electronics

MDPI

Article

Inaudible Attack on AI Speakers

Seyitmammet Saparmammedovich Alchekov [1], Mohammed Abdulhakim Al-Absi [1], Ahmed Abdulhakim Al-Absi [2,*] and Hoon Jae Lee [1,*]

[1] Department of Computer Engineering, Graduate School, Dongseo University, Busan 47011, Republic of Korea; mslchekov@gmail.com (S.S.A.); d0185123@kowon.dongseo.ac.kr (M.A.A.-A.)
[2] Department of Smart Computing, Kyungdong University, Gosung 24764, Republic of Korea
* Correspondence: absiahmed@kduniv.ac.kr (A.A.A.-A.); hjlee@dongseo.ac.kr (H.J.L.)

Abstract: The modern world does not stand still. We used to be surprised that technology could speak, but now voice assistants have become real family members. They do not simply turn on the alarm clock or play music. They communicate with children, help solving problems, and sometimes even take offense. Since all voice assistants have artificial intelligence, when communicating with the user, they take into account the change in their location, time of day and days of the week, search query history, previous orders in the online store, etc. However, voice assistants, which are part of modern smartphones or smart speakers, pose a threat to their owner's personal data since their main function is to capture audio commands from the user. Generally, AI smart speakers such as Siri, Google Assistance, Google Home, and so on are moderately harmless. As voice assistants become versatile, like any other product, they can be used for the most nefarious purposes. There are many common attacks that people with bad intentions can use to hack our voice assistant. We show in our experience that a laser beam can control Google Assistance, smart speakers, and Siri. The attacker does not need to make physical contact with the victim's equipment or interact with the victim; since the attacker's laser can hit the smart speaker, it can send commands. In our experiments, we achieve a successful attack that allows us to transmit invisible commands by aiming lasers up to 87 m into the microphone. We have discovered the possibility of attacking Android and Siri devices using the built-in voice assistant module through the charging port.

Keywords: inaudible attack; smart speaker; assistance; attack; speaker

Citation: Alchekov, S.S.; Al-Absi, M.A.; Al-Absi, A.A.; Lee, H.J. Inaudible Attack on AI Speakers. *Electronics* **2023**, *12*, 1928. https://doi.org/10.3390/electronics12081928

Academic Editor: Cecilio Angulo

Received: 9 March 2023
Revised: 8 April 2023
Accepted: 13 April 2023
Published: 19 April 2023

1. Introduction

Modern smart assistants, designed to respond to the user's voice commands, constantly record the surrounding space, which carries a number of threats. The fact is that audio recordings made by a smartphone or speaker are often sent to the servers of the manufacturing company for quality assessment, where they can potentially become the prey of intruders. This technology is still imperfect, and voice assistants do not always "understand" how to react to the user's command [1,2]. This can lead to the fact that in the smart house, which is controlled by the assistant, the alarm system can turn off or the heating starts up while there is no one there. By default, voice assistants work and collect information about what is happening around it using a microphone. All words that were spoken in the hearing range of the device's microphone are transmitted to the servers of the developer company for recognition and processing. In the event of a data leak, snippets of conversations and information about the user's location may be disseminated. At the same time, the system can be hacked either point wise, damaging the user's gadget, or it is possible to carry out a hacker attack on the manufacturer's services—this is more expensive, since it requires lengthy preparation and appropriate skills [3,4]. As a result, user data—speech samples, personal information, business conversations—end up with third parties who can use it on their own or sell it to fraudsters.

1. Speech input;
2. Device starts recording when the trigger word is heard, e.g., "Hey Siri or Alexa". The LED indicates recording status;
3. Recording is sent to the cloud for processing and stored;
4. Response is sent back; traffic is SSL encrypted.

It is optional that Backend can pass information to external extensions (Skills/Actions) for further processing. Voice assistants are ubiquitous, but not all users see them as a potential source of cyber threats. Devices with this technology can have access to various devices, as well as to the personal data of their owner. Many researchers have discovered and proven vulnerability in voice assistants, due to which fraudsters can manage user accounts, as well as gain access to their voice history and personal data.

