LipPass: Lip Reading-based User Authentication on Smartphones Leveraging Acoustic Signals

Presenter: Li Lu

Li Lu[1], Jiadi Yu[1], Yingying Chen[2], Hongbo Liu[3], Yanmin Zhu[1], Yunfei Liu[1], Minglu Li[1]

Shanghai Jiao Tong University[1]

Rutgers University[2]

Indiana University-Purdue University Indianapolis[3]

Increasing Security Concerns of Mobile Devices

- Mobile Device
  - Pervasive and common
  - Frequent storage medium for sensitive information
    - ID number, CVS code of credit cards
- Concern about privacy leakage in mobile devices
  - 78% users worry about losing sensitive data on their personal devices (Symantec[1])
- User Authentication
  - First guard for privacy on mobile devices
  - Direct and efficient

Existing Authentication Mechanisms

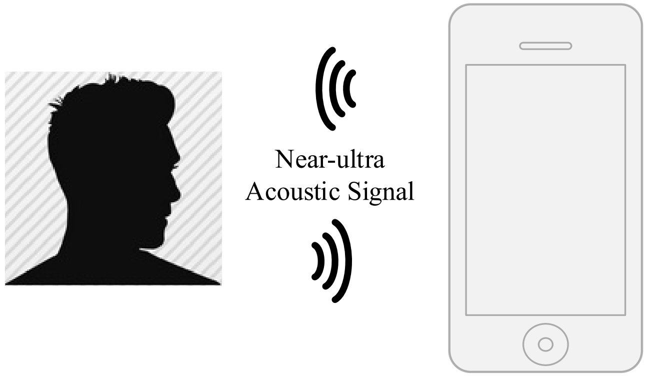
- Password
  - Most widely deployed
  - But hard to remember & vulnerable to stealing attacks
- Biometric-based approaches
  - Fingerprint, Face recognition, Voiceprint
  - Based on physiological characteristics →
    - Vulnerable to replay attacks
    - Susceptible to ambient environments (e.g., lights & noises)
- To deal with the weakness,
  - Behavioral characteristic-based authentication

Outline

- Preliminary
- System Design
- Evaluation
- Conclusion

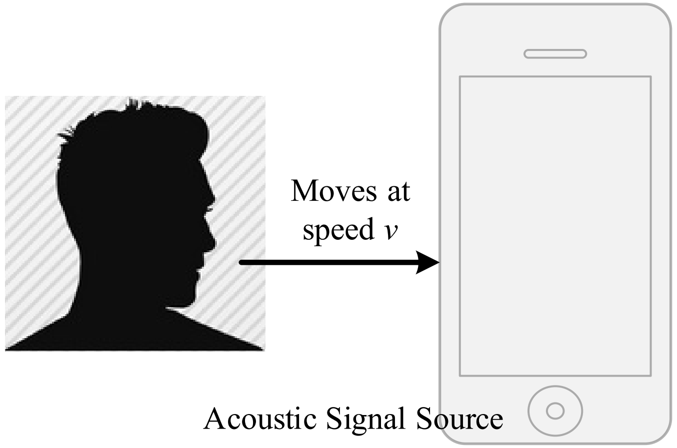
Motivation

- When a user speaks
  - Lip movements
  - Different users → different lip movements
- Capturing lip movements
  - Utilizing audio devices on smartphones
  - Emitting acoustic signal by the speaker, and receiving reflected signal through the microphones
  - Lip movements → Doppler effect of acoustic signals



