SaTC: CORE: Medium: Securing the Voice Processing Pipeline Against Adversarial Audio

Submitted by Patrick Traynor on Mon, 07/06/2020 - 3:55pm

Project Details

Lead PI

Patrick Traynor

Co-PIs

Vincent Bindshaedler
Thomas Shrimpton

Performance Period

Oct 01, 2019 - Sep 30, 2023

Institution(s)

University of Florida

Sponsor(s)

National Science Foundation

Award Number

1933208

Ranked 542 out of 2290 Group Projects.
111 related hits.

In a world in which many new computing devices have limited or no traditional user interface (e.g., smart thermostats, personal digital assistants including Amazon's Alexa, etc), voice interfaces are becoming a primary means of interaction. Such systems not only simplify interaction with conventional devices for traditional users, but also promote broader inclusion for both the elderly and those with disabilities. These interfaces have been made significantly more accurate in recent years through the application of deep learning techniques; however, these techniques are subject to a number of attacks using modified audio. While previous researchers have demonstrated such attacks using significant knowledge of specific deep learning models, our initial work demonstrates that knowledge of signal processing (or how voices are turned into the inputs deep learning models require) can create attacks that work across a wide variety of systems. The work proposed in this grant will allow us to fully characterize the security challenges in the space between signal processing and deep learning, and to develop strong defenses to ensure that these systems can continue to operate in the presence of malicious inputs. A wide range of systems, from the Internet of Things (IoT) to infrastructure such as air traffic control, will benefit from improved resilience to malicious audio.

This effort is focused on the design methods and tools to protect the entire voice processing pipeline. In our view, this naturally segments our efforts into three logical thrusts, beginning with an in-depth analysis of the algorithms used for audio preprocessing and an investigation of comprehensibility metrics from the field of psychoacoustics. These efforts naturally lead into our second thrust, which focuses on the algorithms used in the second step of the audio processing pipeline. Here, we exploit weaknesses in the most popular feature extraction algorithms to produce new attacks, and then develop defenses against such attacks and techniques to protect speaker privacy. Our final thrust investigates the impact of attacks in the two previous thrusts and their impact on the underlying machine learning algorithms. With these insights, we will investigate additional methods of protecting particularly vulnerable layers of models against these attacks. The researchers possess the unique expertise in areas including information security, voice interfaces, adversarial machine learning, privacy-preserving data synthesis, and statistical signal processing.

Patrick Traynor

Related Artifacts

No related artifacts found.

CORE Medium Division of Computer and Network Systems (CNS)