user wants to access an application or service on a website. It is inevitable that this user should make a registration at first. The steps of the registration are shown in Fig.4.

1. User Equipment (UE) makes a registration request from user terminal e.g., a computer or mobile phone to Authentication Sever (AS), then the user fills in his or her credential information and upload his or her voice biometric samples by sensors;
2. When AS receives user utterance, it makes a request that send user credentials to Storage Server (SS);
3. SS responses AS and agrees to receive user credentials;
4. AS transmits user credential to SS and the user enrollment can be implemented within the goal;
5. SS replies AS that transmission is completed.
6. AS replies UE that registration is successful.

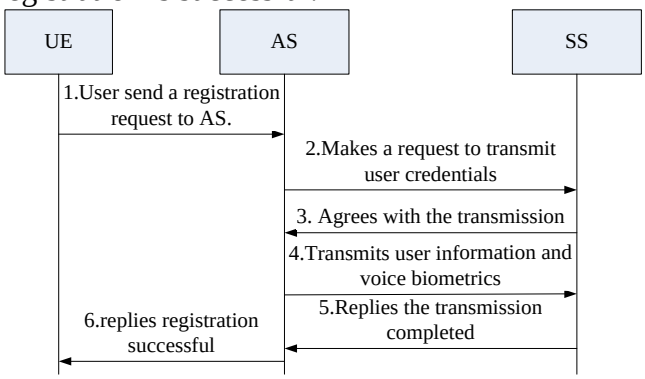

Fig.4 User registration

SSO authentication.

The steps of SSO authentication for an open platform are shown in the Fig. 5.

1. Tester who wants to entry makes a request of logging on by using claimed id and uploading his or her voice biometrics, i.e., utterance acquired by MIC to AS via UE;
2. AS sends a request that fetches the claimed id's information and its corresponding utterance samples to SS;
3. SS responses and sends the claimed id credentials to AS;
4. AS uses the algorithm mentioned in section 3.4 to match the model of the target speaker and tester's, in order to make an authentication for tester. If match successfully, it indicates the identity of the tester is identical to the claim id; vice versa.
5. If authentication is successful, AS sends a token generation request to Application Sever (AppS);

6. User sends a request that accesses certain service or application on the website, and sends encrypted token to AppS by user privacy key.

