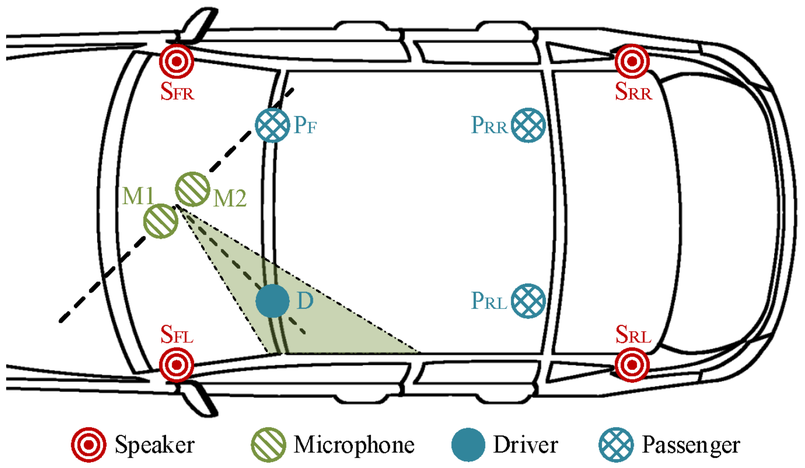
Safeguarding Automatic Speech Recognition Systems on Driverless Vehicles

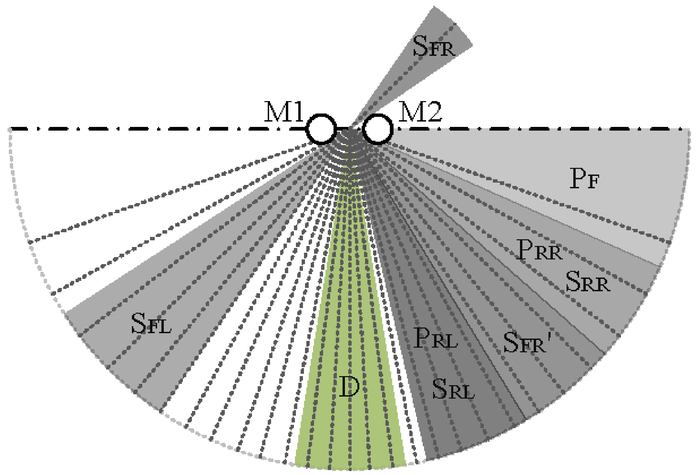
Shu Wang, Kun Sun
{swang47, ksun3}@gmu.edu
Center for Secure Information Systems, George Mason University

Motivation

- Automatic speech recognition (ASR) systems suffer from several adversarial voice attacks.
- Prior work [1] focus on the distinction of voice using the frequency and noise features.
- For driverless vehicle application, we want to find more robust features through the physical attributes of the received signals because of the fixed voice source locations.
- By using a pair of close-coupled microphones, we developed a secure ASR system (SASR) which contains three detection steps.



Location and Orientation of a Dual Microphone



Directions of Voice Sources to Microphones

Technique Approach

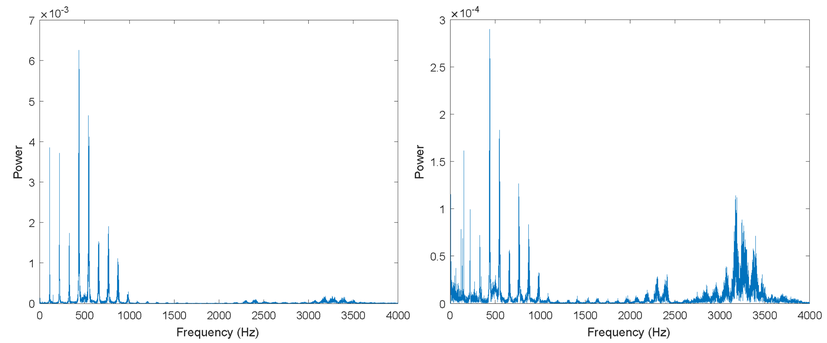
- **Detecting Multiple Speakers**

  Through autocorrelation analysis of linear prediction residual of the received voice, we filter out the multi-speaker signals.
- **Identifying Single Voice Source**

  We measure the voice propagation direction via time difference of arrivals (TDOA).
- **Detecting Voice from Mobile Phone**

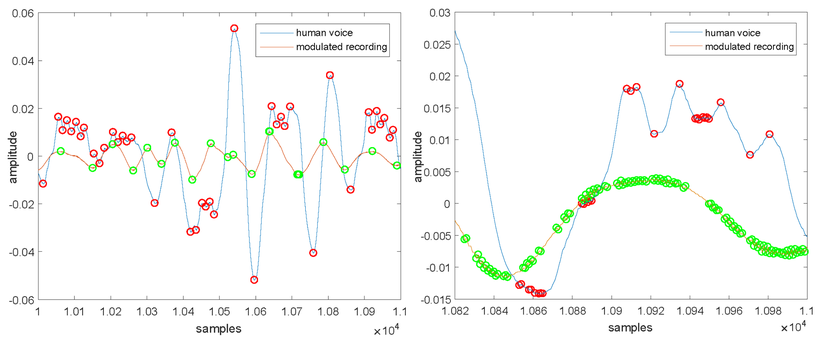
  We distinguish the replay attacks through frequency-domain power spectrum due to the low-frequency energy loss of mobile speakers [2].

  To prevent the modulated voice which can compensates the energy loss in frequency domain, we develop a time-domain double granularity extrema cross-check approach.



(a) Human voice (b) Relay voice

Frequency-domain Power Spectrum



(a) Coarse granularity (b) Fine granularity

Time-domain Double Granularity Extrema

Results

- **Detection Accuracy** for three detection steps is 83.3%, 96.8%, 97.6% respectively.
- **Performance Overhead**

Detection Step	Running Time	Memory
Multi-speaker Detection	134 ms	111 MB
Single Source Identification	33 ms	23 MB
Mobile Replay Detection	47 ms	10 MB
Total Overhead Costs	214 ms	144 MB

Artificial Intelligence and Cybersecurity Workshop
June 4-6, 2019 | University of Maryland

References

[1] M. Zhou, Z. Qin, X. Lin, S. Hu, Q. Wang, and K. Ren. 2019. *Hidden Voice Commands: Attacks and Defenses on the VCS of Autonomous Driving Cars*. IEEE Wireless Communications (2019), 1–6.

[2] Jesús Villalba and Eduardo Lleida. 2011. *Detecting Replay Attacks from Far-Field Recordings on Speaker Verification Systems*. In Biometrics and ID Management, Springer Berlin Heidelberg, Berlin, Heidelberg, 274–285.