September 2020

Protecting Users Against Malicious Attacks That Target Device Microphones

Anonymous

Protecting Users Against Malicious Attacks That Target Device Microphones

ABSTRACT

Voice-activated devices such as smart speakers, smart displays, smartphones, etc. have been shown to be susceptible to attacks that utilize directed laser beams to activate the device. Such attacks can cause the device to respond to queries even in the absence of actual spoken commands from a user. This disclosure describes multiple approaches such as the use of a physical barrier and use of signal processing techniques to analyze signals detected at multiple microphones of a device to mitigate such attacks. The device can be designed with a light blocking geometry to minimize visible light that reaches the microphone membrane. The device can also include a light sensor in the vicinity of the microphone to detect the presence of light inside the microphone port. Output of the light sensor can be used in the device security logic.

KEYWORDS

- Smart speaker
- Smart display
- Acoustically transparent material
- Voice activation
- Spoken input
- Wake word
- Activation word
- Light command
- Light barrier
- Audio injection attack
- Directed laser beam

BACKGROUND

Voice input is a primary mode of interaction for devices such as smart speakers, smart displays, and other appliances. Other devices such as smartphones, computers, tablets, etc. also support voice input. Most such devices include one or many microphones that detect speech input. Many devices are configured to automatically listen for speech input, e.g., a wake word or

activation word, and respond to spoken commands. Such devices have been shown to be vulnerable to attacks using a directed modulated laser beam [1]. To carry out the attack, an attacker directs a laser beam towards the device causing the microphone to activate even in the absence of actual spoken input.

DESCRIPTION

This disclosure proposes two different solutions that help avoid or minimize this microphone vulnerability. A first approach utilizes a physical barrier and is suitable for any device with a microphone. A second approach utilizes signal processing techniques that verifies that input was received at multiple microphones, thus reducing the likelihood of activation by a directed laser beam.

Physical barrier

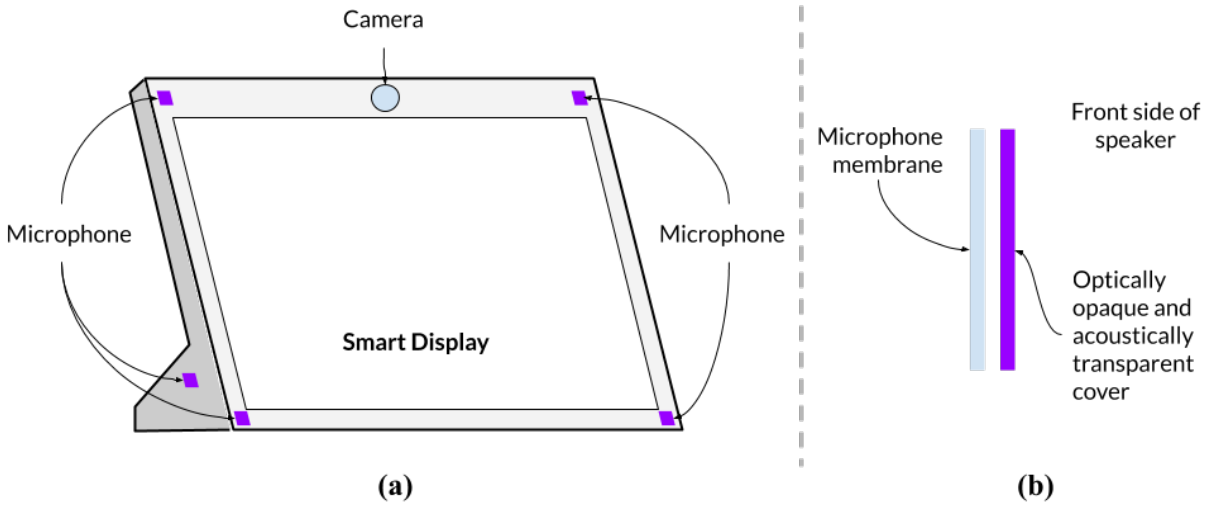

Fig. 1: Blocking light signals with physical barrier

Fig. 1 illustrates a smart display configured with a physical barrier that blocks light signals incident on the microphone(s). In the example of Fig. 1(a), the smart display includes a

camera and multiple microphones. In Fig. 1, four microphones can be seen on the front of the device and one on the side of the device. The use of a physical barrier is suitable for any device with one or more microphones. As shown in Fig. 1(b), the microphone has an outer cover that is a physical barrier in front of the microphone membrane.

