Over-the-Air Speech Recognition Attack

UCLA EE209AS Winter 2018 by Professor. Mani srivastava
Final Project Proposal
Team members: Weikun Han, Zhengshuang Ren

Introduction

Recent progress in intelligent home assistant devices such as Google Home and Amazon Alexa is changing people's daily lives and allows users to interact with their home devices in a smarter and more convenient. Such devices are integrated with speech recognition models mostly based on Deep Learning Neural Networks to recognize users' voice commands. The advantage of using Deep Learning Models is their higher accuracy on recognizing users' commands correctly than traditional machine learning algorithms. However, such devices bring new security concerns since they are operating users' private home devices and transmitting sensitive data and information about users' private personal lives. Vulnerabilities of these devices may be exploited and used to cause users' property loss.

Problem Statement

Recent research has shown that deep learning models are easy to be fooled by attackers to perform untargeted or even targeted attacks by generating adversarial examples to produce wrong recognized commands and to actuate users' home devices in unwanted ways. Moustafa Alzantot and Nicholas Carlini have demonstrated the vulnerabilities of such speech recognition models by generating adversarial examples to perform targeted attacks with high successful rates. Notice in Moustafa's work, the author is performing black-box attacks without knowing the details about the recognition neural network whereas in Carlini's work, the attack is performed in white-box attacks leveraging the structure and details about the network.

However, they achieved the research-purpose attacks by deploying adversarial example files into the home assistant devices, but in practical attacks, over-the-air attacks are more realistic to perform since the attackers may not have physical access to the devices.

Objective

In this project, we proposed a way to generate the adversarial examples before the air channel so that the attackers are able to perform over-the-air attacks based on the adversarial examples they have already have. The main idea of our project is to mimic the air channel characteristics, the characteristics of the speaker used to play the adversarial examples and the microphone on the home assistant devices listening to the commands in order to predict and construct the original adversarial examples which will result in the ones in the existing works after passing through the speaker-air-microphone channel. We leveraged the power of deep learning neural network to mimic the speaker-air-microphone channel to provide high accuracy and avoid the complicated analysis of speaker/microphone circuits and the acoustic air channel.