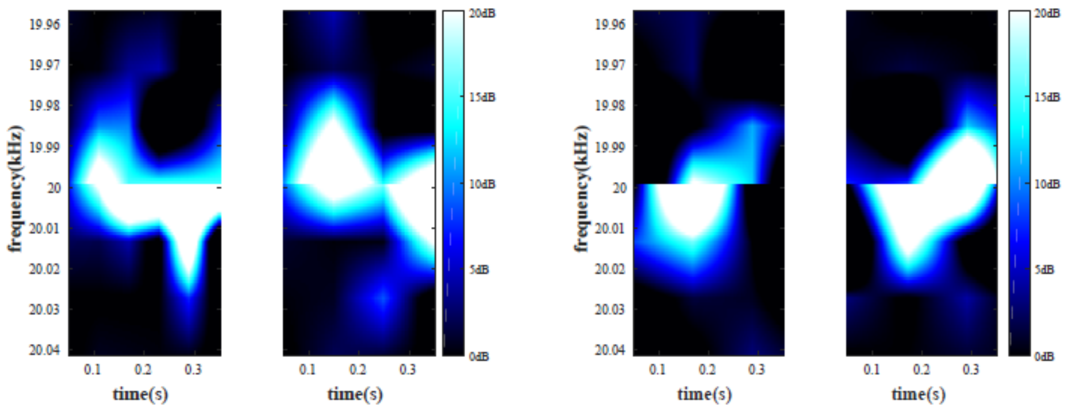
Doppler Effect

- An object moving (at speed *v*) relative to acoustic signal source brings a frequency change
  - $\Delta f = \frac{v}{c} \times f_0$, where c and f_0 are speed and frequency of acoustic signals respectively
- Audio device setting
  - f_0 =20kHz, sampling rate: 44.1kHz
- Time-domain → Frequency-domain
  - 2048-point FFT



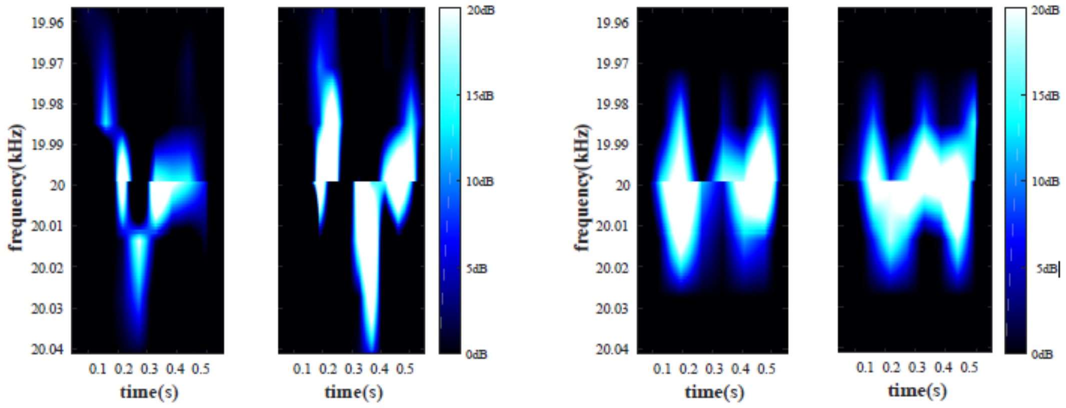
Difference in Doppler Profiles

- When speaking the same passphrase
  - Doppler profiles of different users are significantly different
  - Doppler profiles of the same user are similar
- Doppler profiles caused by lip movements → User authentication



