SAML Single Sign-On Protocol Development Using Combination of Speech and Speaker Recognition

Patrick Telnoni
Sekolah Teknik Elektro dan
Informatika
Institut Teknologi Bandung
Bandung, Indonesia
patrick.telnoni@students.itb.ac.id,
ptelnoni89@gmail.com

Rinaldi Munir
Sekolah Teknik Elektro dan
Informatika
Institut Teknologi Bandung
Bandung, Indonesia
rinaldi@informatika.org

Yusep Rosmansyah
Sekolah Teknik Elektro dan
Informatika
Institut Teknologi Bandung
Bandung, Indonesia
yusep@stei.itb.ac.id

Abstract - Single sign on (SSO) is a centralized authentication system where user acquires access to many services using one credential. However, Single Sign On has one fatal weakness. If one credential has been acquired by an attacker, the attacker can acquire access to many services. SAML SSO protocol, which widely studied also suffers this vulnerability. This research provides economic solution for such problem by combining speech and speaker recognition and apply it to SAML SSO protocol.

Keywords: Single sign-on, biometric, SAML, speech, speaker recognition

I. Introduction

With the increase of service provided on the internet which requires registration (sign-up) thus increase the number of credential possessed by a single user. To solve this problem, a centralized authentication system is required so that one user can access various services using a single credential. This system is called Singe Sign-On (SSO) [5, 11].

The main problem in SSO is the credential used for authentication usually use text based credential such as username and password. Such type of credential can be easily stolen, thus when successfully stolen, an attacker can acquire access to many services with just one authentication process. To prevent impersonator to acquire access to the impersonated user, a unique of credential is required. Unique credential which is difficult to be stolen is biometric. Considering the increased presence of microphone devices from 2006 [18] and low cost censor compared to other biometric acquisition devices we use voice biometrics for SSO system to prevent impersonator to be authorized in SSO System.

II. Background

A. SAML

SAML [2, 14] is a XML based SSO protocol which transport security assertions and corresponding protocol messages using XML format. SAML allows so-called protocol bindings that embed SAML constructs in other structures for transport. For instance, SAML builds the Simple Object Access Protocol (SOAP) with its SOAP over HTTP binding [2]. These so-called profiles contain protocol flows and security constraints for applications of SAML. The example of SAML message is shown in Figure 1.

```
<saml:Subject>
    <saml:NameID  Format="urn:oasis:
names:tc:SAML:1.1:nameid-ormat:
unspecified">
patrick.telnoni@students.itb.ac.id
   </saml:NameID>
    <saml:SubjectConfirmation
Method="urn:oasis:names:tc:SAML:2.0:cm:bearer">
        <saml:SubjectConfirmationData
NotOnOrAfter="20014-05-30T02:44:24.173Z"
Recipient="http://localhost:8080"/>
    </saml:SubjectConfirmation>
</saml:Subject>
```

Figure 1. SAML Assertion example

In contrast to OpenID where every party can be ID Provider, SAML pairs ID Provider to Service Provider. SAML uses HTTP, SMTP, FTP and SOAP to send its assertion. SAML has component as follows:

1) Assertions
   SAML allows for one party to assert security information in the form of statements about a subject. For instance, a SAML assertion could state that the subject is named "Patrick Telnoni", has an email address of patrick.telnoni@s.itb.ac.id, and is a student of STEI ITB.
2) Protocol
   SAML has few protocols to handle request and response.
3) Binding
   SAML Binding has few protocols to carry messages over underlying transport protocol.
4) Profiles
   SAML profiles define how the SAML assertions, protocols, and bindings are combined and constrained to provide greater interoperability in particular usage scenarios.

B. Biometric

Biometric [4] is a method to recognize a person by using one's biologic characteristics and behavior. Biometric is used to recognize someone with assumption that a person is unique physically and behaviorally. Biometric system is widely used to recognize and to restrict access to certain room, information, service and even to cross country borders.

978-1-4799-5100-0/14/$31.00 ©2014 IEEE

Biometric has common components as follows [8]:

1. Capture
   Component to capture user's biometric sample.
2. Reference database
   Store data about registered users on a system, including recorded biometric.
3. Matcher
   Compare biometric data acquired from biometric censor with recorded biometric data in database.
4. Action
   Decision about action that will be executed by system according to results acquired by matcher.

