DAISY

Data Analysis and Information SecurityY Lab

WearID: Low-Effort Wearable-Assisted Authentication of Voice Commands via Cross-Domain Comparison without Training

Presenter: Cong Shi

Cong Shi*, Yan Wang†, Yingying Chen*, Nitesh Saxena‡, Chen Wang*

*WINLAB, Rutgers University, NJ, US

† Temple University, PA, US

‡University of Alabama at Birmingham, AL, US

ACSAC 2020

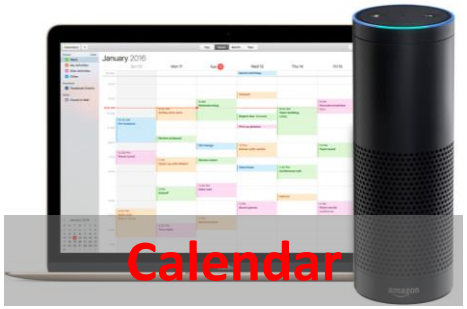
RUTGERS WINLAB

TEMPLE UNIVERSITY

UAB

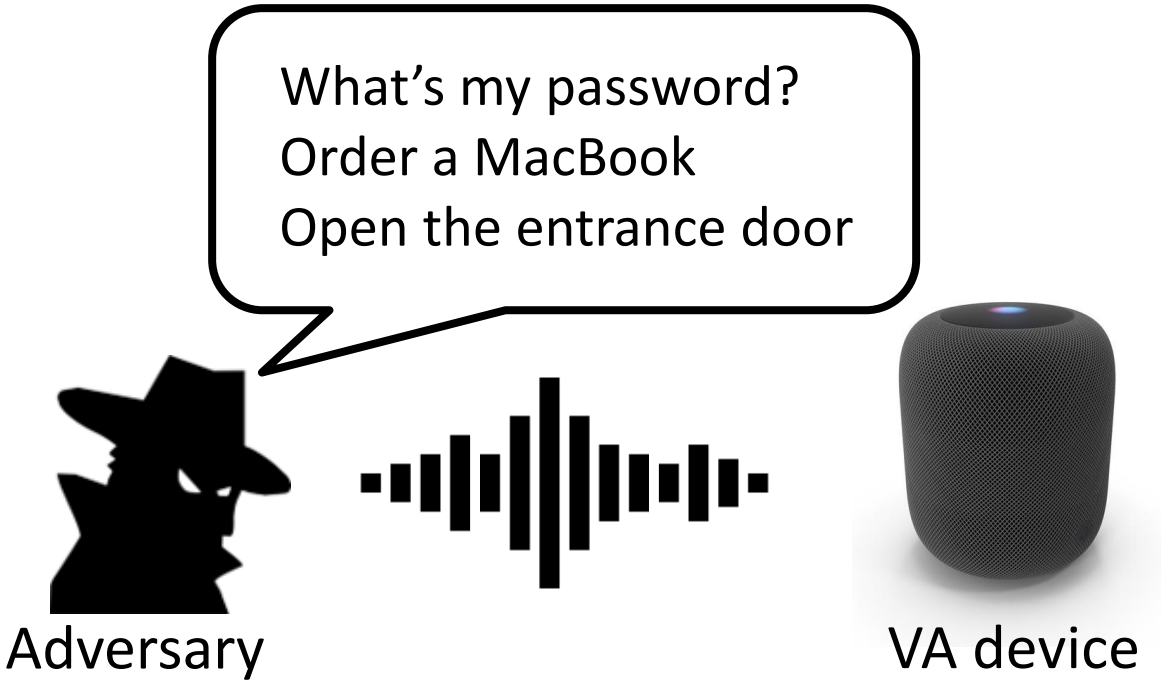
Voice Assistant System

- Speech recognition technologies enable smart and IoT devices to understand natural language and take voice commands
- Voice assistant (VA) systems facilitate numerous daily tasks



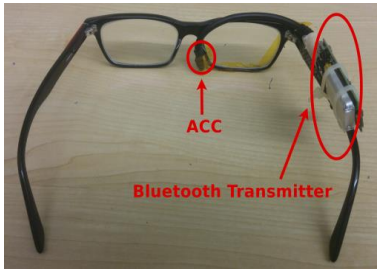
Highly Critical Voice Commands

- Growing trend of using critical voice commands to access sensitive information and functionalities
- Lure adversaries into faking the user's voice commands



Existing Solutions

- Voiceprint-based technologies
  - Rely on acoustic features
  - Prone to acoustic attacks (e.g., replay attacks, hidden voice commands)
- Two-factor authentication
  - Audio CAPTCHA, replay calls/messages, or virtual buttons
  - Require significant user efforts and prone to carless behaviors
- Solutions using dedicated/specialized devices
  - Use multiple microphones or high-sampling rate accelerometers [1]
  - Lead to additional cost and energy consumption



[1] Huan Feng, Kassem Fawaz, and Kang G. Shin. "Continuous authentication for voice assistants." In Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking, pp. 343-355. 2017.

