Securing IoT and Edge Devices under Audio Adversarial Attacks

PIs: Yingying (Jennifer) Chen[1], Bo Yuan[1], Jian Liu[2]

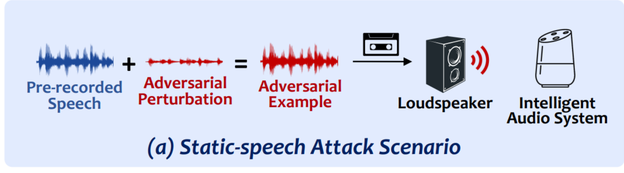
[1]Rutgers University, [2]The University of Tennessee, Knoxville

[1]http://www.winlab.rutgers.edu/~yychen/, https://sites.google.com/site/boyuaneecs/, [2]https://web.eecs.utk.edu/~jliu/

Audio Adversarial Machine Learning Attacks

- Voice commands provide a convenient interface to control and interact with IoT and edge devices
- Voice assistant systems have been shown to be vulnerable to various types of attacks, including emerging adversarial machine learning attacks
- Develop practical audio adversarial attacks against deep learning models for speech recognition and speaker identification, and further design defense techniques

Audio Adversarial Machine Learning Attack



Challenges

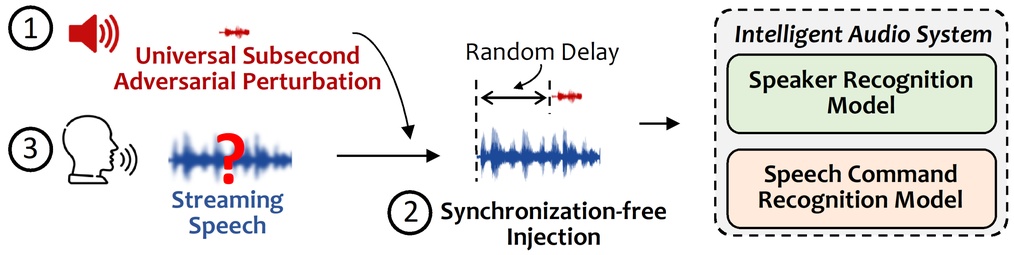
- Investigate audio adversarial attacks under practical timing constraints (e.g., attack against live speech)
- Achieve a holistic solution to defend against audio adversarial attack and other machine-induced audio attacks (e.g., replay attack, voice synthesis attack)
- Enable the defense to generalize new acoustic environments with reduced training cost

Scientific Impacts

- Fully understand and demonstrate the feasibility of audio adversarial attacks in the physical world
- Develop defending strategies in practical environments to build attack-resilient voice-controllable IoT and edge systems
- Lead to develop a new computing paradigm in audio-based adversarial machine learning in both theoretic foundations as well as safety-critical audio-oriented emerging applications

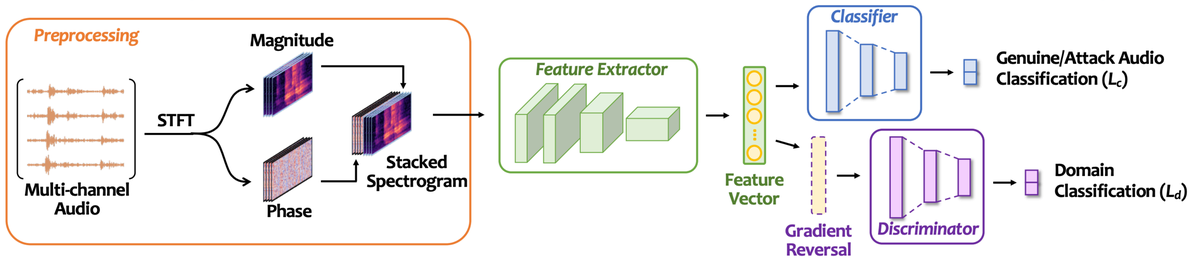
Approaches

Universal, Synchronization-free, and Targeted Audio Adversarial Attacks via Subsecond Perturbations



- Only need to add a very short adversarial perturbation
- Generate input-agnostic universal adversarial perturbations that can make arbitrary audio input to be mis-recognized
- Incorporate the varying time delay into the optimization process to generate synchronization-free perturbations
- Incorporate the main sources of physical distortions during over-the-air propagation to achieve practical attacks

Detection of Machine-induced Audio Attacks with Microphone Array



- Dissect existing machine-induced audio attacks, including replay attacks, voice synthesis attacks, hidden voice commands, inaudible attacks, and audio adversarial examples
- Design a holistic defense strategy leveraging multi-channel audio recorded by the microphone array equipped on voice assistant systems (e.g., Google AIY, ReSpeaker 4-Mic Linear Array, ReSpeaker Core V2)
- Build a deep learning model and adopt the unsupervised domain adaptation framework to achieve environment-independent detection

Broader Impacts

- Advance the defense techniques against audio adversarial attacks
- Include curriculum development, outreaching to K-12 students
- Interact and share results directly with industrial voice service providers