C. Speaker Recognition

Speaker recognition [10] is a method to recognize voice's owner based on stored voice pattern. Voice biometric used in this research related to speech processing.

In speaker recognition, the conducted processes are capture voice, feature extraction, pattern matching and decision regarding the result of the matching process. Figure 2 shows general speaker recognition process.

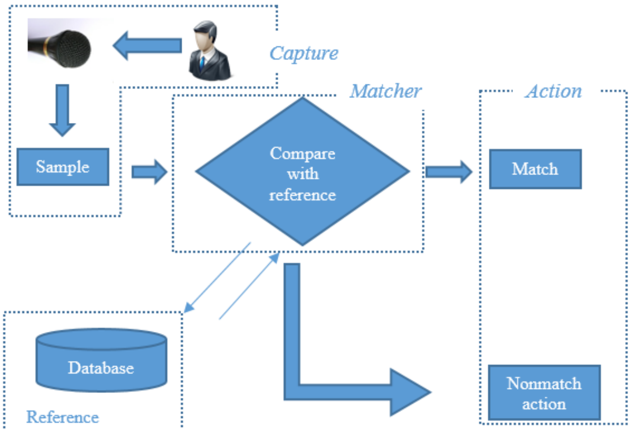


Figure 2 Speaker recognition process

D. Speech Recognition

Speech recognition [13] is a method to recognize the words spoken by user. Speech recognition is quite similar to speaker recognition. The key factor that distinguishes speaker recognition and speech recognition is matching result of feature extraction. In speec recognition, the result of feature extraction is matched against acoustic model. Result of feature extraction also matched against language model to predict the next word after the currently spoken word recognized [12]. In general, speech recognition process is shown by Figure 3.

III. Related Works

There is related work about SAML using two way authentication. In [16], SAML is combined with two way authentication using XACML instead of biometric. XML based Web service standard for communicating access control policies between services. However, because XACML is based on XML, it suffers vulnerability to XML Signature Wrapping [17]. However, research in [17] also proposes practical countermeasures that are easily implemented.

There are many related works preformed to use biometric for authentication, but related work about using biometric for Single Sign-On is still rare. Therefore, we studied related work about SAML and biometric. Related works about biometric also performed to test various biometric. In [4], various biometrics are tested. The tested biometrics are fingerprint, iris, face, voice, handwriting and hand vein pattern. In [4], the biometric which gave best result are voice and handwriting.

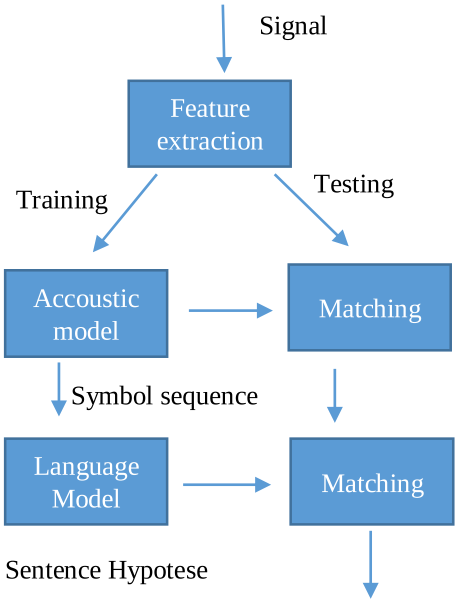


Figure 3 Speech recognition process

In [6], the tested biometrics are hand vein patterns. In this research, test is preformed to 100 users and 0,1% acquired false acceptance rate for 98,5% genuine acceptance rate. However, considering the price of infrared thermal camera, we decide not to use hand vessel pattern.

Another research performed in [12] uses keystroke for authentication.. Keystroke rhythm for each user is unique. Still, keystroke is not widely developed into popular programming language like other biometric such as voice or face biometric.

Another research in [7] performed to design Biometric Authentication as a Service (BioAAS). In this research, SAML is technical aspect is not the main concern. The main concern of this research to build BioAAS with data protection using the regulation Europe regulation about data protection.

We didn't find related work about SAML using biometric. Therefore, we develop SAML SSO Protocol using biometric to overcome impersonation. We choose voice biometric by considering the economic aspect and performance of biometric based on related work in [4, 12].

IV. Analysis and Design

SAML protocol SSO is more widely studied protocol compared to OpenID and OAuth. In research [1], SAML is known to have vulnerability as is shown by Figure 4. In Figure 4, communication between ID Provider and Service Provider can be intercepted by the interceptor and can be inserted malicious content.