Attack Model

- Attacks on user's absence (i.e., audible attacks)
  - *Random attacks*: use the adversary's own voice
  - *Impersonation attacks*: exploit speech synthesis techniques to produce voice commands
  - *Replay attacks*: replay pre-recorded voice commands of the legitimate user

- Co-location attacks (i.e., inaudible attacks)
  - *Hidden voice command attacks*: encode pre-recorded voice commands as background noises
  - *Ultrasound attacks*: modulate pre-recorded voice commands into ultrasound frequency bands

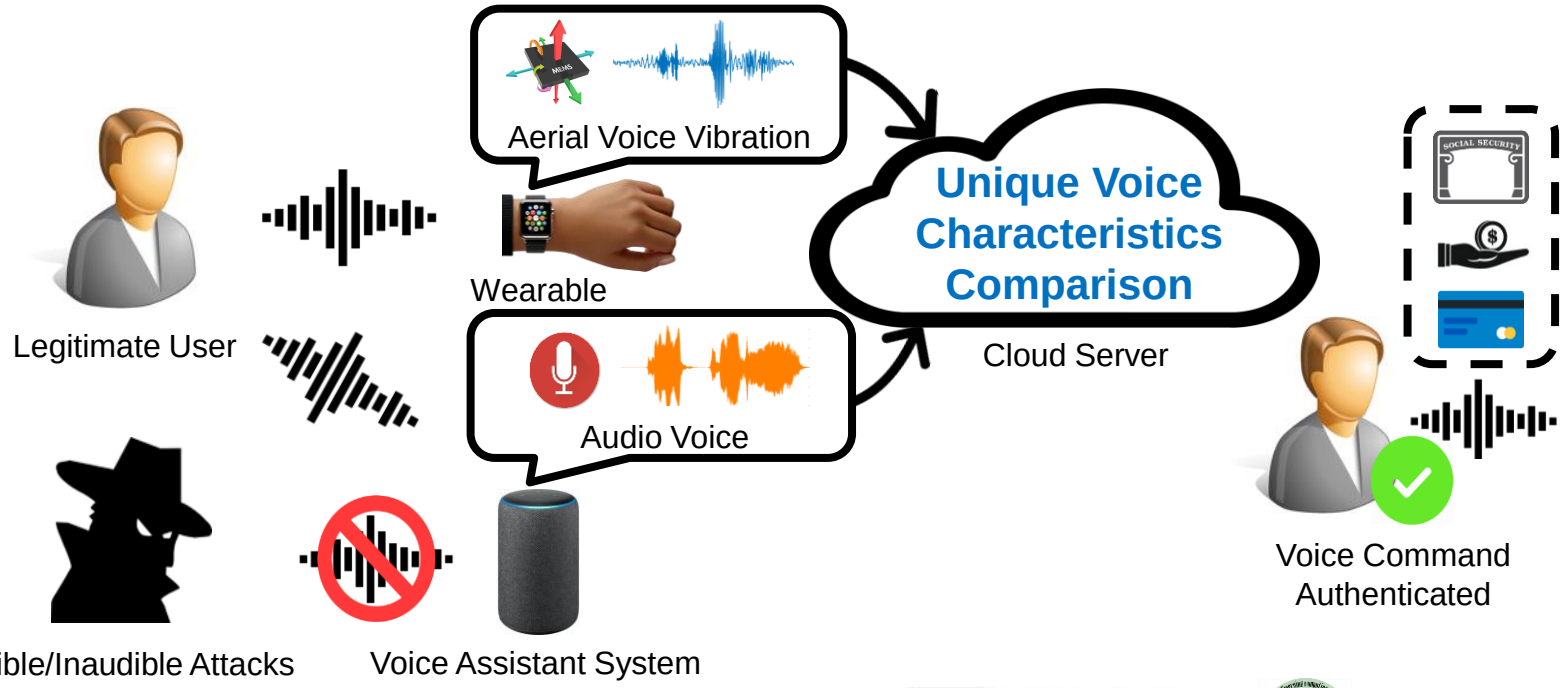
Our Idea: Capturing Aerial Speech Vibration

- Explore the wearable's motion sensor to harness aerial speech vibrations corresponding to live human speeches
- The unique response and short response distance of the motion sensor prevents both audible and inaudible attacks

