Research of Voice Assistants Safety

Nikita Burym, Mikhail Belenko and Pavel Balakshin

ITMO University, Kronverksky Pr. 49, bldg. A, St. Petersburg, 197101, Russian Federation

Abstract

Internet-connected gadgets with voice assistants are becoming more popular due to their convenience for everyday tasks such as asking about the weather forecast, playing music or controlling other smart things in the house. However, such convenience comes with a privacy risks: smart gadgets have to constantly listen in order to activate when the "wake word" is spoken, and are known to transmit recorded audio from their environment and record it on cloud servers. Specifically, this article focuses on the privacy risks associated with using smart gadgets with voice assistants.

Keywords

voice assistants, smart speakers, gadgets, privacy, IoT, voice command, voice recording, wake word

1. Introduction

In the field of information technology, the means of interaction between a user and a technical system called an interface. Interfaces are different and implemented by different means and methods. One of the most important tasks in the development of modern technical systems is to provide the most intuitive and natural user interface.

One of the natural forms of human interaction is speech. The voice interface is one of the key parts of human-machine interaction, allowing improve the existing user interface, as well as provide a more convenient way of human-computer interaction. Google's voice assistant [1] and Apple's Siri voice assistant [2] are prime examples, highlighting the urgent need to introduction speech technologies such as speech recognition and voice interfaces.

Today, voice assistants are in demand in the customer support segment. Most large companies use communication services to improve the quality of customer service.

Voice assistants are increasingly appearing in the form of home appliances that surround us. Notable examples from this area are smart speakers Google Nest, previously named Google Home [3], Amazon Echo Dot [4], Apple HomePod [5] and Harman/Kardon Invoke [6], which provide a voice control interface. People can use them to turn music on and off, ask for weather forecast, tell them to adjust the room temperature, order goods online, and much more.

To provide better user experience, most devices with a built-in voice assistant use an always-listening mechanism that receives voice commands all the time [7]. In particular, users are not

Proceedings of the 12th Majorov International Conference on Software Engineering and Computer Systems, December 10–11, 2020, Online & Saint Petersburg, Russia

CHVRCHES@mail.ru (N. Burym); mikael0bmv@gmail.com (M. Belenko); pvbalakshin@gmail.com (P. Balakshin)

0000-0002-4343-6408 (N. Burym); 0000-0002-5060-1512 (M. Belenko); 0000-0002-9421-8566 (P. Balakshin)

required to press or hold a physical button on devices before speaking commands. However, this advantage may expose users to security threats due to the openness of voice channels.

2. Issue 1: Unauthorized access

Most voice assistants require a voice command or the so-called “wake word” to initiate user interaction. For different devices, these voice commands are different, for example, for the Google voice, you need to say the phrase “OK Google” [1, 3], and for an Amazon voice assistant, the phrase “Hello Alexa” is required [4], after which the assistants are informed that the user is ready to ask a question or tell commands. As a result, anything that said on radio, television, or during normal human dialogue can accidentally wake an assistant. On the one hand, this may seem harmless, but some voice assistants transfer the recording or its text version and other data to the cloud server to execute the user’s command. This data can be stored on cloud servers for quite a long time and used by a voice assistant company, for example, to check the quality of speech recognition, which leads to the following risks:

1. **Voice assistants can store more information than intended.** They should only record voice after they hear the wake word, but they may react to similar words and speech on TV or wake for no reason.
2. **Employees can get access to personal information, because they are the ones who check the quality of the voice assistants.** They can find out personal data, for example, a medical history told to monitor a patient’s condition in a hospital or cash card information pronounced for a purchase in the Internet.
3. **Criminals can take advantage of the data.** Like other information collected, voice recordings that locating on a cloud server are at risk of hacker attacks. They can be stolen and used, for example, to simulate the user’s voice to hack devices protected by biometrics [8].

3. Issue 2: Anyone can control the device

It should be noted that voice assistants are designed to be at the center of the Internet of Things ecosystem. Thus, while they allow users to access the Internet and execute various commands, they can also communicate and control all other smart gadgets in the home.

Recently, a new method has been found that allows you to control Apple’s Siri voice assistant using inaudible Ultrasonic waves [9]. Ultrasonic waves are sound waves with a frequency higher than a human can hear [10]. However, smart gadget microphones can record these higher frequencies. This method can activate the voice assistant and make it use various functions of the smartphone, for example, a phone call, transfer commands to other devices in the IoT ecosystem without touching the gadget, since the assistant thinks you are saying a command and proceeds to transmit important information. For this reason is advised not to connect the device to any IoT security solutions such as smart door locks, as hackers can use a voice assistant to instruct the device to unlock the front door and enter in the house.