(a) User 1 (b) User 2

Hello



(a) User 1 (b) User 2

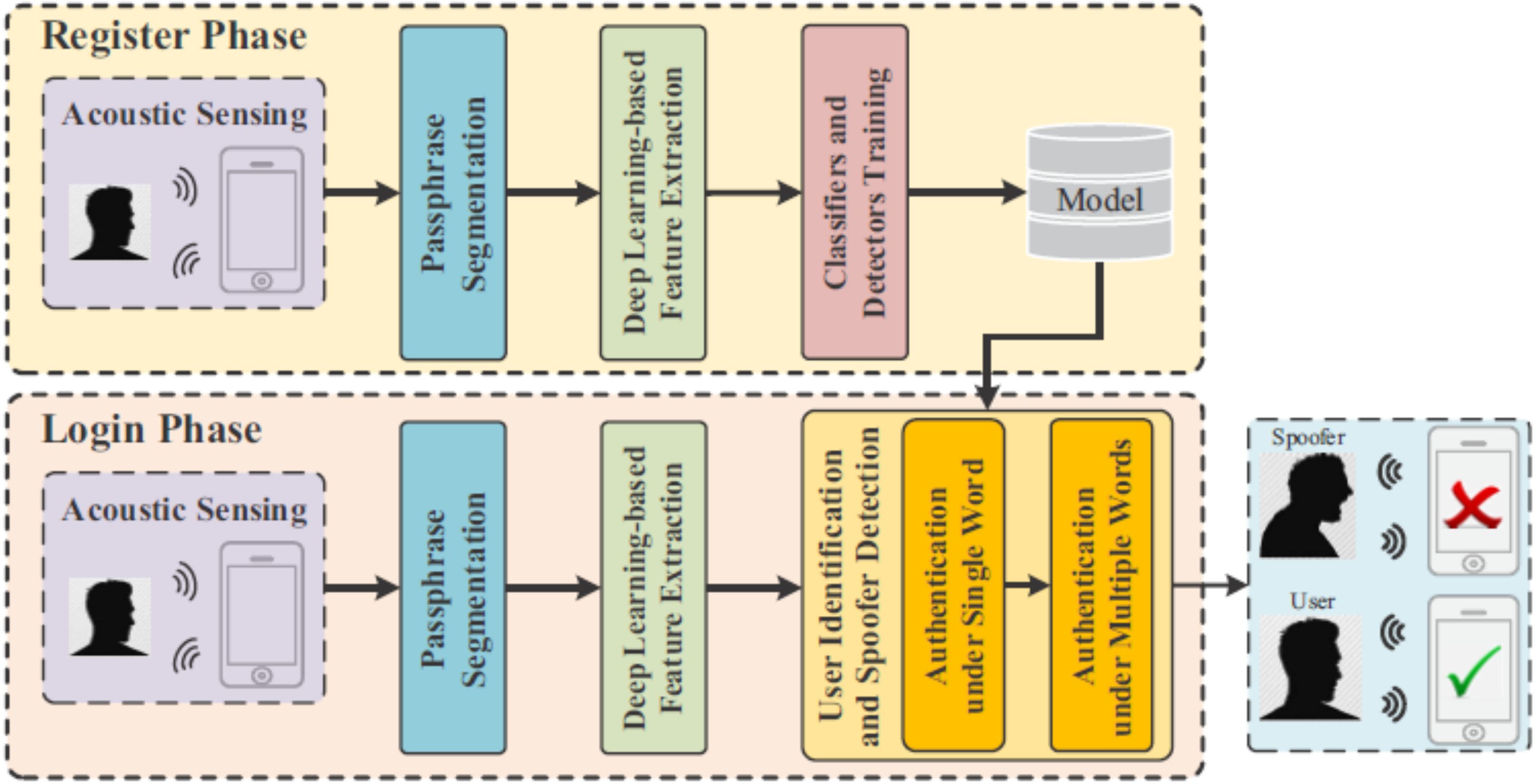
World

Outline

- Preliminary
- System Design
- Evaluation
- Conclusion

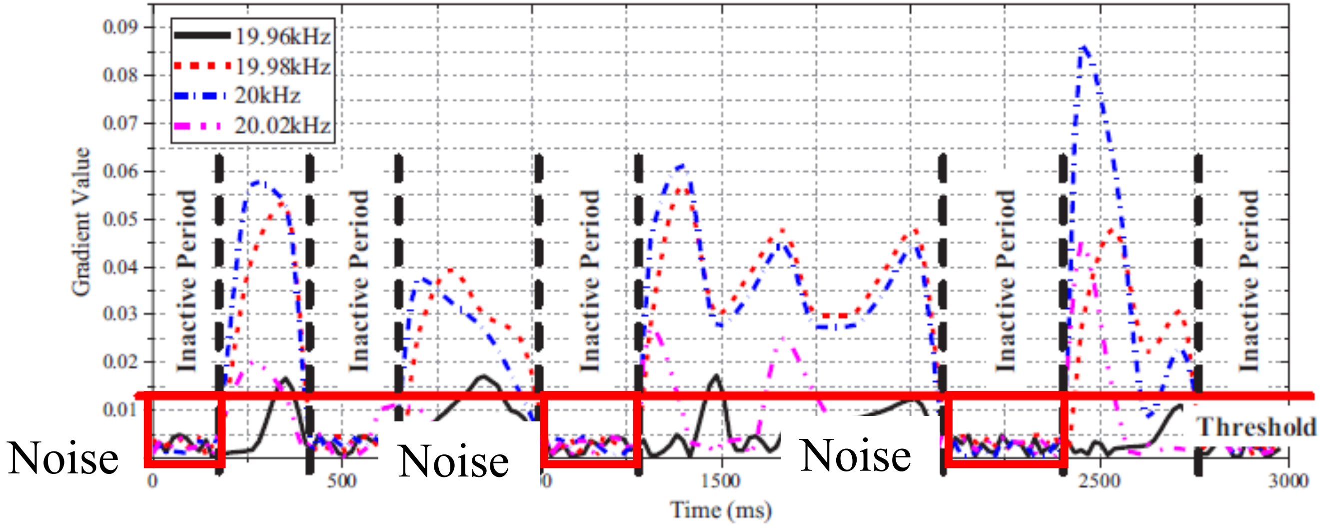
Overview

- **Two Phases:**
  - Register & Login
- **Four Processes:**
  - Passphrase Segmentation
  - Deep Learning-based Feature Extraction
  - Classifiers and Detectors Training
  - User Identification and Spoofer Detection



Passphrase Segmentation

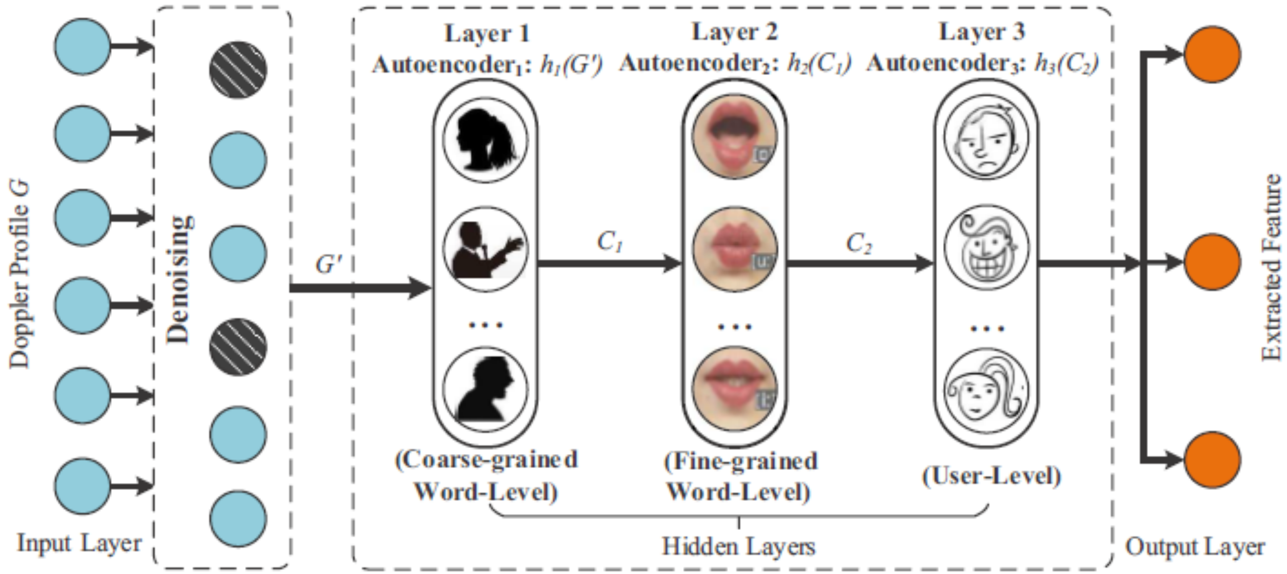
- A passphrase → several words
  - There is usually a short interval between two successive words
- Speaking words vs. Intervals between words
  - Speaking → significant Doppler effect caused by lip movements
  - Interval → only white noises
- Threshold-based approach



Deep Learning-based Feature Extraction

- From acoustic signal episode of each word
  - Extract efficient and reliable features
- Three-layer autoencoder-based Deep Neural Network
  - Non-linear feature extraction
  - Abstract compressed representations through unsupervised manner