In Figure 4, the red color represents action and result performed by an interceptor. This vulnerability from Figure 4 can lead to attacks as follows:

1. Cross site scripting (XSS),
2. Cross Site Resource Forgery (CSRF),
3. Unvalidated Redirects and Forwards.

Attack number 1,2 and 3 is listed OWASP Top 10 Vulnerability list 2013 [13]. Attack number 1 and 3 is an attack which can lead access and credential theft. In SSO, with single credential, user can acquire access to many services using single credential. Therefore, after user's credential successfully stolen, attacker can acquire access to many Service Providers by authenticated once. In [9], it is suggessted to combine various technique of authentication to obtain better security. This techniques is called two factor authentication. There are three type of credential mentioned in [9]:

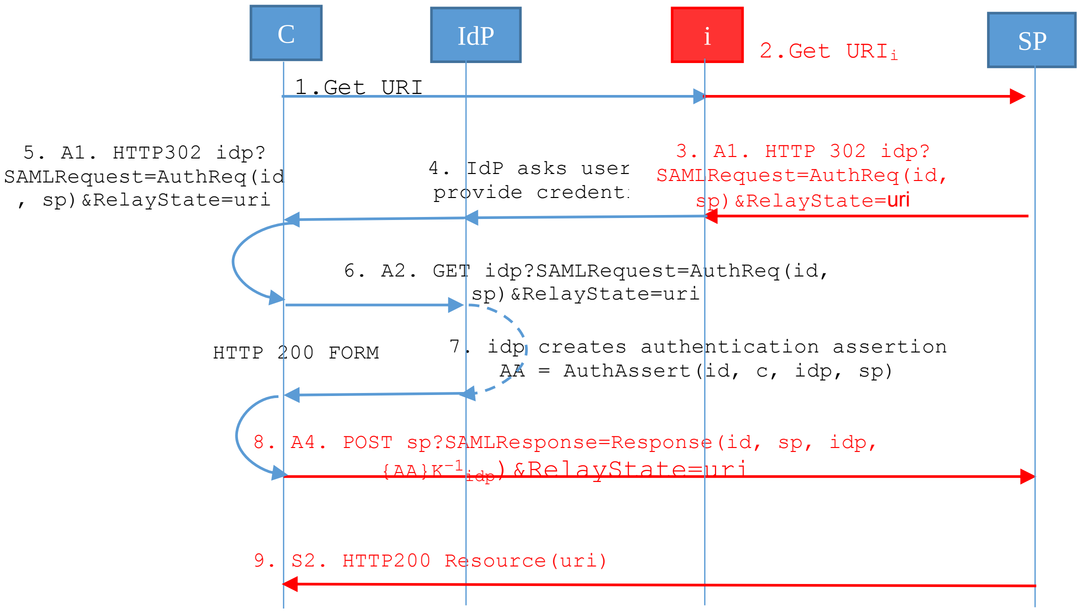

C
IdP
i
SP
1.Get URI
2.Get URI$_i$
3. A1. HTTP 302 idp? SAMLRequest=AuthReq(id, sp)&RelayState=uri
4. IdP asks user provide credent
5. A1. HTTP302 idp? SAMLRequest=AuthReq(id , sp)&RelayState=uri
6. A2. GET idp?SAMLRequest=AuthReq(id, sp)&RelayState=uri
HTTP 200 FORM
7. idp creates authentication assertion AA = AuthAssert(id, c, idp, sp)
8. A4. POST sp?SAMLResponse=Response(id, sp, idp, {AA}K$^{-1}_{idp}$)&RelayState=uri
9. S2. HTTP200 Resource(uri)

Figure 4. SAML Vulnerability

1. Something you know, typically username and password.
2. Something you have, type of credential using things like smartcard an ATM
3. Something you are, using something that user is, typcally biometric.

Type number 1 can be easilly obtained by an attacker using may method. Type number 2 requires more cost than type number 3. To use type number 2, we need card manufacturer, which is hard for us to find. Thus, we choose type number 3. We will choose biometric with cheap censor.

Based on the analysis, biometric is added for authentication at ID Provider. Designed SSO will use speaker recognition to overcome access theft. We choose voice biometrics for consideration as follows:

1. The microphone has become embedded hardware in every computer and smartphone.
2. The microphone is cheap as censor for biometric
3. Voice biometrics is more difficult to penetrate than face biometric. Nowadays, face biometric is more penetrable using 3d printer.