Our Goal: Developing a low-effort training-free voice authentication system leveraging aerial speech vibrations

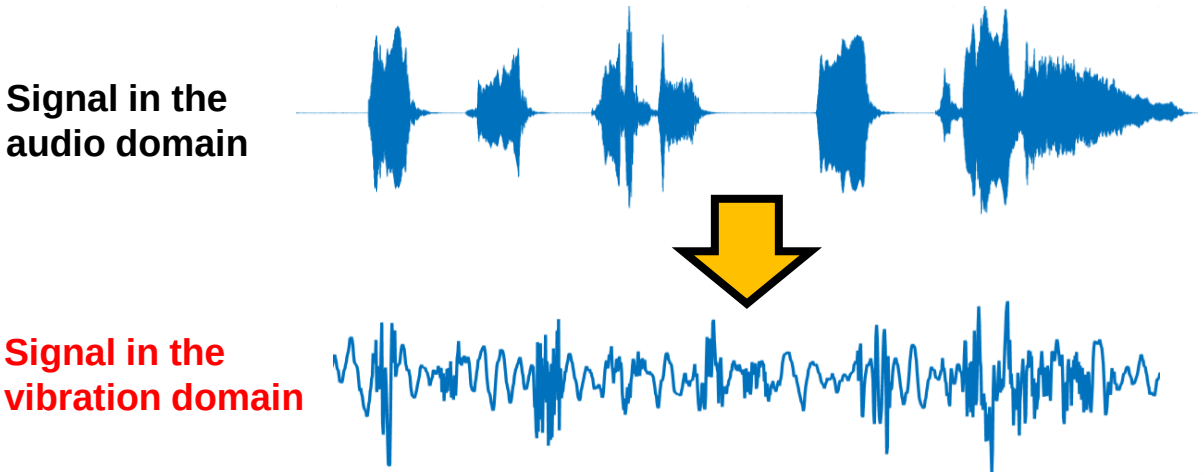
Our Contributions

- Propose **a cross-domain user authentication system**
  - Compare the aerial speech vibration (i.e., in the *vibration domain*) and the audio speech (i.e., in the *audio domain*) on the VA system's cloud
  - Do not require any hardware modifications
  - Do not require privacy-sensitive voice templates