To protect against this kind of attacks, voice assistant companies are encouraged to develop software for the phone that analyzes the received signal to discriminate between ultrasonic waves and genuine human voices [9].

4. Issue 3: Misactivating of smart speakers

Researchers from Northwestern University and Imperial College London conducted experiments during which it turned out that smart speakers in which voice assistants are embedded can be activated when watching TV series and spy on users [11].

The purpose of this work was to find out if smart speakers record random sounds from the environment, and if so, how and when it does. The researchers also tried to identify what false wake words typically misactivate the voice assistants, certain types of dialogue, and other factors. In particular, the work answered the following questions:

1. **How often do smart speakers misactivating?** It is characterized by how often the smart speaker is incorrectly activated during a conversation. The more cases of incorrect activation, the higher the risk of unexpected audio recordings.
2. **How long does smart speaker recording environmental sound after misactivating?** Prolonged misactivation, represent a higher privacy risk than a short one, as more data (such as context and conversation details) is recorded over a long period.
3. **Are there certain TV shows that cause more misactivations than others do? If yes, why?** Each TV show you select contains different conversational characteristics (accent, context, etc.). It measures which ones cause the most misactivation in order to understand which characteristics correspond to the increased risk of misactivation.
4. **What words that do not wake up properly like “Hey Alexa” or “Okay Google” are constantly causing misactivation?** This will help find undocumented wake words or sounds that should be avoided by the user.

During the experiment, the researchers played the content of the American entertainment company Netflix for 134 hours next to smart speakers. They selected TV series of different genres with many dialogues and watched if the phrases from the dialogues in the series could activate voice assistants in Google Home Mini, Apple Homepod, Harman Kardon (Cortana) and two generations of Amazon Echo Dot, as each version has a different number of microphones for recognition, which can affect the accuracy of speech recognition. Researchers repeated the tests several times in order to determine which words not intended to wake up the assistant regularly activate smart speakers [11]. Table 1 shows the smart speakers tested and their characteristics.

As it turned out during the experiment, the assistant is misactivating up to 19 times a day, while Siri and Cortana are most likely to misactivate and record environmental sounds. Most often, assistants were misactivating when watching the TV series “Gilmore Girls" and “The Office".

Researchers have also identified some patterns in which words not intended for the assistant can activate it. For example, these turned out to be words that rhyme with the words of activation (in particular, Amazon Echo mistook the phrase “kevin’s car” for “Alexa”). Table 2 provides a list of some pattern phrases of misactivating of voice assistants [11].

Table 1
Smart Speakers and their characteristics

Device	Assistant	Wake word
Google Home Mini	Google Assistant	“OK/Hey Google”
Apple Homepod	Apple Siri	“Hey Siri”
Amazon Echo Dot	Amazon Alexa	“Alexa”,“Amazon”,“Echo”,“Computer”
Harman Kardon Invoke	Microsoft Cortana	“Cortana” (US only)

5. Testing: Misactivating of Yandex’s Alice

In this part of work, a research conducted misactivating of the voice assistant “Alice” in Russian. This assistant activates by the wake phrases “Listen, Alice”, “Alice”, and “Hello, Alice” [12].

The start of the researching consisted of choosing consonant words and phrases to test for misactivations. In total, 25 words and phrases were chosen for dictating. It can be concluded that the voice assistant “Alice” has a good speech recognition module because dictating consonant words does not cause misactivations.

The next stage involved verification misactivaions without the participation of a speaker. For this task cartoon in Russian was chosen based on the fairy tale “Alice’s Adventures in Wonderland” with duration about 1 hour and 3 minutes and the first season of the TV series “The Alienist”. No misactivations were detected during the playback of the series. During the playback of the cartoon the assistant was activated 9 times and only for the word “Alice”. However, it should be noted that it was triggered by the sound reproduced through the speakers, and not by the real human voice, which confirms the fact that there was no analysis of the incoming signal. In addition, after activation the voice assistant recorded the sounds of the environment with sound of the cartoon, in which there could potentially be dialogues of people or confidential information.

Based on the results of this part of work, it can be concluded that the voice assistant “Alice” has a good speech recognition module, since dictating consonant words does not cause misactiovaions. Also, it can be noticed that this voice assistant often recognizes the Russian word for “fox” as “Alice” in recognizing commands mode (wake mode).