7. AppS sever handles operations user required;

8. AppS returns a result of the operation back to UE.

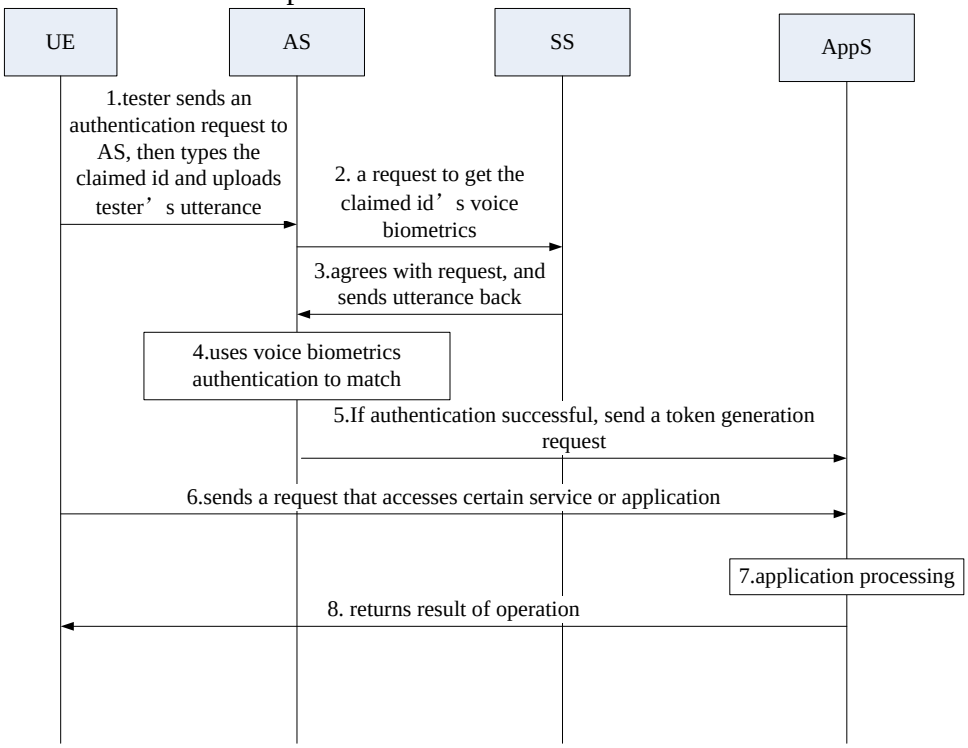

Fig.5 SSO authentication

5. Summary

This paper proposes a novel voice biometric-based authentication web authentication with a secure mechanism. After the paper analyzes the design and the security requirements of the scheme, the framework of the proposed scheme has been shown. And then a voice biometric authentication model has been demonstrated in detail. Also, the reasons why the scheme use voice biometrics rather not other biometrics, e.g, face, fingerprint, finger vein etc, are explained. To emphasis the availability of proposed methods, its SSO implementation is given. In future, the problem that voice biometrics may be easily recorded by other mobile, recorder, or even user own mobile phone with control of the malicious software or Trojan when user makes an authentication should be taken into account.

References

[1] S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz, “fusion of face and speech data for person identity verification”, IEEE Transaction on Netrual Network, Vol. 10, No. 5, 1065-1074, 1999.

[2] A. Kong, D. Zhang and M. Kamel, “A survey of palmprint recognition”, Pattern Recognition, Vo. l 42, No. 7, p. 1408-1418, 2009.

[3] Z. Li, D. Sun, D. Liu, H. Liu, “Two Modality-Based Bi-Finger Vein Verification System”, The 2010 International Conference on Signal Processing, p.1690-1693, 2010.

[4]W. M. Campbell, “Speaker recognition: A tutorial”, The IEEE Proceeding, Vol. 85, No. 9, 1437-1462, 1997.

[5]J.Daugman,“Combining Multiple Biometrics”, Available on http://www.cl.cam.ac.uk/users/jgd1000/combine/combine.html.

[6] W. M. Campbell, "Speaker recognition: A tutorial", The IEEE Proceeding, 1997, 85(9):1437-1462.

[7] J. P. Campbell, W. Shen, W. M. Campbell, et al., "Forensic speaker recognition", IEEE Signal Processing Magazine, 2009,26(2):95-103.

[8] L. G. Kersta, "Voiceprint identification",Nature, 1962, 196(4861): 1253–1257.

[9] R. Lawrence, "A tutorial on hidden markov models and selected applications in speech recognition", Proceedings of the IEEE, 1989,77(2):257-286.

[10] I. Shahin, and N. Botros, "Speaker identification using dynamic time warping with stress compensation technique", in Proc.of the IEEE Southeastcon'98, 1998: 65-68.

[11] M. Plumpe, T. Quatieri, D. Reynolds, Modeling of the glottal flow derivative waveform with application to speaker identification, IEEE Trans. Speech Audio Process., vol. 7, no.5, 569-586, 2006.

[12] S. Zhong, MSc Thesis: Speaker Segmentation and Verification, Nanyang Technological University, 2008.

[13] L. Hong and A. K. Jain, "Integrating faces and fingerprints for personal identification", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.20, No.12, p.1295-1307, 1998.

[14] D. Reynolds, T. Quatieri, R. Dunn, Speaker verification using adapted Gaussian Mixture Models, Digital Signal Processing 10, 19-41, 2000.

[15] T. Kinnunen, and H. Li, "An overview of text-independent speaker recognition: from features to supervectors", Speech Communication, 2010, 52(1): 12-40.

[16] E. Vale, A. Alcaim, Adaptive weighting of suband-classifier responses for robust text-independent speaker recognition, Electronics Letters,Vol.44, No.21, 2008.

[17] J. Byous, Single Sign-on Simplieity with SAML an Overview of Single Sign-on Capability Based on the Seeurity Assertions Markup Language(SAML)Speeifieation． 2002