Using voice biometric, processes in Figure 4 modified into Figure 5. The green part in Figure 5 represents proposed solution. Further detail about process is shown by Figure 6.

Voice biometric also can be easily breached using high definition recording. Therefore, in designing the SSO system, we add speech recognition so that when a user's voice has been successfully recorded by an attacker, the speech recognition works as guard thus prevent the attacker from being authenticated and authorized to ID Provider. At ID Provider, user is prompted to pronounce specific keywords. While user pronounce the keywords, user's voice is recorded. If user passes the speech recognition process, information from the authentication form (username, password and recorded voice) will be sent to server. If not, user will be prompted to re-pronounce the keywords.

From the proposed solution, we define three use cases for the system as follows:

1. Registration, define registration process including recording user's voice for training purpose.
2. Login, define authentication process using username, password and speech and speaker recognition.
3. Logout, define how user perform logout from active session.

V. Implementation

In the implementation phase, we use HTML5 to conduct voice recording and speech recognition. HTML5 is chosen as front-end component because HTML5 has become popular client side technology with many functionalities including accessing the client's microphone. While HTML5 is applied to access user's microphone and display speech recognition result, the speech recognition process is conducted by Google server. For speaker recognition process, we use Modular Audio Recognition Framework (MARF) which has been widely used for research purpose.

Figure 7 describes workflow of designed SSO. Each step described as follows:

1. User request access to service provider.
2. Service Provider redirects user to ID Provider. Id Provider prompts user to provide username and password. In this step, user also prompted to mention generated keyword.
3. ID Provider contact google to perform speech recognition
4. Google send speech recognition result
5. ID Provider performs login process using username and password. ID Provider also performs speaker recognition. After authentication and speaker recognition succeed, ID Provider redirect user to services provider

Figure 8 shows authentication form in ID Provider which contains username and password field and speech recognition field.

MARF is configured based on [15], where Endpoint (Preprocessing method), Linear Predictive Coding (Feature Extraction method), Chebysev Distance (Classification method) give highest acceptance rate (82.76%). However, since MARF framework we use in this research is not complete, we use Raw (Preprocessing method)Fast Fourier Transform (Feature Extraction method), Eucledian Distance (Classification method) which give 75.86% acceptance rate.