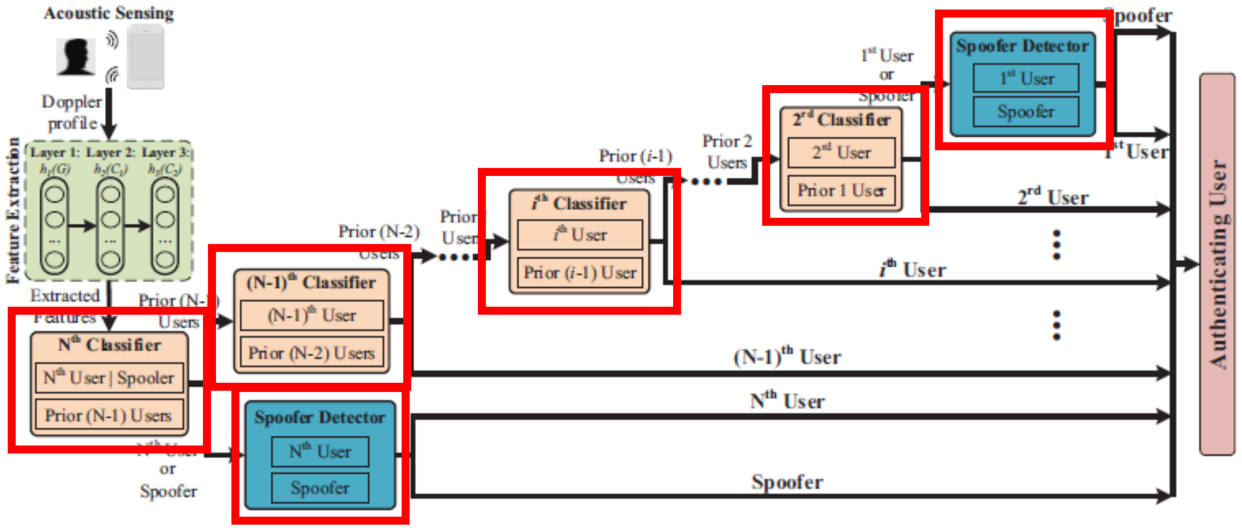
- 1^{st} Layer: coarse-grained word-level
- 2^{nd} Layer: fine-grained word-level (e.g., phoneme level)
- 3^{rd} Layer: user-level



Classifiers and Detectors Training

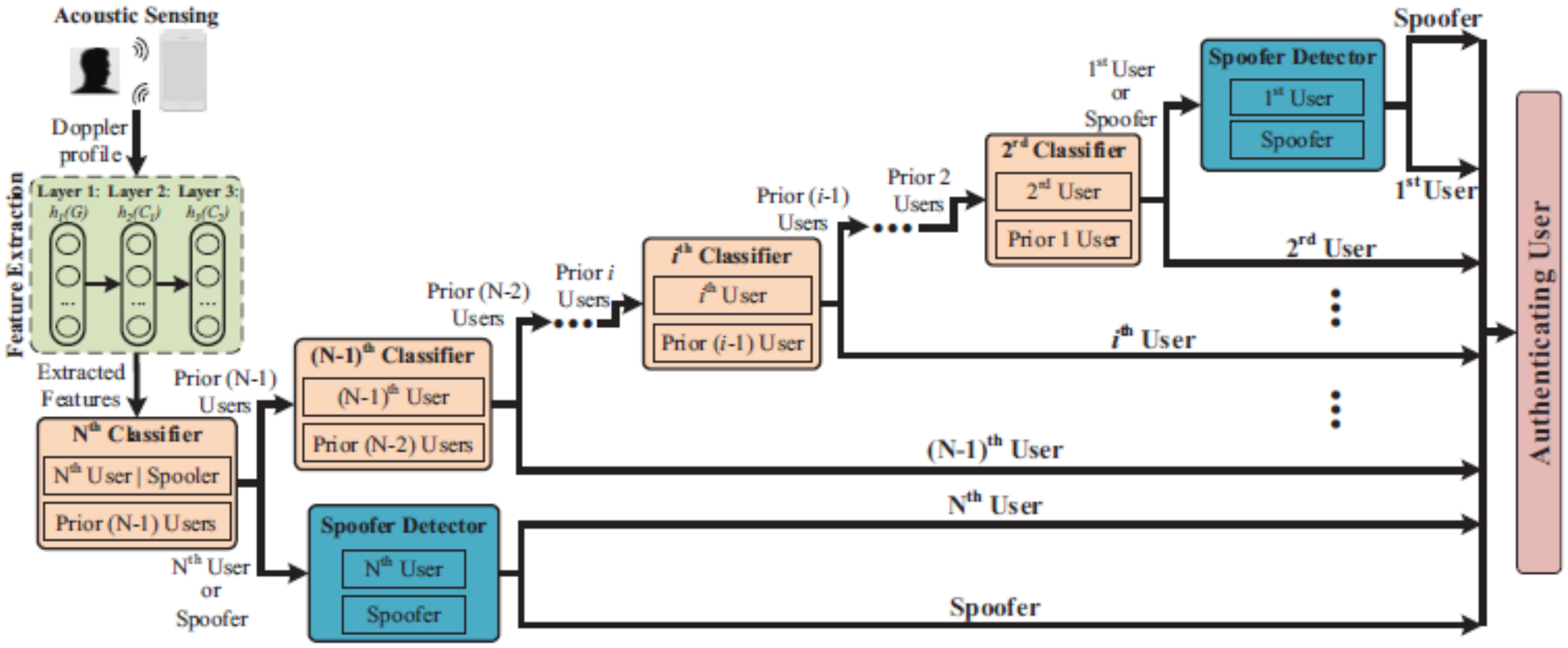
- Multi-user Classifier & Spoofer Detector Training
  - SVM (Support Vector Machine) & SVDD (Support Vector Domain Description)
- Users register to the system sequentially
  - Reconstruct a classifier whenever a new user registers → significant computational complexity
  - Multiple binary classifiers training