AI speakers have a complex structure, which usually includes at least a smart speaker (e.g., Amazon Echo, Google Home) and a cloud-based voice personal assistant (e.g., Alexa, Google Assistant). A typical request works as shown in Figure 1: In order to start a conversation with the voice assistant, we must first wake the voice assistant up with the words "Ok Google" or "Siri" according to its model. The moment a person utters the words of awakening and begins to communicate with the speaker, voice activation occurs. After displaying a characteristic window or highlighting, we can ask our question. As a result, our request is sent to the server to determine what we said, the data is processed and the response flies back. All this happens in seconds or even fractions of seconds, but with rare exceptions, this is how it works. After the command is processed on the server, the command is either sent back as a return command or a search query is generated, the results of which form a voice response (in the case of a smartphone, also a visual one). In the first case, the light turns on, the door opens, etc., and in the second, the weather is announced or, for example, the latest news. In addition, working through the server allows us to make smart home gadgets more versatile, as we can ensure their work at a distance. For example, turn on the kettle when approaching the house, turn on the lights in the garage in advance, or turn on the robotic vacuum cleaner while working [5,6].

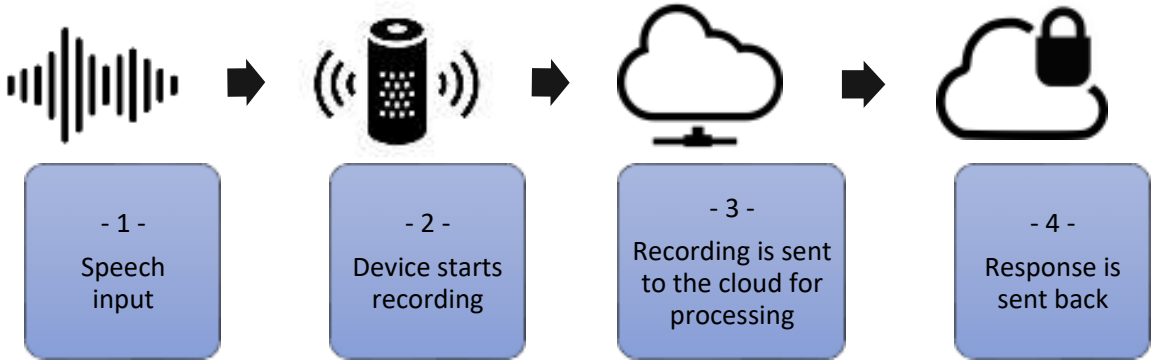


Figure 1. AI speaker's workflow.

In many models, once the intent is determined, the provider sends the user's request to the cloud for storage. User skills include, for example, the ability to play music, check for updates, control other smart home devices, and shop [7].

There are currently about 100,000 Alexa skills [8] and about 5000 Google Assistant skills [9]. Built-in skills that offer services (such as updates and weather forecasts) and third-party skills provided by third-party developers utilizing development skills are the two primary categories of skill kinds (e.g., smart home devices, coffee machines). It is important to note that third-party skills are usually hosted on a remote web service host controlled by a third-party skill developer. Finally, any output produced by the skill data is sent back to the provider, which generates a speech response, which is returned to a smart response that reproduces the data for the user.

1.1. Scammers Can Hack a Voice Assistant

Voice command technology was supposed to make our lives easier, but security experiments say there are some risks involved in using it. The fact is that Voice Assistants are dangerous because they constantly listen to the surrounding space, waiting for commands, and hear everything that happens. At the same time, most of these programs transfer part of the audio stream to the servers of manufacturing companies, in which information is stored indefinitely and can cause serious harm in the future. Further, the fact is that many commands can only be heard by a machine. As shown in Figure 2, as a person is busy watching their favorite series on TV, the voice assistant will be able to hear something else.

Figure 2. We cannot be sure who is giving voice commands to our assistant.

Every year, several home gadget users report cases of strangers hacking into their voice assistants and transmitting commands. These incidents can be considered the first warning signs. How many cases of such fraud are still unknown? It is impossible to even imagine. Moreover, who is to blame?

