Adversarial Machine Learning Against Voice Assistant Systems

RUTGERS
WINLAB | Wireless Information Network Laboratory

Catherine Mathews, Matt Kokolus, Raymond Huang, David Man
Advisor: Dr. Yingying Chen

Objective

- Study the security of voice assistant systems under adversarial machine learning
- Develop a system to generate hidden voice commands to attack voice assistants
- Use a drone to carry loudspeaker to play hidden voice commands to a voice assistant system

Background

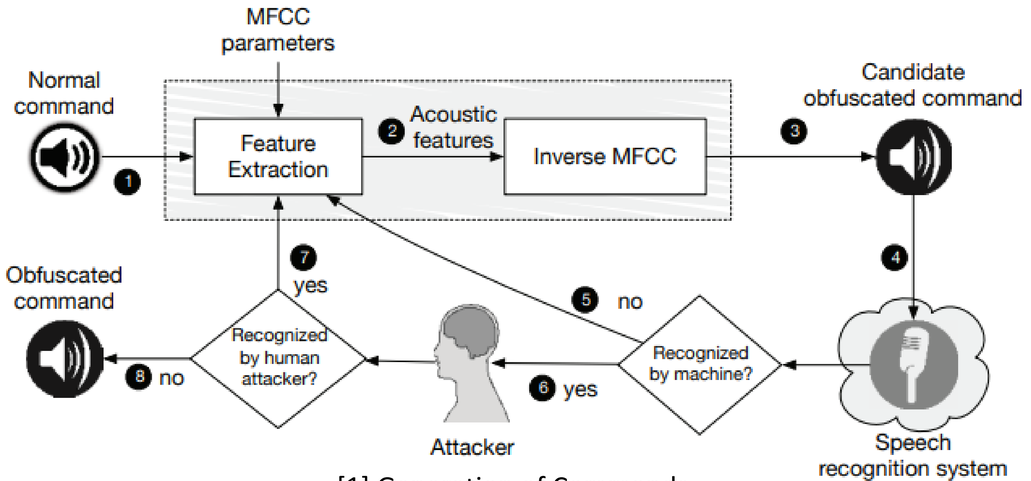
- Hidden Voice Commands
  - Audio samples that have been altered to fool speech recognition systems
  - Interpreted as common commands by voice assistants, but unrecognizable to human listeners

Drone Progress

- Objective: Use Holy Stone HS700 drone to carry loudspeaker that will play commands to attack voice assistant system
- Completed setup and tested flight of Yuneec Tornado H920

Generating Hidden Voice Commands

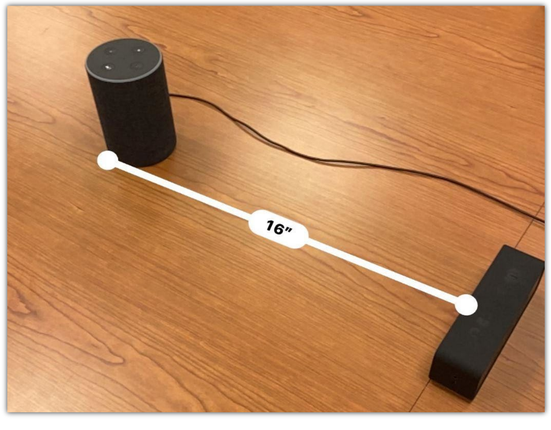
1. Convert audio input into mel-frequency cepstral coefficients (MFCCs) while adding noise
2. Invert the MFCCs back into an audio file while adding additional noise
3. Test if audio is interpretable by humans and/or by voice assistants
4. Repeat process until audio is interpretable to only voice assistants



[1] Generation of Commands

Experiment Procedure

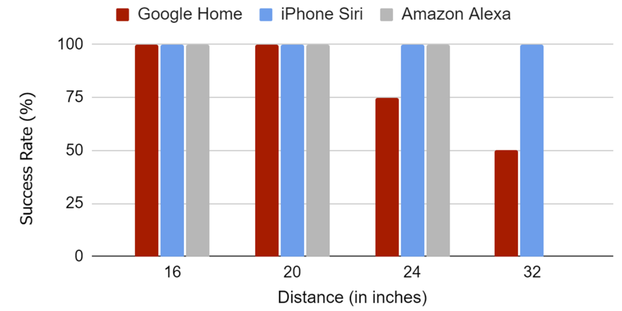
- Recorded our own voice commands
- Obfuscated each command and played it through a speaker facing a voice assistant device
  - Gradually increased distance between speaker and device
  - Measured speaker volume, room noise, command success
- Systems: Google Home, Amazon Alexa, Apple's Siri
- Commands: "What's the time", "What's the weather like today", "Set a timer for 10 minutes"



Experiment Results

- Google Home
  - Recognized all commands at distances up to 22 inches
- Amazon Alexa
  - Recognized all commands at all measured distances (less than 32 inches)
- Apple's Siri
  - Recognized all commands at distances up to 11 feet

Success Rate vs. Distance



Future Work

- Generate more commands that are less recognizable to humans
- Attach drone to loudspeaker and carry out commands from drone

Acknowledgements

We would like to thank our advisor Dr. Yingying Chen and our mentors Cong Shi, Bin Hu, Kailong Wang, Wenjin Zhang, and Linqi Xiao for their support and guidance throughout this project.

References

[1] Carlini, Mishra, Vaidya, Zhang, Sherr, Shields, Wagner, & Zhou. (2016, August). Hidden Voice Commands. https://www.usenix.org/system/files/conference/usenixsecurity16/sec16_paper_carlini.pdf

[2] Li, Shi, Xie, Liu, Yuan, & Chen. (2020, March). Practical Adversarial Attacks Against Speaker Recognition Systems. https://winlab.rutgers.edu/wp-content/uploads/2021/06/Practical-Adversarial-Attacks-Against-Speaker-Recognition-Systems.pdf

WINLAB