How voice assistants follow inaudible commands

October 23 2019

The researchers can conceal secret messages for voice assistants in any audio file, those including speech, music and ambient noise – e.g. birds' twittering. Credit: Roberto Schirdewahn

An attack against speech recognition systems with manipulated audio

files used to work only via a data interface. Now, all it takes is playing the secret messages via loudspeakers.

Researchers can conceal voice commands for machines that are inaudible to the human ear in any audio file. Speech recognition systems understand those commands perfectly well. In September 2018, researchers from the Horst Görtz Institute for IT Security at Ruhr-Universität Bochum reported such attacks against the speech recognition system Kaldi, which is integrated in Alexa. Originally, those so-called adversarial examples could only be performed via a data interface; today, they work perfectly well over the air. A detailed article on those attacks and potential countermeasures can be found in Bochum's science magazine Rubin.

In order to integrate secret messages into audio files, the researchers take advantage of the psychoacoustic model of hearing. "As long as the ear is busy processing a sound at a specific frequency, humans are incapable of hearing other sounds at low volume for a few milliseconds," explains Lea Schönherr from the research group Cognitive Signal Processing, headed by Professor Dorothea Kolossa. These frequencies are where researchers hide the secret commands for machines. To the human ear, the additional information sounds like random static noise; but it changes the meaning of the message for the voice assistant.

Taking the room into consideration

Originally, the attack could only be performed directly via the data interface; today, loudspeakers will do. This is more complicated, as the sound is affected by the room in which the file is played. Accordingly, when creating manipulated audio files, Lea Schönherr takes the so-called room impulse response into consideration. It describes how a room reflects and changes the sound. Room impulse responses can be simulated using dedicated computer programs.

"The attack can be tailored to a specific room setup in which it is played," elaborates the communication engineer. "However, we have recently performed a generic attack, which does not need any prior information about the room, but still works equally well or even better over the air." In the future, the researchers are planning to run tests with voice assistants available in the market.

Closing the security gap

Since speech recognition systems aren't currently deployed in any safety-critical applications but are mainly used for convenience, adversarial examples cannot do a lot of damage yet. Therefore, there's still time to close this security gap, according to the researchers from Bochum. In the Cluster of Excellence Casa, short for Cyber Security in the Age of Large-Scale Adversaries, the research group Cognitive Signal Processing, which developed the attacks, collaborates with the Chair for System Security headed by Professor Thorsten Holz, whose team is designing the countermeasures.

MP3 principle as countermeasure

IT security researcher Thorsten Eisenhofer intends to teach the speech recognition system to eliminate any ranges in the audio signals that are inaudible to humans and to hear only the rest. "We cannot prevent audio files being manipulated by attackers," he says. His goal is to rather force an attacker to place the manipulation into audible ranges; thus, attacks could no longer be easily hidden. Eisenhofer uses the MP3 principle for this purpose.

MP3 files are compressed by deleting any ranges that are inaudible to humans—and this is what the defense strategy against adversarial examples is aiming at. Consequently, Eisenhofer combined Kaldi with an MP3 encoder that cleans up the audio files before they reach the

speech recognition system. The tests have shown that Kaldi did indeed no longer understand the secret messages, unless they were moved into the human hearing range. "At this point, the audio files were considerably changed," explains Thorsten Eisenhofer. "The static in which the secret commands are hidden could be distinctly heard."

More information: Lea Schönherr, et al. Imperio: Robust Over-the-Air Adversarial Examples for Automatic Speech Recognition Systems. arxiv.org/abs/1908.01551: arXiv:1908.01551v2 [cs.CR]

Provided by Ruhr-Universitaet-Bochum

Citation: How voice assistants follow inaudible commands (2019, October 23) retrieved 23 April 2024 from https://techxplore.com/news/2019-10-voice-inaudible.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.