- Assume i^{th} user registers to the system
  - Train a binary classifier through one-versus-rest manner (i.e., i^{th} user & other i-1 users)
  - Train a spoofer detector through SVDD (i.e., i^{th} user & spoofers)



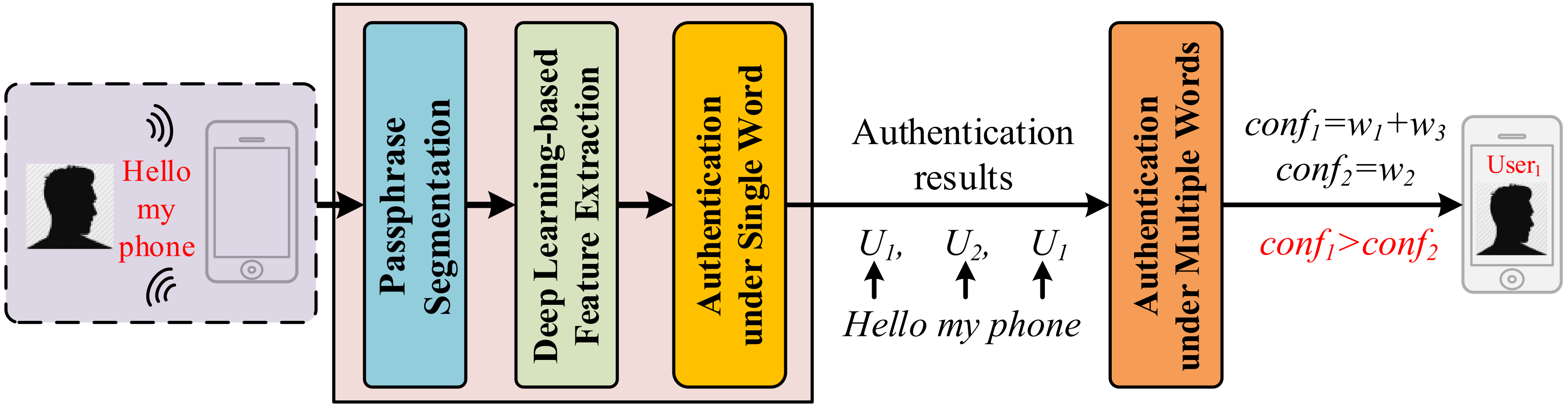
User Identification & Spoofer Detection

- Authentication under single word
  - Binary tree-based authentication



User Identification & Spoofer Detection (Con.)