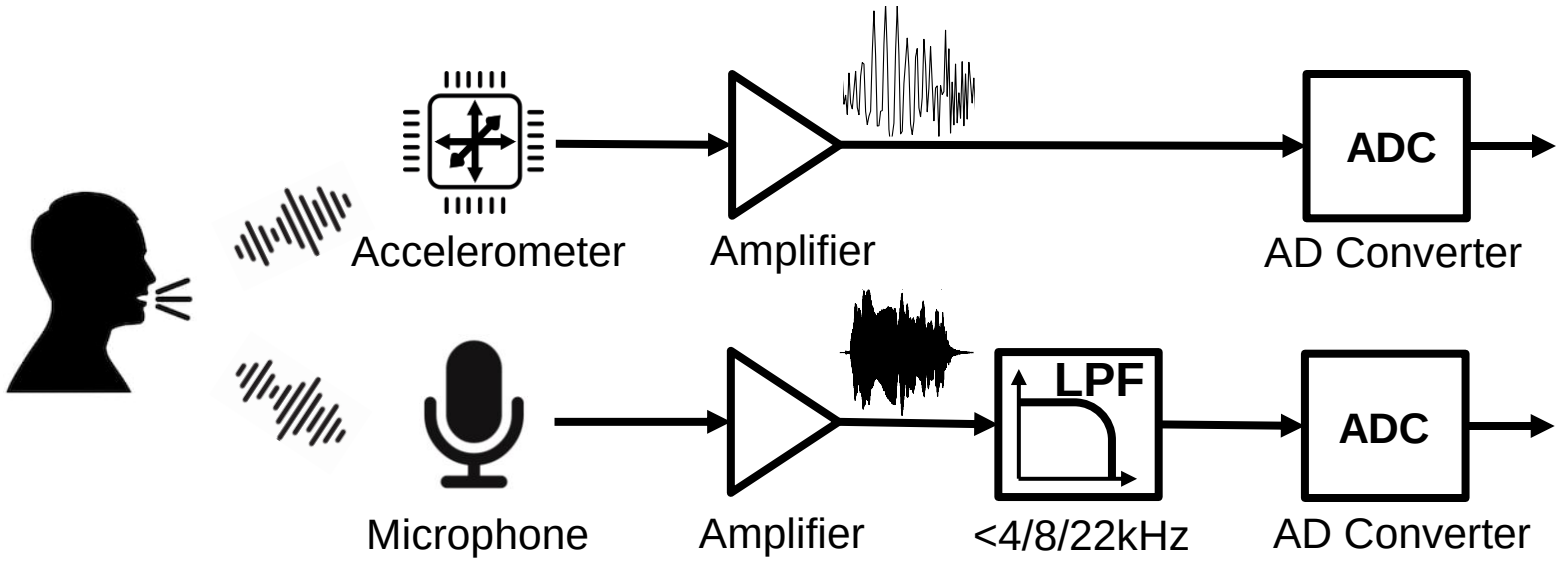
Our Contributions

- Provide **enhanced security**
  - Leverage the unique vibration domain responses (e.g., short response distance of accelerometers) to prevent audio domain attacks
- Model the **cross-domain relationship**
  - Derive the unique spectral relationship between the audio and vibration domains
  - Convert audio signals to low-frequency vibration signals



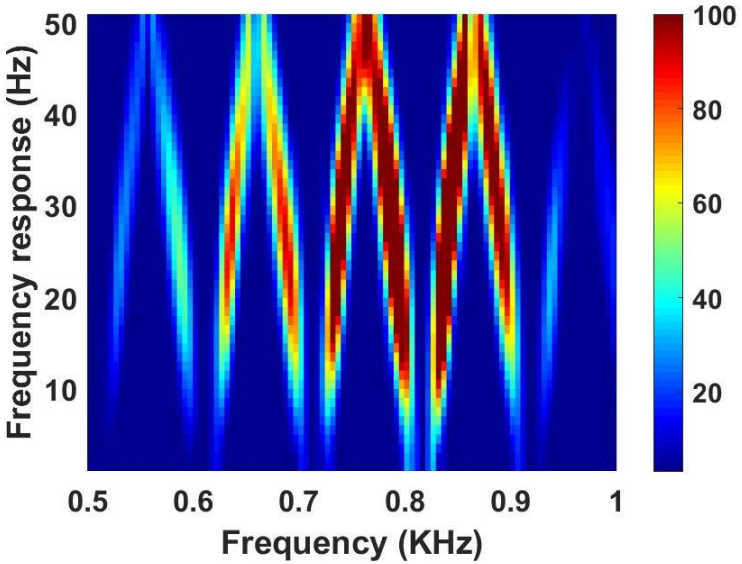
Accelerometer vs. Microphone

- Microphone exploits a pressure-sensitive diagram to capture sound and utilizes a Low Pass Filter (LPF) for denoising
- Accelerometer measures sound in terms of the vibration of the inertial mass and can capture up to $4kHz$ vibration signals

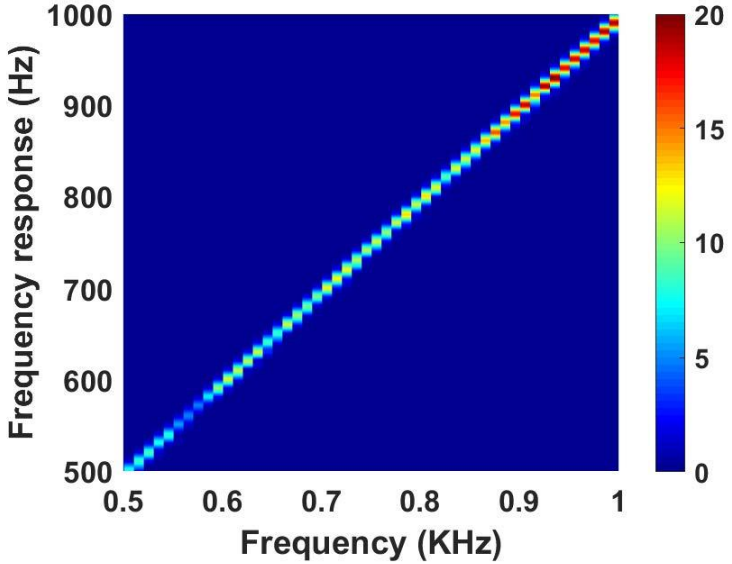


Response in Vibration Domain

- Vendors of wearables limit the sampling rate of accelerometers to below $200Hz$ for saving energy
  - Lead to signal aliasing : $f_{alias} = |f - Nf_s|, N \in Z$
  - Result in unique frequency selectivity of the wearable



Accelerometer's response to a chirp signal ($0.5kHz$~$1kHz$)

