# Poster: Natural Language Voice based Authentication Mechanism for Smartphones

Ubaid Ur Rehman
Kyung Hee University
Yongin-si, Gyeonggi-do, Republic of Korea
ubaid.rehman@khu.ac.kr

Sungyoung Lee
Kyung Hee University
Yongin-si, Gyeonggi-do, Republic of Korea
sylee@oslab.khu.ac.kr

## ABSTRACT

We have designed and implement a *random text dependent voice based authentication* protocol for smartphones. The objective was to provide an efficient and reliable authentication mechanism that ensure prevention against the emerging attacks. In this paper, we have focused on the architecture, protocol, and prevention against replay attack only.

## CCS CONCEPTS

• **Security and privacy** → *Biometrics*; • **Computer systems organization** → *Distributed architectures*;

## KEYWORDS

Smartphones; Authentication; Voice; Replay Attack

**ACM Reference Format:**
Ubaid Ur Rehman and Sungyoung Lee. 2019. Poster: Natural Language Voice based Authentication Mechanism for Smartphones. In *The 17th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '19), June 17–21, 2019, Seoul, Republic of Korea.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3307334.3328645

## 1 INTRODUCTION

With the emergence and tsunami of smart devices, most people carry their smartphone, smartwatch, and computer tablet everywhere. These devices have different types of application that contain personal information such as emails, bank accounts, pictures, contacts, and social networking profiles. The misuse of this personal information may lead to serious consequences such as identity theft, financial and data loss. According to a survey conducted by *Pew Research Center* [1], a limited number of smartphone users prefer screen lock, which includes password, pin, pattern, fingerprint, and iris scanning. These authentication mechanisms are really challenging for the user, especially the elder community. As password, pattern, and pin needs to be memorized. Moreover, due to the low accuracy of the existing fingerprint and iris scanning mechanism on the smartphone, it is also dependent on the password, pattern, or pin.

In order to resolve this challenge, the authentication paradigm has been shifted to the voice user interface [3]. Therefore, it is popular in recent technologies such as smart wearable devices, smart vehicles, smart home, and virtual assistants. Moreover, it is conveniently used during driving and exercising. The existing literature shows that voice authentication is vulnerable to replay attack because the attacker may use a pre-recorded voice sample of a victim, which leads to identity spoofing[2].

## 2 NATURAL LANGUAGE VOICE BASED AUTHENTICATION MECHANISM

Figure 1 shows the architecture along with the communication protocol workflow of our designed *natural language voice based authentication mechanism using random text dependent* approach. Also, the detail of each component is described as follows:

- *User Smartphone:* For the prototype implementation, we have developed an android application that unlocks the smartphone's screen using our designed *random text dependent voice based authentication* protocol. In order to interact with the user, the application uses the built-in microphone, touchscreen, display, and speaker.
- *Authentication Server:* The authentication server deals with the generation, exchange, storage, and use of unique identity. It generates a *256 bits* unique identity upon a registration request from the user's smartphone. The unique identity is bind with the device identity. Moreover, it is used by the *AES 256* for encrypting and decrypting the allocated identity storage unit.
- *Identity Storage Unit:* The identity storage unit store the user's recorded voice sample along with the pitch of voice segments and Mel-Frequency Cestrum Coefficients (MFCC). Furthermore, it contains detail information about each recorded sample, which includes the sample rate, bits per sample, number of channels, device identity, current sample, total samples, start event, stop event, timer event, time period, tag, user data, type, and label.
- *Random Function:* The random function select records or words that cannot be predicted reasonably. It consists of Records Selector and Words Selector. The records selector select the random voice sample from the identity storage unit. While the Words Selector selects random words from the publicly available English words list.
- *Identity Tagging:* It assigns a tag to the voice sample along with the specific keyword text and time stamp.
- *Recorded Voice Decision Maker:* This module generates a notification to the user regarding a recorded voice sample. Based on the user response, the module process the request accordingly.