- Authentication under multiple words
  - Strengthen robustness of authentication result
- An example ($User_1$ & $User_2$ register to the system):
  - A user speaks ‘Hello my phone’ to login
  - Three labels (i.e., U_1, U_2, U_1) can be obtained through the approach above
  - Calculate two confidences for two users (i.e., $conf_1>conf_2$)
  - The user is identified as $User_1$



Outline

- Preliminary
- System Design
- Evaluation
- Conclusion

Experiment Setup

- 48 volunteers in 4 real environments respectively
  - Volunteers: 24 males and 24 females, whose ages range in [18,52]
  - Environments: lab (bright and quiet),
    station (bright but noisy),
    dark lab (quiet but dark),
    pub (dark and noisy).
- 10 passphrases:
  - Each of them contains 1-10 words
  - Each word contains >4 phonemes

Lab

Station

Dark Lab

Pub

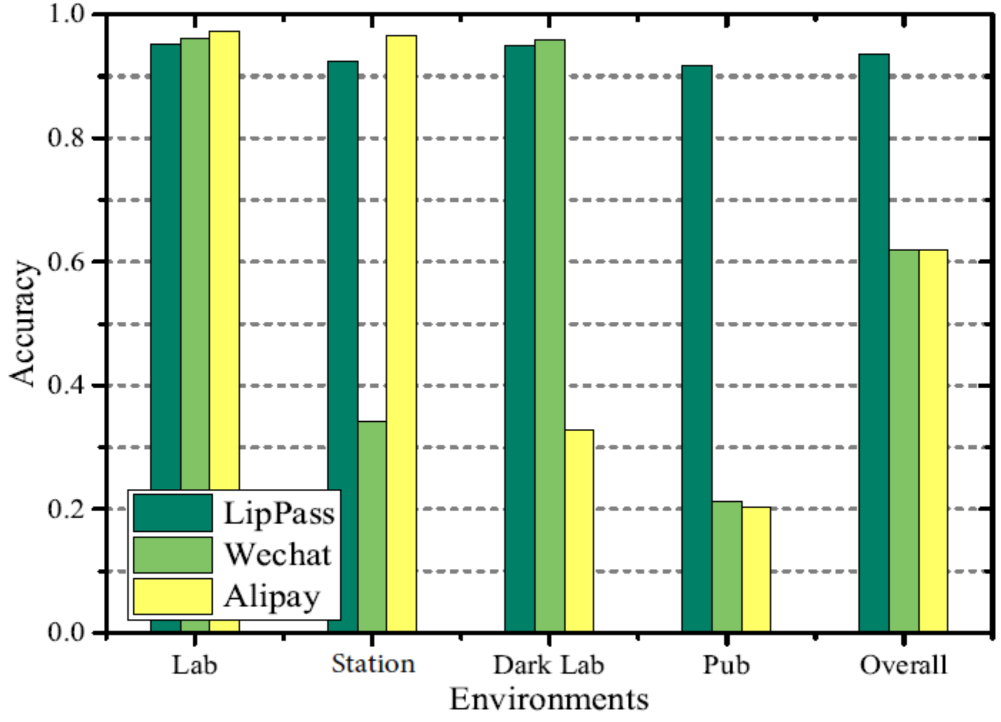
Overall Performance

- Achieve over 80% accuracy in identifying registered users
- Average 90.2% accuracy in user authentication
- Average 93.1% accuracy in spoofer detection

Ground Truth \ Authentication Results	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8	U_9	U_{10}	Spoofer
U_1	0.837	0.033	0.006	0.024	0.029	0.005	0.050	0.000	0.010	0.000	0.006
U_2	0.020	0.857	0.024	0.030	0.031	0.000	0.010	0.006	0.013	0.006	0.003
U_3	0.010	0.012	0.871	0.024	0.010	0.006	0.010	0.047	0.000	0.004	0.006
U_4	0.024	0.000	0.010	0.925	0.000	0.006	0.012	0.000	0.003	0.010	0.010
U_5	0.006	0.000	0.010	0.000	0.933	0.020	0.000	0.009	0.010	0.000	0.012
U_6	0.020	0.006	0.000	0.010	0.010	0.930	0.000	0.018	0.000	0.000	0.006
U_7	0.020	0.006	0.000	0.000	0.010	0.030	0.900	0.000	0.010	0.012	0.012
U_8	0.011	0.012	0.020	0.010	0.010	0.006	0.006	0.910	0.000	0.006	0.009
U_9	0.012	0.010	0.010	0.010	0.006	0.000	0.002	0.006	0.938	0.000	0.006
U_{10}	0.020	0.010	0.020	0.000	0.010	0.000	0.000	0.010	0.000	0.920	0.010
Spoofer	0.016	0.010	0.012	0.006	0.010	0.000	0.003	0.006	0.000	0.006	0.931