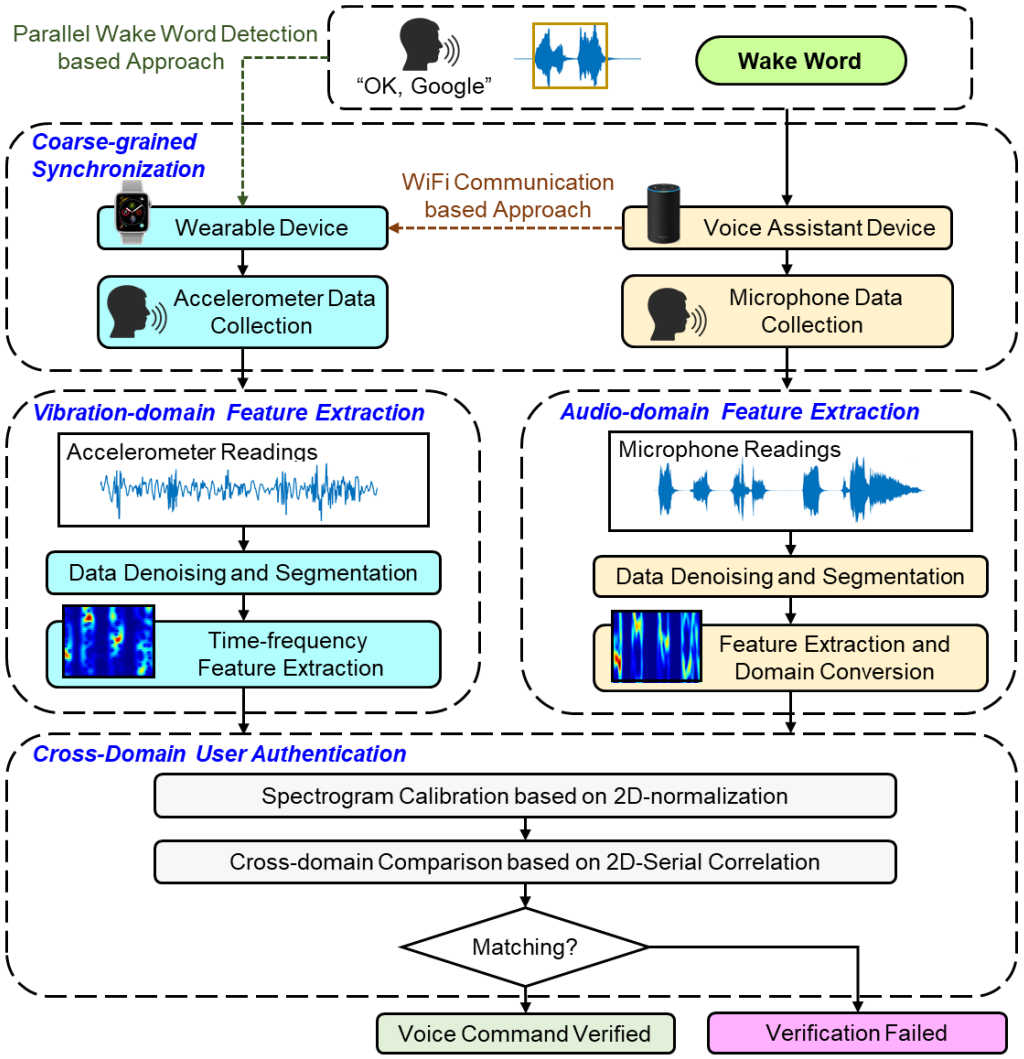


Microphone's response to a chirp signal ($0.5kHz$~$1kHz$)

Challenges

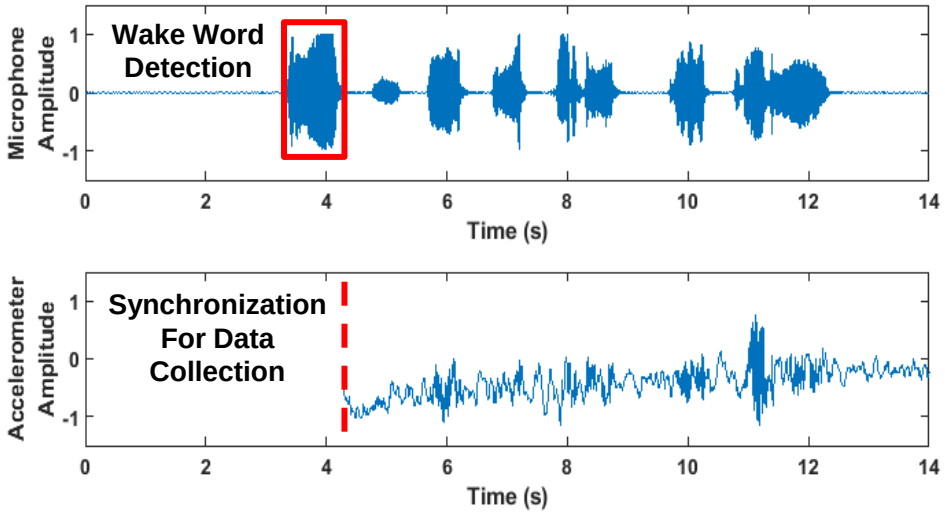
- The weak response of wearable's accelerometer to human voice make it difficult to extract aerial speech vibrations
- The heterogeneous hardware designs and huge sampling rate gap make any direct comparison infeasible
- Synchronization of the data collected in totally different hardware is difficult