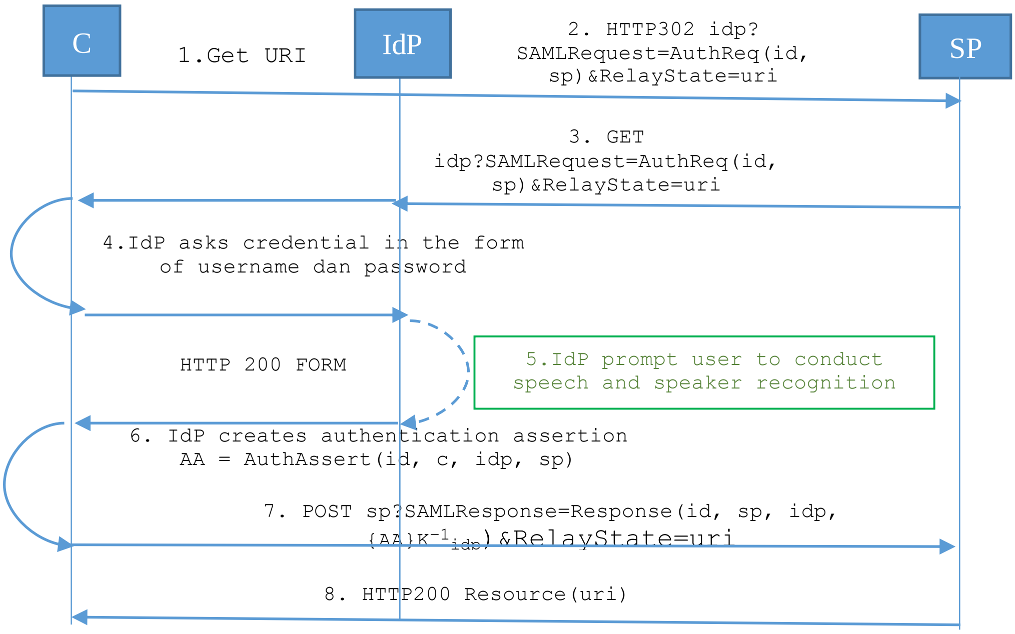

C
IdP
SP
1.Get URI
2. HTTP302 idp?
SAMLRequest=AuthReq(id,
sp)&RelayState=uri
3. GET
idp?SAMLRequest=AuthReq(id,
sp)&RelayState=uri
4.IdP asks credential in the form
of username dan password
HTTP 200 FORM
5.IdP prompt user to conduct
speech and speaker recognition
6. IdP creates authentication assertion
AA = AuthAssert(id, c, idp, sp)
7. POST sp?SAMLResponse=Response(id, sp, idp,
{AA}K-1idp)&RelayState=uri
8. HTTP200 Resource(uri)

Figure 5.Proposed solution

VI. Testing and Explanation

The environment to perform testing is described as follows:
Hardware environtment consist of:

1. Prosesor: Intel® Core™ i5-2430M CPU @ 2.30 GHz
2. Memory: 4 GB
3. 16 bit microphone embedded in ASUS A43SA laptop

Software environtment consist of:

1. Operating System: Windows 8
2. IDE: Eclipse Kepler
3. Programming Language: Java 1.7.0 using Servlet
4. Apache Tomcat 7.0
5. Google Chrome Web browser version 34.0.1847.131 m

For infrastructure environment we use ADSL Internet connection 2 Mbps.

Number of user to perform test is three users. Each user has three recorded voice samples as training data. In total, there are 9 voice samples for training data.

We set test envirommnent into two condition, quiet environtment and noisy/crowded environtment. In quiet environment, we set room as quiet as possible. For noisy/crowded environment, we conduct test in office hours so the room filled with lots of people. For each environtment, we performed 10 test of each use case.

Based on the test, almost all defined use cases run well. However, main concern in this research is about prevent access theft in SSO using biometric. Therefore, we focus our analysis on login use case.

We use acquire 100% (10 of 10) of successful authentication in quiet environment and 90% (9 of 10) of successful authentication in noisy environment. Because the main The failure upon perform login use case will be explained as follows:

a. Failure on speech recognition process

Explanation:

User pronounce correct words, but speech recognition process returns different result.

Analysis:

i. Configured language for speech recognition process is english, while mother language of registered user is Indonesia. This cause misspelling which leads to inaccuracy.

ii. User pronounce prompted keywords in quick tempo.

iii. Noise from surroundings is carried to Google server for speech recognition process.

b. Failure on speaker recognition process

Explanation:

Voice sent from authentication form is genuine voice of user which match username and password, but upon speaker recognition is performed in ID Provider, the process give dffierent result from the expected one.

Analysis:

i. When ID Provider record user’s voice, environment around user is quite noisy. “Raw” preprocessing method in MARF is applied in speaker recognition process may be the cause. It means no filter applied to the recorded voice.

ii. Keyword only consists of three words and this affect the duration of recorded voice

iii. Audio format of recorded voice sent from HTML5 and audio format in ID Provider server is different. There is a quiality difference between recorded voice in HTML5 and in ID Provider’s server.

VII. Conclusion

Conclusions from this research are described as follows:

1. Using biometric to prevent access theft on SSO is quite effective.
2. The key factor for using biometric is censor’s cost and social acceptance.
3. User’s environment is the key factor in speech and speaker recognition success rate. Ideal condition is required to obtain optimum result.

VIII. Future Research Suggestion

For future research, we propose suggestions as follows:

1. Create dedicated server for speech recognition purpose, because Google’s speech recognition server can be shut down whenever Google wants.
2. Appy filter to recorded voice.
3. Use face biometric instead of voice biometric, because voice biometric’s stability depends on environment.