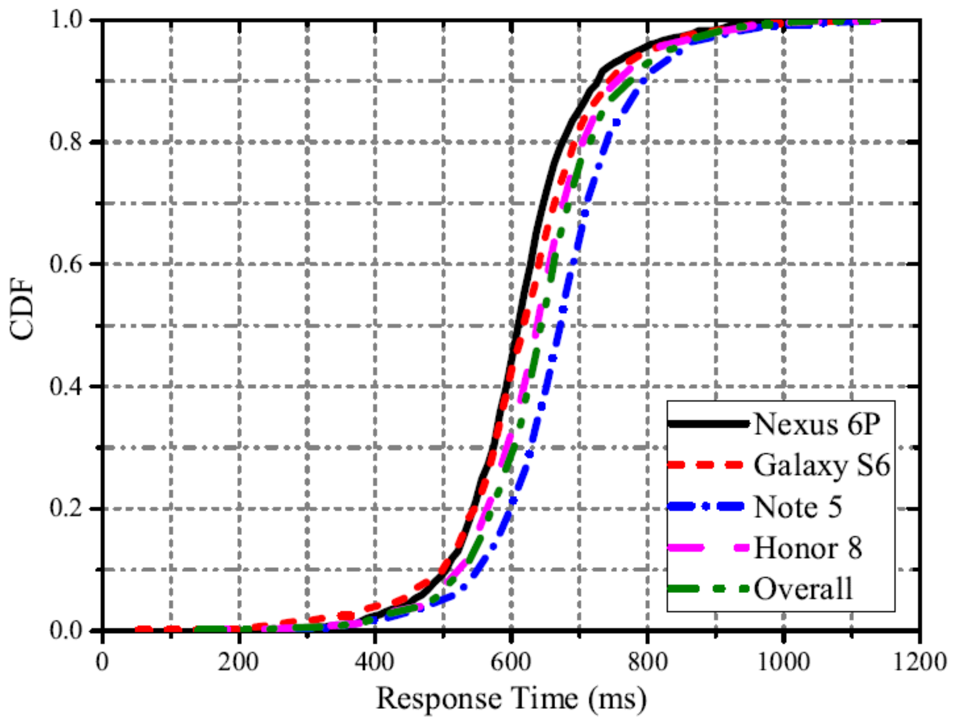
Comparison with other Authentication System

- Ideal environment (Lab)
  - LipPass: 95.3% vs. Wechat: 96.1% & Alipay: 97.2% (similar performance)
- Noisy environment (Station)
  - LipPass: 92.4% vs. Wechat: 34.3% (significantly better than Wechat)
- Dark environment (Dark Lab)
  - LipPass: 94.9% vs. Alipay: 32.9% (significantly better than Alipay)
- Worst environment (Pub)
  - LipPass: 91.7% vs. Wechat: 21.3% & Alipay: 20.4% (better than other two approaches)



Response Time

- Response time = Login Time – End Speaking Time
- CDF of response time
  - 90% of volunteers are with less than 0.8s
  - Average response time: 0.64s
- LipPass is responsive



Outline

- Preliminary
- System Design
- Evaluation
- Conclusion

Conclusion

- Observation:
  - reveal the feasibility of utilizing Doppler profiles induced by lip movements for user authentication
- Contribution:
  - Propose a lip reading-based user authentication system
  - Design a deep learning-based method to abstract high-level behavioral characteristics of lip movements
  - Develop a binary tree-based authentication approach to identify each individual
- Evaluation: evaluate performances of *LipPass* in four real environments
  - Achieve 90.2% accuracy in user authentication
  - Achieve 93.1% accuracy in spoofer detection

Thank you!

Q & A