System Overview



Synchronization and Data Preprocess

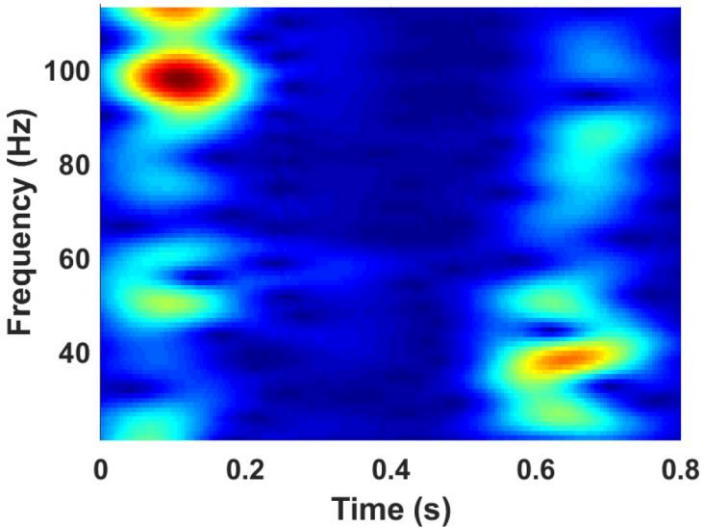
- Coarse-grained synchronization
  - Based on WiFi: VA device sends a triggering message to synchronize the data collection process on the wearable
  - Based on the wearable's accelerometer: detect the wake word in parallel with the VA device to trigger data collection
- Data processing: Apply a high-pass Butterworth filter with a cut-off frequency of $20Hz$ to remove human motion artifacts



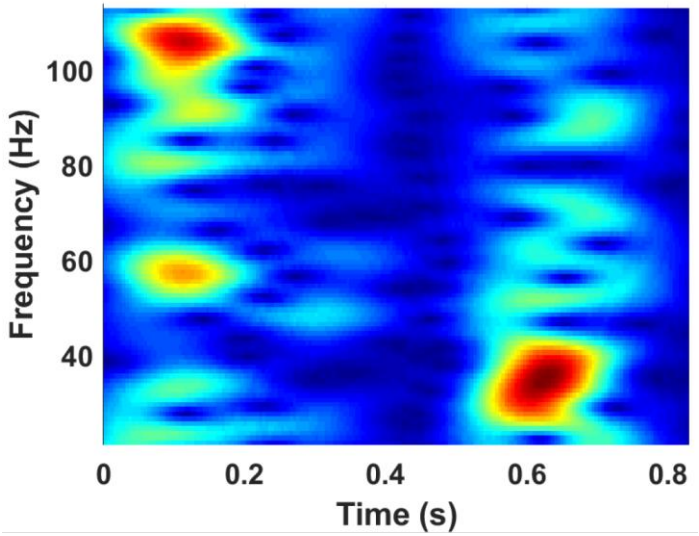
Feature Domain Conversion

- Explore Short Time Fourier Transform representations (i.e., spectrogram) as features for audio and vibration domains
- Convert audio spectrogram into vibration domain:

$$\hat{S}_{mic}(t, f_w) = \sum_{n=-inf}^{inf} S_{mic}(t, win(|f_m + n \times \omega|))$$



Spectrogram of vibration signals for "Alexa"

