When speech assistants listen even though they shouldn't

July 3 2020, by Julia Weiler, Donata Zuber

This voice assistant does not only react to the trigger word “Amazon,” but is also activated by the phrase “and the zone.” Credit: RUB, Marquard

Researchers from Ruhr-Universität Bochum (RUB) and the Bochum Max Planck Institute (MPI) for Cyber Security and Privacy have

investigated which words inadvertently activate voice assistants. They compiled a list of English, German, and Chinese terms that were repeatedly misinterpreted by smart speakers as prompts. Whenever the systems wake up, they record a short sequence of what is being said and transmit the data to the manufacturer. The audio snippets are then transcribed and checked by employees of the respective corporation. Thus, fragments of very private conversations can end up in the companies' systems.

Süddeutsche Zeitung and NDR reported on the results of the analysis on 30 June 2020. Examples yielded by the researchers' analysis can be found at unacceptable-privacy.github.io.

For the project, Lea Schönherr from the RUB research group Cognitive Signal Processing, headed by Professor Dorothea Kolossa at the RUB Horst Görtz Institute for IT Security (HGI), collaborated with Dr. Maximilian Golla, previously at HGI, now at MPI for Security and Privacy, as well as Jan Wiele and Thorsten Eisenhofer from the HGI Chair for Systems Security headed by Professor Thorsten Holz.

Testing all major manufacturers

The IT experts tested the voice assistants by Amazon, Apple, Google, Microsoft, and Deutsche Telekom, as well as three Chinese models by Xiaomi, Baidu, and Tencent. They played them hours of English, German, and Chinese audio material, including several seasons from the series "Game of Thrones," "Modern Family," and "House of Cards," as well as, news broadcasts. Moreover, professional audio data sets that are used to train smart speakers were also included.

All voice assistants were equipped with a light sensor that registered when the activity indicator of the smart speaker lit up, thus, visibly switching the device into active mode indicating that a trigger occurred.

The setup also registered when a voice assistant sent data to the outside. Whenever one of the devices switched to active mode, the researchers recorded which audio sequence had caused it. They later manually evaluated which terms had triggered the assistant.

The researchers used their setup to analyse eleven different smart speakers, including devices by Amazon, Apple, Google, Microsoft, and Deutsche Telekom. Credit: RUB, Marquard

False triggers identified and generated

Based on this data, the team created a list of over 1,000 sequences that incorrectly trigger speech assistants. Depending on the pronunciation, Alexa reacts to the words "unacceptable" and "election," while Google

reacts to "OK, cool." Siri can be fooled by "a city," Cortana by "Montana," Computer by "Peter," Amazon by "and the zone," and Echo by "tobacco."

In order to understand what makes these terms false triggers, the researchers broke the words down into their smallest possible sound units and identified the units that were often confused by the voice assistants. Based on these findings, they generated new trigger words and showed that these terms also activated the voice assistants.

"The devices are intentionally programmed in a somewhat forgiving manner, because they are supposed to be able to understand their humans. Therefore, they are more likely to start up once too often rather than not at all," concludes Dorothea Kolossa.

Using light sensors, they registered when the indicator LEDs of the speakers lit up. Credit: Maximilian Golla

Audio snippets are analyzed in the cloud

The researchers analyzed in more detail how the manufacturers evaluate false triggers. A two-stage process is most common. First, the device analyzes locally whether the speech it perceives contains a trigger word. If the device suspects that it has heard the trigger word, it begins to upload the current conversation to the manufacturer's cloud for further analysis with more computing power. If the cloud analysis identifies the term as a false trigger, the voice assistant remains silent, only its indicator LED lights up briefly. In this case, several seconds of audio recording may already end up at the corporation, where they are transcribed by humans in order to avoid such false triggers in the future.

"From a privacy point of view, this is of course alarming, because sometimes very private conversations can end up with strangers," says Thorsten Holz. "From an engineering point of view, however, this approach is quite understandable, because the systems can only be improved using such data. The manufacturers have to strike a balance between data protection and technical optimisation."

More information: GitHub: Unacceptable, where is my privacy? Exploring Accidental Triggers of Smart Speakers: unacceptable-privacy.github.io/

Provided by Ruhr-Universitaet-Bochum