The physical barrier can be in front of the microphone (on the outer side of the body of the device) or inside the microphone port (inside the body of the device). The cover is produced out of an optically opaque material which is acoustically transparent. Such a cover, when placed appropriately, can block light (e.g., a directed laser beam or other light) without interfering with the function of the microphone (receiving spoken input). For example, the cover can be made of any suitable material such as open cell foams, fabrics, porous polymers, etc. The cover made from this type of material is designed to minimize attenuation and interference with the microphone response as a function of frequency for different possible placements of the device.

Curved port design is another form of a physical barrier. The microphone ports on the device can be designed with a light blocking geometry (such as placing a right angle in port geometry), which is used to minimize the transmission of visible light into the microphone membrane/ASIC. In addition, the angled part of the port can be designed to be reflective, further minimizing light dispersion beyond the edge of the angle.

Another form of protection is placement of an additional sensor, such as a light detector in the vicinity of the microphone. Such additional sensor can be used to indicate the presence of strong light inside the microphone port. This signal can be used in the security logic of the device.

Signal processing

A second approach is to apply signal processing techniques from the signals detected by multiple microphones of a device, e.g., the four microphones of the smart display of Fig. 1. The signal processing involves determining whether the signals detected by the different microphones are similar. For example, if the directed laser beam hits a single microphone or a subset of microphones, some microphones do not detect the signal intended by the attacker. The mismatch between the signals can be interpreted as a possible attack and the device is not activated in response to the input. In contrast, when a user in the vicinity of the device provides a spoken command, the audio is detected at all microphones, which is deemed safe and the device is activated.

The device is configured to perform self-diagnostics to ensure all microphones are operable and are capable of receiving audio signals. Upon successful completion of the self-diagnostics, and during operation of the device, comparison of the signals detected by the multiple microphones can be made based on:

- **Signal coherence:** signals in the lower end of the speech spectrum (<1kHz) typically show a high degree of coherence, if an acoustic source is stimulating the sensors. Low coherence and complete absence of signal on one of the microphones can indicate the source is of non-acoustic origin.
- **VAD (voice activity detection)** is calculated on each microphone. The voice activity is highly correlated when the signal is of acoustic origin and the SNR (signal to noise ratio) is favorable, indicating a genuine activation.
- **ASR (automated speech recognition)** is run on the audio detected at each microphone and the outputs are compared. A true acoustic source (e.g., spoken audio input) with a

good SNR typically yields similar, or highly correlated results while malicious activation (e.g., by use of a laser beam) is unlikely to yield such a result.

The signal processing technique protects against attackers that target a single or subset of the microphones. The technique is especially suitable for situations when microphones are placed on two or more surfaces of the device, e.g., two or more of front, back, sides, top, and bottom of the device since this type of placement makes simultaneous modulated light attack on all microphones impossible from a single direction. The smart display of Fig. 1 includes microphones on the front and sides of the device and is suitable for use of the signal processing techniques.

CONCLUSION

Voice-activated devices such as smart speakers, smart displays, smartphones, etc. have been shown to be susceptible to attacks that utilize directed laser beams to activate the device. Such attacks can cause the device to respond to queries even in the absence of actual spoken commands from a user. This disclosure describes two approaches - the use of a physical barrier and use of signal processing techniques to analyze signals detected at multiple microphones of a device - to mitigate such attacks. The device can be designed with a light blocking geometry to minimize visible light that reaches the microphone membrane. The device can also include a light sensor in the vicinity of the microphone to detect the presence of light inside the microphone port. Output of the light sensor can be used in the device security logic.

REFERENCES

1. Hackers Can Use Lasers to 'Speak' to Your Amazon Echo
2. Sugawara, Takeshi, Benjamin Cyr, Sara Rampazzi, Daniel Genkin, and Kevin Fu. "Light Commands: Laser-Based Audio Injection Attacks on Voice-Controllable Systems." (2019).