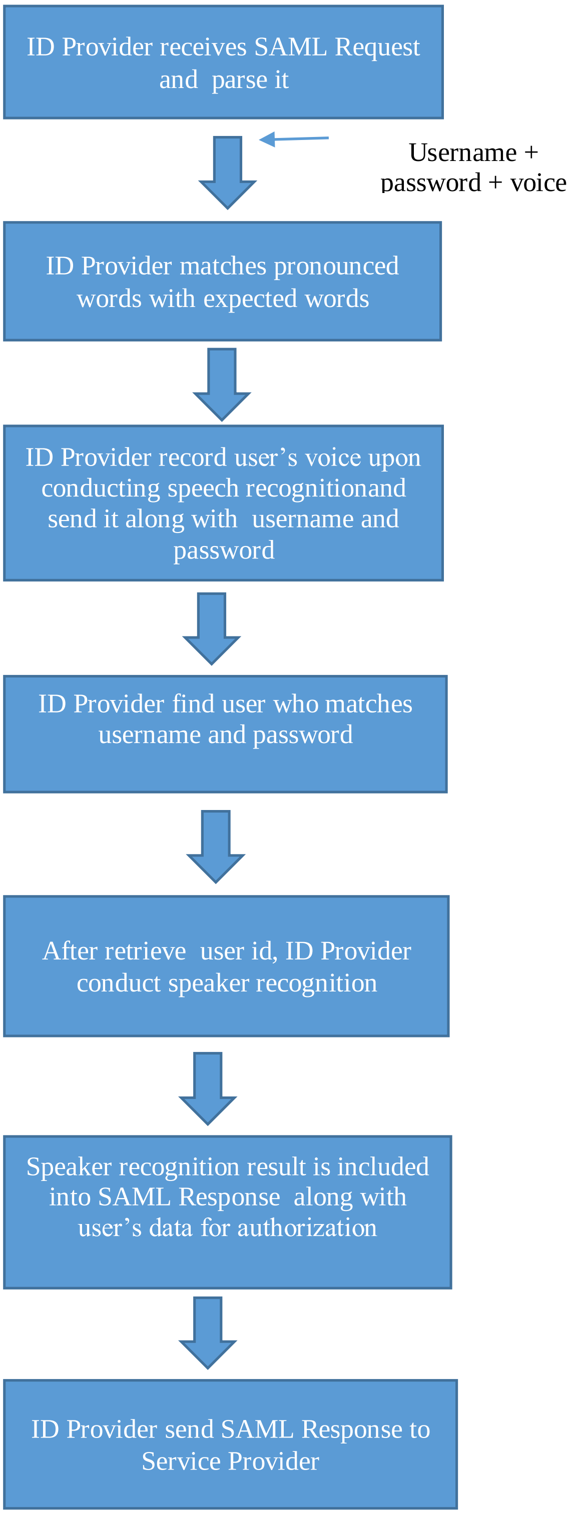

ID Provider receives SAML Request and parse it
Username + password + voice
ID Provider matches pronounced words with expected words
ID Provider record user’s voice upon conducting speech recognitionand send it along with username and password
ID Provider find user who matches username and password
After retrieve user id, ID Provider conduct speaker recognition
Speaker recognition result is included into SAML Response along with user’s data for authorization
ID Provider send SAML Response to Service Provider

Figure 6. Proposed solution (Detailed)

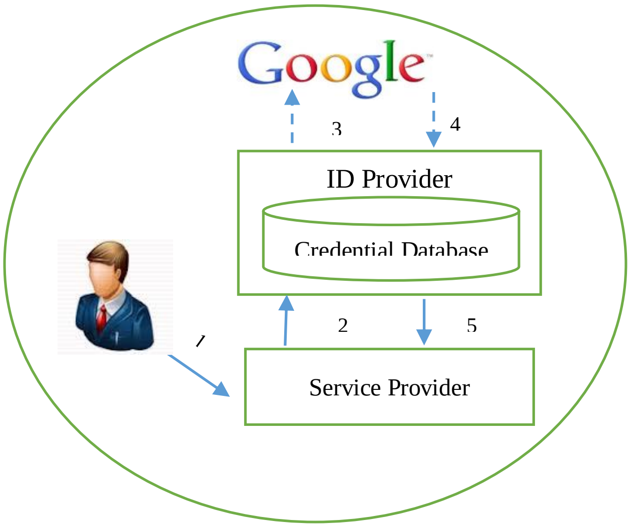
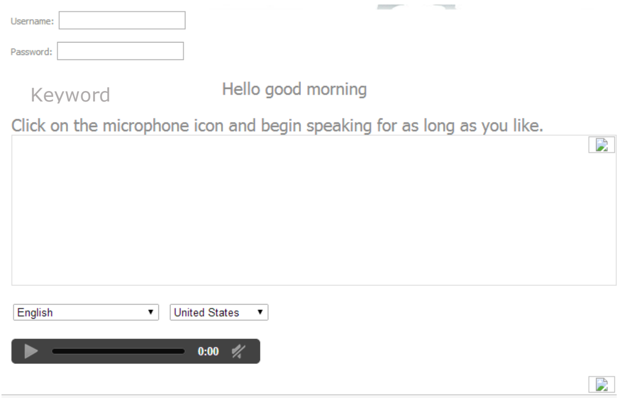


Figure 7. Flow process of designed system



Username:

Password:

Keyword Hello good morning

Click on the microphone icon and begin speaking for as long as you like.