Manufacturers of voice assistants blame weak user passwords and lack of two-factor authentication for all the troubles.

Yes, of course, weak duplicate passwords can be fertile ground for scammers. Yet, even voice assistants with advanced security protocols can fall victim to hacking in some way. Researchers at institutions in the US, China, and Germany have successfully exploited encrypted audio files over the past couple of years to convince AI-based voice assistants such as Siri and Alexa to carry out their instructions.

All of this underscores the fact that fraudsters can easily gain access to an autonomous device, forcing it to act in their own interests. This is not only opening all kinds of websites, but also making purchases, making money transfers, opening doors at home or gaining access to even more important information.

In fact, the attacker becomes us, and the more functions our voice assistant controls, the more we are at risk. These are the consequences of using voice assistants for fraudulent purposes.

1.2. AI Recognizes Speech That Humans Cannot Hear

This is a colossal security problem, rooted in the fact that artificial intelligence has a much more sensitive "hearing" than humans. For example, people cannot identify every single sound in the background noise of a restaurant hall, but artificial intelligence systems can easily cope with this task [10,11]. AI speech recognition tools can also process audio frequencies outside the range we can hear [12].

This fact gives the scammers an advantage. They get at least two options for submitting "inaudible" commands to programs. First, they can issue commands against a background of white noise, as American students did at Berkeley and Georgetown [13] in 2016. Then, the students played hidden voice commands in the video and through speakers to connect to voice devices, open websites, and put gadgets in flight mode. Yes, these are not the worst consequences, but, nevertheless, they highlight the vulnerability of voice assistants, which we did not even know about (Figure 2). There is another variant of using "inaudible" commands, and in practice it was used by researchers from the Ruhr University in Bochum in Germany [14]. In September of the year 2019, they conducted an unusual experiment: they used loud noises to hide the commands for the voice assistant in their background. In a short demo video, people and the popular Kaldi speech recognition tool could hear a woman reading a business news bulletin. However, the background data has a built-in command that only Kaldi can recognize: "Deactivate the security camera and unlock the front door." Just imagine: we are sitting at home in the evening, after work, some acquaintance throws up a cool video for us. We watch it and do not suspect that intruders may already have our data. Yes, in the realities of our country, voice assistants usually do not manage security systems and we will still need to look for a suitable house, but it is quite possible to steal information about passwords, as well as financial data, in this way. Experts say that in theory, this approach can be scaled, for example, with the help of broadcasts. The scope of application of voice assistants to scammers is not limited to this. One can launch what researchers at Zhejiang University in China [15] call a "DolphinAttack." As part of such an attack, commands are generated and transmitted at a frequency that is beyond the range of human hearing. This type of attack relies on ultrasonic transmission, which means the attacker must be near the targeted devices. Researchers in Zhejiang province have used this technology to use a locked iPhone to make phone calls. They said the "DolphinAttack" could also force a voice-activated device to take photos, send texts, and visit websites. This can lead to malware installation, identity theft, fraudulent purchases, and possibly extortion or blackmail.

1.3. Voice Assistant Working Process—Command and Voice Recognition

Every voice assistant has at least a microphone and speaker: the first is needed to hear your commands, the second to answer you. Depending on the model, the number of microphones, their directivity (usually 360 degrees), sensitivity and other parameters may differ, but this does not affect the principle of work. The voice assistant connects to the Internet, and it cannot work without constant access to the network. You tune your voice assistant to a phrase that "wakes it up"; this is a command that makes it "listen" to whatever you say. For example, by saying Alexa, you will wake up the assistant from Amazon and it will start listening to everything you say. There is no shutdown command as such, the device simply falls asleep when it realizes that the dialogue has ended. You can wake it up by repeating the Alexa command. There were many jokes in the United States about the fact that women whose names are the same as the device from Amazon were unlucky, they would have it working all the time. In fact, in Alexa, just like in other voice assistants, you can change the command word to any other. The advice from the developers is simple: not to keep this word short and to be well recognized even in noisy conditions. The device stores all settings in local memory, the buffer and voice recognition system are also located there. A home voice assistant can be thought of as a simplified version of a smartphone, which may or may not have a display. It is important that the voice recognition system is in most cases local, and it is the device that processes the sound and recognizes it. In some cases, when the device cannot independently recognize and decrypt the voice, it sends the recording to the cloud, where recognition takes place, since the servers have higher performance and large databases for checking and selecting words. As a rule, most manufacturers use combined systems, voice recognition occurs locally, is sent to the server that must process it, it is sent as conditional text or already a command. For a number of requests involving local action, and it is immediately implemented. For