6. Conclusion

The fast introduction of smart voice assistants in homes, businesses and public places has raised a number of concerns from privacy advocates. While these devices offer comfortable voice interaction, their microphones always listen for the wake words. As smart speakers become more common in everyday life, there is an urgent need to understand the behavior of this ecosystem and its impact on consumers. In this work several security vulnerabilities in smart speakers and voice assistants were reviewed. The main disadvantage of modern voice assistants is without physical presence-based access control, they can receive voice commands even when there are no people nearby.

Table 2
List of some misactivating patterns among repeatable misactivations

Words	Some patterns	Some examples from the subtitles
OK/Hey Google	Words rhyming with “Hey” or “Hi” (e.g., “They” or “I”), followed by hard “G” or something containing “ol”	“Okay ... to go”, “maybe I don’t like the cold”, “they’re capable of”, “yeah ... good weird”, “hey ... you told”, “A-P ... I won’t hold”
Hey Siri	Words rhyming with “Hey” or “Hi” (e.g., “They” or “I”), followed by a voiceless “s”/“f”/“th” sound and a “i”/“ee” vowel	“Hey ... missy”, “they ... sex, right?”, “hey, Charity”, “they ... secretly”, “I’m sorry”, “hey ... is here”, “yeah. I was thinking”, “Hi. Mrs. Kim”, “they say ... was a sign”, “hey, how you feeling”
Alexa	Sentences starting with “I” followed by a “K” or a voiceless “S”	“I care about”, “I messed up”, “I got something”, “it feels like I’m”
Echo	Words containing a vowel plus “k” or “g” sounds	“Head coach”, “he was quiet”, “I got”, “picking”, “that cool”, “pickle”, “Hey, Co.””
Computer	Words starting with “comp” or rhyming with “here”/“ear”	“Comparisons”, “I can’t live here”, “come here”, “come onboard”, “nuclear accident”, “going camping”, “what about here?”
Amazon	Sentences containing combinations of “was”/“as”/“goes”/“some” or “I’m” followed by “s”, or words ending in “on/om”	“it was a”, “I’m sorry”, “just ... you swear you won’t”, “I was in”, “what was off”, “life goes on”, “have you come as”, “want some water?”, “he was home”
Cortana	Words containing a “K” sound closely followed by a “R” or a “T”.	“take a break ... take a”, “lecture on”, “quartet”, “courtesy”, “according to”

References

[1] Google Voice Assistant, https://assistant.google.com. Last accessed 10 October 2020

[2] Apple Siri Voice Assistant, https://www.apple.com/ru/siri. Last accessed 10 October 2020

[3] Google Nest, https://en.wikipedia.org/wiki/Google_Nest_(smart_speakers). Last accessed 12 October 2020

[4] Amazon Echo Dot, https://www.amazon.com/Echo-Dot/dp/B07FZ8S74R. Last accessed 12 October 2020

[5] Apple HomePod, https://www.apple.com/homepod. Last accessed 12 October 2020

[6] Harman/Kardon Invoke, http://www.harmansound.ru/product/harman-kardon-invoke-black. Last accessed 12 October 2020

[7] Xinyu Lei, Guan-Hua Tu, Alex X. Liu, Chi-Yu Li, Tian Xie: The Insecurity of Home Digital Voice Assistants - Amazon Alexa as a Case Study. In: 2018 IEEE Conference on Communications and Network Security (CNS)

[8] Prospects and problems of voice assistance https://blog.dti.team/voice-assistants-3/. Last accessed 18 October 2020

[9] Qiben Yan, Kehai Liu, Qin Zhou, Hanqing Guo, Ning Zhang: SurfingAttack: Interactive Hidden Attack on Voice Assistants Using Ultrasonic Guided Waves. Computer Science & Engineering, Washington University in St. Louis

[10] Ultrasound, https://en.wikipedia.org/wiki/Ultrasound. Last accessed 19 October 2020

[11] Daniel J. Dubois, Roman Kolcun , Anna Maria Mandalari, Muhammad Talha Paracha, David Choffnes, Hamed Haddadi: When Speakers Are All Ears: Characterizing Misactivations of IoT Smart Speakers. In: Proceedings on Privacy Enhancing Technologies, vol. 2020 (4), pp. 255-276

[12] Yandex Alice, https://yandex.ru/alice. Last accessed 25 October 2020