*MobiSys '19, June 17–21, 2019, Seoul, Republic of Korea*

ACM ISBN 978-1-4503-6661-8/19/06.
https://doi.org/10.1145/3307334.3328645

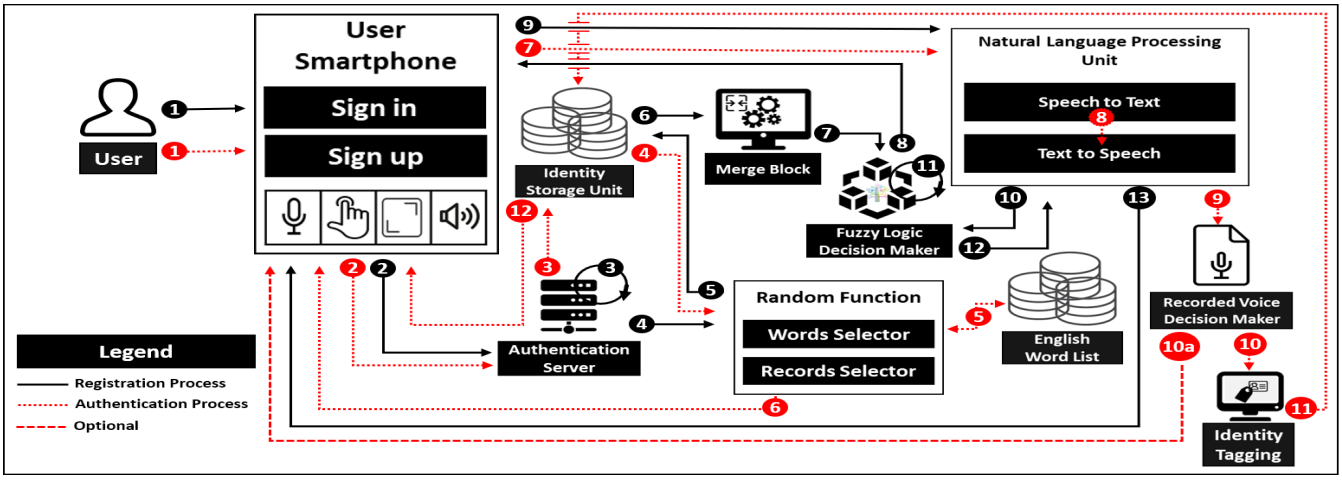


**Figure 1: Natural Language Voice based Authentication Architecture and Protocol Workflow**

- *Natural Language Processing Unit:* The natural language processing unit consists of two submodules: *i) Speech to Text* take the user voice input and convert it into text. *ii) Text to Speech* take text as an input and convert into speech.
- *Fuzzy Logic Decision Maker:* This module compares the user's input voice sample with the previously stored voice sample based on the MFCC features along with the corresponding text and takes a decision using the fuzzy logic approach. In our designed prototype implementation, if the similarity index is greater than 70% then the user gets authenticated.
- *Merge Block:* It combines the randomly selected keywords into one block. Moreover, it combines the Mel-Frequency Cestrum Coefficients (MFCC) of the randomly selected keywords and shows a visual representation of the selected keywords features.
- *English Word List:* We have used the publicly available English words list from *WordNet*, which consists of *1,44,884* words with the length of *1* to *47* alphabets word. In order to prevent a brute-force attack, we have discarded the words that are less than *5* alphabets, which give us a wordlist of *1,42,967*.

## 2.1 Protocol Workflow

- *Registration:* The user has to click on the *Sign Up* button, which generate a registration request and sent to the *Authentication Server*. Based on the registration request, it generates a unique identity and also request the *Identity Storage Unit* for allocation of isolated storage. With this the *Random Function* triggers the *Words Selector* and select *five* random words from the *English Word List*. For every iteration, the user has to speak the word displayed on the screen. Then *Natural Language Processing Unit* process the spoken word accordingly and confirm from the user with the help of *Recorded Voice Decision Maker*. Upon the user's confirmation, the *Identity Tagging* assign a specific tag to the recorded voice sample and store in the *Identity Storage Unit*. Similarly, for the remaining *four* keywords, the same steps from *6* to *12* will be repeat. Once the registration process gets completed, a notification will be generated on the user's smartphone and voice based authentication will be activated to unlock the screen.
- *Authentication:* The user has to click on the *Sign In* button, which generates an authentication request and sent to the *Authentication Server*. It extracts the unique identity and requests the *Records Selector* to decrypt the user's assigned storage and randomly select *five* voice sample text along with its corresponding MFCC features. After the selection, the *Merge Block* generates two separate blocks, the one that contains the *five* selected voice sample text separated by single space. The other block contains the MFCC features in the same sequence. Both blocks are shared with the *Fuzzy Logic Decision Maker*, which display the text block to the user. Once user says the words, the *Natural Language Processing* unit converts the spoken words into text and send to the *Fuzzy Logic Decision Maker*, which take a decision based on the similarity index and inform the user accordingly. Moreover, the selection of words will be random upon each login attempt. Therefore, a prerecorded voice sample will be useless that prevent replay attack.

## ACKNOWLEDGEMENT

This research was supported by IITP grant funded by the Korean government (MSIT) (No. 2017-0-00655) and Ministry of Science and ICT, Korea, under the ITRC support program (IITP-2017-0-01629) supervised by the IITP and NRF-2016K1A3A7A03951968.

## REFERENCES

[1] Monica Anderson. 2017. Many smartphone owners don't take steps to secure their devices. https://www.pewresearch.org/fact-tank/2017/03/15/many-smartphone-owners-dont-take-steps-to-secure-their-devices/
[2] Logan Blue, Hadi Abdullah, Luis Vargas, and Patrick Traynor. 2018. 2ma: Verifying voice commands via two microphone authentication. In *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*. ACM, 89–100.
[3] Jeanette Rose, John Windle, Ryan Schuetzler, and Ann Fruhling. 2018. Evaluation of Voice Authentication for Patient Health Record Access. (2018).