example, when the user says "set the alarm for 8 am," the system executes the command locally without going to the cloud. The same applies to the settings of other smart home devices, including, for example, changing the temperature in the thermostat.

Many people mistakenly believe that if a voice assistant locally recognizes commands and a voice, then this data will forever remain in it, but this is not the case. It all depends on the manufacturer and the brand of the device, but all data is always written after the command word is transmitted to the manufacturer's servers, where they are stored. This data can be used to fine-tune the recognition system and for other purposes; for example, the police can request it as part of a criminal investigation, and the manufacturer will provide it. However, you need to understand that the voice assistant does not record everything that happens around it 24/7: it records only voice excerpts after the command word. The next important point is the languages in which voice assistants can speak. For example, for Amazon, these are English and German. In the case of Amazon, the Alexa service is a convenient way to buy something on the Amazon website, so the distribution area is limited to those countries where the service has the largest audience. For the same Google Home plans, on the contrary, which is to be wherever possible, we will see the expansion of Google Home in all global markets and it will start appearing in all languages. Although so far it is represented in exactly the same number of countries as its direct competitor, here, Google acts as a catch-up and therefore is in no hurry to be everywhere, since other competitors are unlikely to. Below we will discuss why this is so.

So, you said the command word, and your device "woke up", recording your voice to recognize it. The first and most important element is voice recognition and its conversion to text. Typically, the current voice assistants are configured in one primary language. For example, if the device supports both English and Korean, it will not be able to speak two languages at the same time; it will get confused and ask you to repeat phrases. In future devices, this moment will be easily resolved, as voice assistants will learn to determine the language in which you speak. However, for now, they can only analyze the context in which you pronounce the phrase (all this happens in the cloud) in order to replace some words with English. For example, before, when we said the phrase "in what year the album was released, the reverse side of the moon at pink floyd", it was recognized exactly the way we wrote it. In the Google Speech API, today it is recognized differently: "In what year was Pink Floyd's album" The Other Side of the Moon "released." The reason is that Google uses neural networks to analyze not only the phrase itself, but also the said context. That is, the system tries to predict what exactly you want, and what is the meaning of your question.

1.4. Request Processing, Scripts, and Their Execution

After the voice assistant recognizes the voice and turns it into text, it sends it to the cloud (or executes a simple local command, as in the example with the alarm clock). The dignity of a voice assistant directly follows from the number of operations (commands) that it is able to recognize and process correctly. It is the use cases that become the second problem for the wide and rapid distribution of voice assistants around the world (Figure 3).

In this paper, we focus on the following research questions. Why are there many attacks, but no solutions yet? What are the characteristics of many attacks? What threat models do users have in relation to AI speakers?

To answer these exploration questions, we have run our experiments to show and demonstrate the correctness of all the work. However, we provide a detailed description of the LightCommands' technology. It is a vulnerability in MEMS microphones that allows attackers to remotely send inaudible voice commands to voice assistants, such as Google Assistant and Apple Siri, using a laser pointer. Vulnerabilities affect smart speakers, tablets, and smartphones. We have shown that a low-power laser pointer, modulated by a laser driver and an audio amplifier, shines on the microphone, and even from several tens of meters the microphone interprets this light as the sound of a command. The absence of a PIN required to authenticate most commands eliminates the need to bypass any additional security.