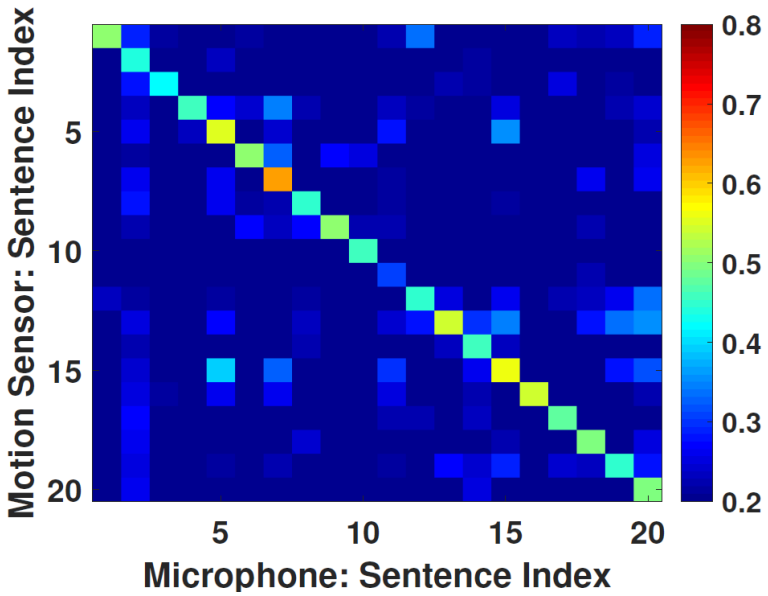


Converted spectrogram of audio signals for "Alexa"

Cross-Domain Similarity Comparison

- Explore 2D-normalization to resolve the scale differences
- Calculate cross-domain similarity: $Corr(\hat{S}_{mic}, S_{acc}) = \frac{A \times V}{\sqrt{A^2 \times V^2}}$

$$s.t., A = \sum_t \sum_f (\hat{S}_{mic}(t,f) - \mu), V = \sum_t \sum_f (S_{acc}(t,f) - \mu)$$



Cross-domain similarity among 20 voice commands

Experimental Setup

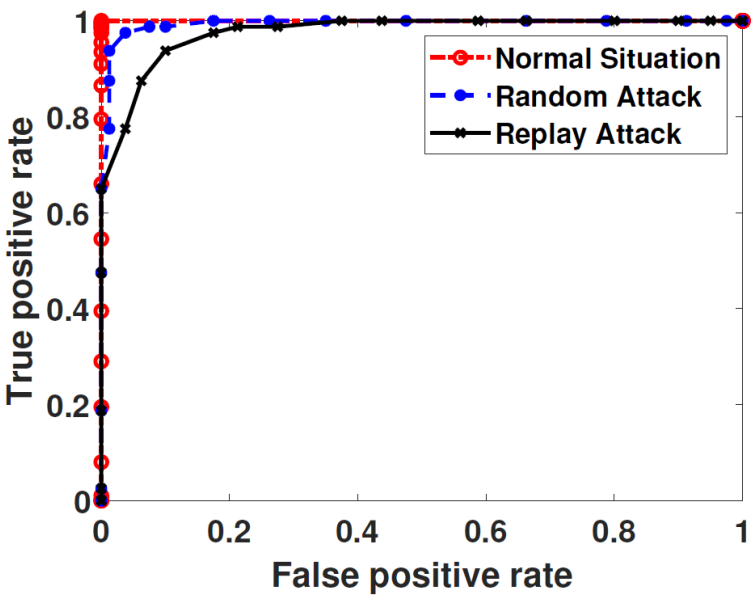
- **Smartwatch**
  - Huawei 2 sport (100Hz)
  - LG W150 (200Hz)

- **Setup**
  - Subject wears the smartwatch
  - Subject stands 1m to the VA device (simulated with a Nexus 6 smartphone)
  - $80dB$ sound pressure level

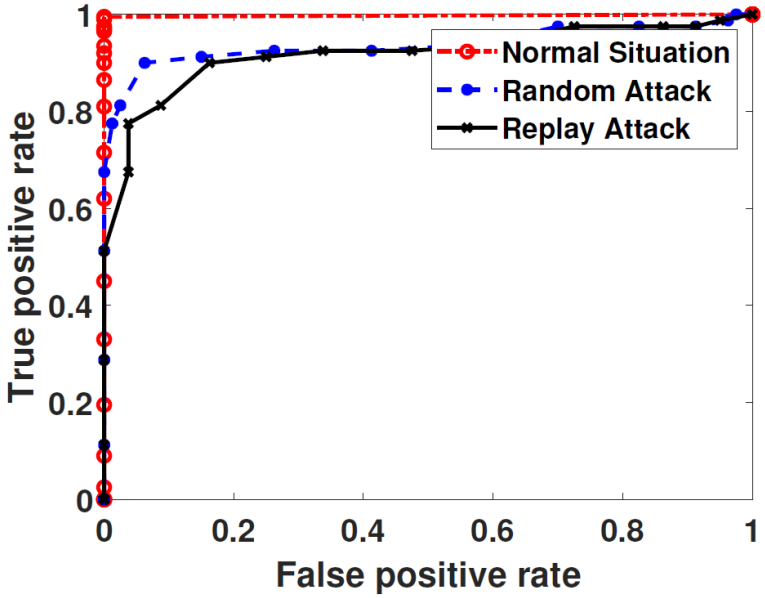
- **Data collection**
  - 10 participants
  - 20 critical voice commands
  - 10 hidden voice commands

