VoxGrid
A Mobile Voice Verification System

Ranyel Bryan L. Maliwanag, University of the Philippines

“IT Security for the Next Generation”
International Round, Delft University of Technology
11-13 May, 2012
The Netherlands
User Authentication

Sign in

Username

Password

Sign in

Stay signed in

Can't access your account?

Facebook Login

Email address:

Password:

Keep me logged in

Log in or Sign up for Facebook

Forgotten your password?
User Authentication Types
Our Solution
Vox
Why Voice?

- Unique
- No special hardware needed
Grid
System Architecture

SPEAKER VERIFICATION
CLIENT-SERVER ARCHITECTURE

CLIENT
- USERNAME AND PASSWORD INPUT
  - NO
  - YES
    - INPUT FIVE DIGIT CODE
      - SPEECH RECOGNITION
        - YES
        - NO
        - ACCEPT LOG-IN
        - DENY ACCESS LOG-IN
      - ENCRIPTION
        - FEATURE EXTRACTION
          - CAPTURE INPUT SPEECH
  - DENY ACCESS LOG-IN

SERVER
- ESTABLISH CONNECTION
  - USER INFO
    - USER EXISTS YES / NO
  - FIVE DIGIT CODE
    - CODE MATCHES YES / NO
  - SEND ENCRYPTED F.V AND USER INFO
    - DECISION
      - PATTERN MATCHING
        - OBTAIN MODEL (USER INFO)
          - RETURN HYPOTHESIZED MODEL
  - OBTAIN BACKGROUND MODEL
  - SCORE NORMALIZATION
    - RETURN BACKGROUND MODEL

DATABASE
- CHECK IF USER EXISTS
  - EXISTS YES / NO
- CHECK IF CODE MATCHES
  - MATCHES YES / NO
VoxGrid
Enrollment Process

1. Input username and password
2. User info: Username and password exist in database
   - Yes: Proceed
   - No: Validate five digit code as second password
3. Capture input speech 3 times
4. Speech utterance feature vector sequence for 3 input voice samples of the speaker
5. Feature extraction
6. Encryption of features and five digit code using AES
7. Receive encrypted features and user info from client
8. Store models in the database with corresponding user info
9. Training of feature vectors
10. Send to server
11. Send registration result to client
Verification Process

1. **Input Username and Password**
   - If correct, go to **Username and Password in the Database**.
   - If incorrect, go to **Input 5-Digit Code for Speech Recognition**.

2. **Username and Password in the Database**
   - If matched, go to **5-Digit Code Matches in the Database**.
   - If not matched, go to **Encrypt Features Using AES**.

3. **Encrypt Features Using AES**
   - Go to **Capture Speech Utterance**.

4. **Capture Speech Utterance**
   - Go to **Send to Server**.

5. **Send to Server**
   - Go to **Decryption**.

6. **Decryption**
   - If code matches, go to **Obtain BG Models in Database**.
   - If code doesn’t match, go to **Server Sends a Deny Access Message**.

7. **Obtain BG Models in Database**
   - Go to **Speech Recognition Result**.

8. **Speech Recognition Result**
   - If score normalizes, go to **Decision**.
   - If not normalized, go to **Threshold**.

9. **Decision**
   - If above threshold, go to **Verification Result**.
   - If below threshold, go to **Server Sends Result to Client**.

10. **Verification Result**
    - If accepted, go to **Client Accepts or Denies Log-in Attempt**.
    - If denied, go to **Client Denies Log-in Access**.

11. **Client Denies Log-in Access**
    - Go to **Server Sends Result to Client**.

12. **Server Sends Result to Client**
    - End process.

13. **Server Sends Client an Access to Speaker Verification**
    - End process.
Results: Client Model (FRR)
Noisy Environment

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Accept</th>
<th>Reject</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker 1</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>Speaker 2</td>
<td>90%</td>
<td>10%</td>
</tr>
<tr>
<td>Speaker 3</td>
<td>80%</td>
<td>20%</td>
</tr>
<tr>
<td>Speaker 4</td>
<td>70%</td>
<td>30%</td>
</tr>
<tr>
<td>Speaker 5</td>
<td>60%</td>
<td>40%</td>
</tr>
<tr>
<td>Speaker 6</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Speaker 7</td>
<td>40%</td>
<td>60%</td>
</tr>
<tr>
<td>Speaker 8</td>
<td>30%</td>
<td>70%</td>
</tr>
<tr>
<td>Speaker 9</td>
<td>20%</td>
<td>80%</td>
</tr>
<tr>
<td>Speaker 10</td>
<td>10%</td>
<td>90%</td>
</tr>
<tr>
<td>Speaker 11</td>
<td>0%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Results: Impostor Model (FAR)
Noisy Environment

Speaker 1  Speaker 2  Speaker 3  Speaker 4  Speaker 5  Speaker 6  Speaker 7  Speaker 8  Speaker 9  Speaker 10  Speaker 11

- Reject
- Accept
Analysis of Results

- **False Rejection Rate**

\[
FRR(T) = \frac{N_i(T)}{N_{ic}(T)} \times 100
\]

\[
FRR(0.85) = \frac{25}{110} \times 100 = 22.72\%
\]

- **False Acceptance Rate**

\[
FAR(T) = \frac{N_c(T)}{N_{ct}(T)} \times 100
\]

\[
FAR(0.85) = \frac{8}{110} \times 100 = 7.27\%
\]
Analysis of Results

Total Success Rate:

\[
TSR = 100\% - \frac{N_i(T) + N_c(T)}{Total\,\,Accesses} \times 100
\]

\[
TSR = 100\% - \frac{28 + 5}{220} \times 100 = 85\%
\]
Limitations

- Varying microphones on different devices in different environments
- Recorded voices *
- Sickness
Conclusions and Future Work

We are able to achieve an 85% Total Success Rate given the data set we have which consists of eleven speakers.

Future work:

- Results analysis on increasing number of users in the system
- Possible Campus-wide Computerised Registration System Integration
- Cross-platform implementation
Thank You

Ranyel Bryan L. Maliwanag, University of the Philippines

“IT Security for the Next Generation”
International Round, Delft University of Technology
11-13 May, 2012
The Netherlands