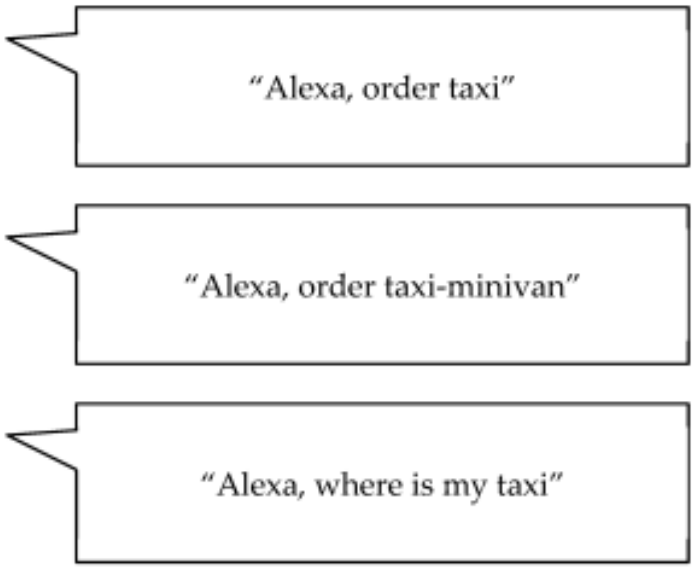


Figure 3. Standard templates.

1.5. How the Paper Is Structured

In this paper, we describe research methods, procedures, tools, etc., so that research can be evaluated and/or replicated. In Section 1, we describe in detail the process of the study itself and the sequence of reasoning, as a result of which theoretical and practical conclusions were obtained. Further, Section 2 describes similar attacks that have been carried out up to this day with different researchers. We describe their stages and stages of research, as well as the rationale and shortcomings of the results obtained in comparison with our attacks. Section 3 describes our first attack, which is carried out with a conventional laser pointer. We describe in detail our research, which is provided in a visual form: in the form of a table, diagram, etc. In Section 4, we present the new attack that we came up with in our lab during our research. In this section, we will explain how a small public charger can become a big problem for smartphone owners. In the discussion section, which is Section 5, we state the significance of our scientific research, first of all, from a subjective point of view. We interpret the results based on the combination of our experience, background knowledge, and scientific capacity, giving several possible explanations. Finally, in Section 6, we provide a brief overview of our completed case study. We conclude the research work and talk about what has already been done. In this section, we confirm the relevance of the study and evaluate the effectiveness of the selected methods to achieve the goal.

2. Related Works

2.1. Injection of Malicious Commands

Many past studies have shown the security of AI speakers, discovering vulnerabilities that allow attackers to send unauthorized voice commands to devices. In particular, [2,16] developed malicious applications for smartphones that play synthetic audio commands to nearby AI speakers without any special permissions from the operating system. While these attacks convey commands that are readily visible to a human listener, other works [13,17,18] have focused on masking commands in audio signals, trying to make them incomprehensible or invisible to human listeners, while remaining recognizable to speech recognition models.

2.2. Inaudible Voice Commands

Later work aims to completely hide the voice commands from the listener. A paper under the name "Backdoor: Making microphones hear inaudible sounds" [19] demonstrates that high-frequency sounds not audible to humans can be recorded using conventional microphones. Subsequently, Song and Mittal [20] and DolphinAttack [15] extended [19]

by sending inaudible commands to AI speakers by modulating the word on ultrasound carriers. Using the non-linearity of the microphone, the signal modulated to the ultrasonic carrier is demodulated by the target microphone into the audible range, restoring the original voice command without being noticed by humans.

Nevertheless, those attacks have distance limitations from around 2 cm to 175 cm due to the low power of the transmitter.

Shockingly, expanding the power produces a discernible recurrence part containing the secret voice order, since the transmitter is likewise influenced by a similar non-linearity found in the getting mouthpiece. Resolving the issue of distance restriction, Roy et al. [21] relieved this impact by parting the sign into various recurrence components and playing them through a variety of 61 speakers. In any case, the return of the discernible release actually restricts the assault reach to 25 feet (7.62 m) in open space with actual obstacles (such as windows), and assimilation of ultrasonic waves noticeable all around further decreases the reach by constricting the communicated signal.