Voice Command Examples
"What's on my calendar for tomorrow"
"Where is my next appointment"
"List all events for January 1st"
"How much is a round-trip flight to New York"
"Remember that my password is 'money'"
"What is my password"
"Add 'go to the grocery store' to my to-do list"
"What's on my shopping list"
"Track my order"
"Read me my email"
"Call my mother"

Defending Audible Attacks



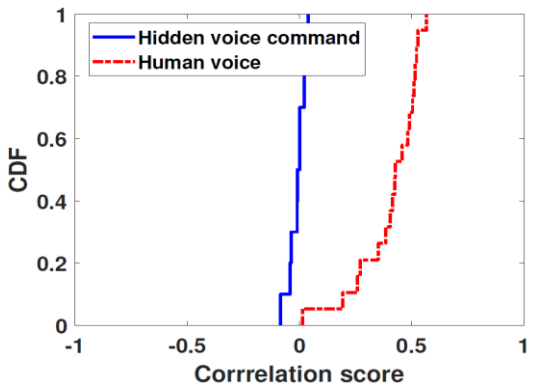
Huawei 2 sport



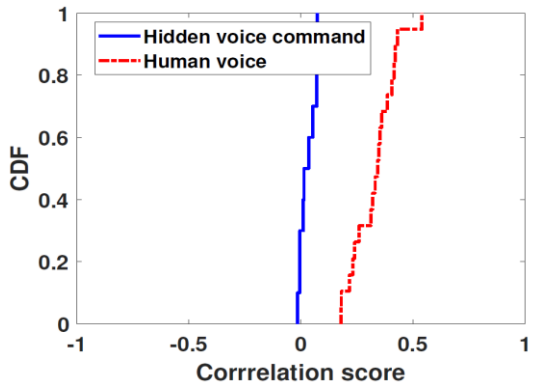
LG W150

- Over **99.6%** TPR and close to **0%** FPR for both watches
- Over **0.94 and 0.89** AUCs under random attacks for Huawei 2 sport and LG W150
- Over **0.91** and **0.88** AUCs under relay attacks for the two smartwatches

Defending Inaudible Attacks

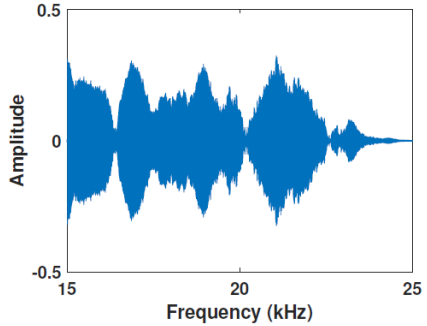


Huawei 2 sport

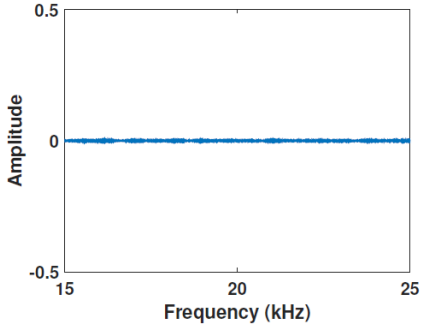


LG W150

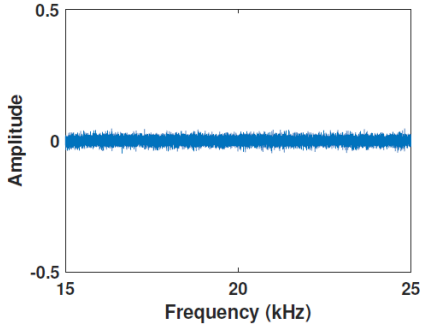
❖Cross-domain similarities are low for hidden voice commands



VA device



Huawei 2 sport



LG W150

❖The ultrasound chirp can not impact accelerometer readings

Conclusion

- Proposed WearID, a wearable-assisted low-effort training-free user authentication scheme for VA systems
- Explored wearable devices' motion sensor to harness aerial speech vibrations to verify highly critical commands
- Developed cross-domain comparison for training-free and privacy-preserving authentication
- Demonstrated the effectiveness and robustness of WearID against various attacks through extensive experiments

Limitations and Future Work

- **Improving accuracy and usability**
  - Using more sensitive wearables
  - Improving the authentication algorithms
- **Defending replay attack in vibration domain**
  - Exploring frequency-selective patterns of the accelerometer
- **Deployment feasibility**
  - Removing environmental factors

DAISY

Data Analysis and Information SecuritY Lab

Cong Shi

cs1421@scarletmail.rutgers.edu

http://winlab.rutgers.edu/~cs1421/