English

United States

Figure 8. Authentication Form in ID Provider

REFERENCE

[1] A. Armando, R. Carbone, L. Compagna, J.Cuellar, G. Pellegrino and A. Sorniotti. “An authentication Flaw in Browser-based Single Sign-On Protocols: Impact and Remediations.” *Computers & Security*, vol. 33, pp.41-58, Mar.2013.

[2] T. Grob. "Security Analysis of The SAML Single Sign-on Browser/Artifact Protocol." Computer Security Applications Conference, 2003. Proceedings. 19th Annual , 2003, pp. 298 - 307.

[3] J. Somorovsky, A. Mayer, J. Schwenk, M. Kampmann and M.Jensen.. “On Breaking SAML: Be Whoever You Want to Be”, *21st USENIX Security*, vol. 33, pp. 397-412, Aug. 2012.

[4] D.Liu, Z.J. Zhang, N.Zhang. “A biometrics-based SSO authentication scheme in Telematics,”. *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discover*, 2012, pp. 191-194.

[5] V. Radhaa & D.H. Reddy. “A Survey on Single Sign-On Techniques”. *Procedia Technology*, 2012, pp. 134-139.

[6] A. Kumar, M. Hanmandlu and H. M. Gupta. “Online Biometric Authentication Using Hand Vein Patterns”. *Computational Intelligence in Security and Defense Applications*, 2009.

[7] C.Senk and F.Dosler. “Biometric Authentication as a Service for Enterprise Identity Management Deployment: A Data Protection Percpective,” Availability, Reliability and Security (ARES) Sixth International Conference, 2011, pp.43-50.

[8] J.N. Pato, J, L.I. Millet. *Biometric Recognition: Challenges and Opportunities*. Washington, D.C.: The National Academies Press, 2010, pp.1-8.

[9] N.Daswani, C. Kern, C. and A. Kesavan. “Foundations of Security What Every Programmer Needs to Know,” 1st ed., New York: Appress , 2007, pp.7-22.

[10] J.P. Campbell. “Speaker Recognition: A Tutorial”. *Proceedings of The IEEE*, vol. 85, no. 9, pp. 1437-1462, 1997.

[11] J.D. Clerq. “Single Sign On,” in *Windows Server 2003 Security Infrastructures: Core Security Features*, 1st ed., USA: Digital Press , 2004, pp.533-579.

[12] J. Roth, X.Liu, A.Ross and D.Metaxas .(2013). “Biometric Authentication via Keystroke Sound”. Biometrics (ICB), 2013 International Conference on *IEEE*, 2013, pp.1-8.

[13] K. Samudravijaya. “Speech and Speaker Recognition: A Tutorial”. *Tata Institute of Fundamental Research*.**,** 2001.

[14] OASIS. (2008). “Security Assertion Markup Language (SAML) V2.0 Technical Overview. OASIS.

[15] I. Clement, S. Mokhov, D. Nicolacopoulos and S. Sinclair. “Experimentation Results” in Modular Audio Recognition Framework v.0.3.0.6 (0.3.0 final) and its Applications,” The MARF Research and Development Group, Montreal, Rep. Qc, Dec, 2007.

[16] S. Fugkeaw, P. Manpanpanich and S. Juntapremjitt, “Adding SAML To Two-Factor Authentication And Single Sign-On Model For Dynamic Access Control,” in Information, Communications & Signal Processing, 2007 6th International Conference, Singapore., 2007, pp. 1-5.

[17] J. Somorovsky, A. Mayer, J. Schwenk, M. Kampmann and M. Jensen, “On Breaking SAML: Be Whoever You Want to Be,” *Proceeding Security'12 Proceedings of the 21st USENIX conference on Security symposium*, 2012, pp. 397-412.

[18] A. Kounoudes, V. Kekatos, and S. Mavromoustakos, “Voice Biometric Authentication For Enhancing Internet Service Security,” *Information and Commmication Technologies*, 2006, Damascus, pp. 1020-1025.