Most of today's voice assistants are based on cloud-based platforms that either require a reliable Internet connection to function or pose serious privacy risks [22,23]. In Ref. [24], the authors of provided an overview of the state-of-the-art components currently on the market to create a senior voice assistant that is open source, privacy-by-design, and fully implemented in a functional way. The authors chose to use an open data source to obtain nutritional information for packaged foods, a task that is relevant to health.

2.3. Skill Squatting Attack

The last space of work centers around befuddling discourse acknowledgment frameworks by making them confused and effectively giving voice orders. These alleged skill squatting attacks [4,25] work by taking advantage of predispositions in the acknowledgment of comparative sounding words, diverting clients to malignant applications without their insight.

2.4. Hidden Attack

One of the attacker's goals is to stealthily attack voice assistants. The attacker will send voice commands at ultrasonic frequencies that are inaudible to humans, and at the same time reduce the volume of the device to such an extent that it is difficult for users to notice the voice responses from the assistant [26].

Inaudible sound: voice assistants hear what we cannot hear. There are more than a billion voice-controlled devices in use in the world, according to a report from the profile site voicebot.ai. Most of them are smartphones, but other speech-capable devices are gaining popularity. One in five Americans, for example, already have a smart speaker at home that responds to spoken language.

2.5. Ultrasound: Machines Hear, Humans Do Not

One way to carry out such an attack is to use ultrasound, that is, a very high-pitched sound that is inaccessible to the human ear. In [15], researchers from Zhejiang University presented a method of covert control of voice assistants called DolphinAttack ("DolphinAttack"—this is the name scientists gave to their development since dolphins can emit ultrasound). Researchers converted voice commands into ultrasonic waves; a person can no longer distinguish the received high frequencies, but microphones of modern devices pick them up. The highlight of the method is that when the sound is converted into an electrical impulse in the receiving device (for example, a smartphone), the original signal containing the voice command is restored. The mechanism here is the same as when the voice is distorted during its recording—that is, this is not a specially designed function of the device, but a feature of the conversion process itself. As a result, the attacked gadget hears and executes a voice command, which opens up wide opportunities for abuse. The researchers successfully replicated the attack on major voice assistants from major

manufacturers, including Amazon Alexa, Apple Siri, Google Now, Samsung S Voice, and Microsoft Cortana.

2.6. Chorus of the Speakers

One of the weak points of the "DolphinAttack" (from the point of view of the attacker) is its small radius of action—the sound must be heard a meter from the attacked device or closer. In [27], the authors from the University of Illinois at Urbana-Champaign managed to increase the distance. They divided several frequency ranges converted into an ultrasonic system, which was played by different speakers—there were more than 60 of them in the experiment. The hidden voice commands performed by this "chorus" of the device were disassembled from a distance of up to seven meters, despite the background noise. In such conditions, the "DolphinAttack" has a much better chance of success.

In [28], another principle was used by specialists from the University of California at Berkeley. They managed to quietly "embed" voice commands in other audio fragments and "trick" the Mozilla Deep Speech recognition system. The modified recording for the human ear hardly differs from the original one, but the program hears exactly the hidden command in it. They prepared and successfully conducted an experiment with two different recordings. In the first example, in the phrase "Without the dataset the article is useless" they "hid" the command to go to the site of the "attackers": "OK, Google. Open evil.com" (OK, Google, browse to evil.com). In the second, in an excerpt from Bach's cello suite, scientists added the phrase "Speech can be embedded in music."

2.7. Protection against Inaudible Attacks

Manufacturers are already thinking about safeguards for voice-activated gadgets. For instance, seeing processing traces in the received signal and altering its frequency helps protect against ultrasonic assaults. Although Google, which has already put these security measures to the test on its assistant, sincerely cautions that this protection can be circumvented using voice recording and that, with the right acting training, the timbre and manner of a person's speech can be faked, it would be ideal to teach all smart gadgets to identify the owner by voice [29].

However, researchers and manufacturers still have time to search for solutions: as we have already said, while it is possible to quietly control voice assistants only in laboratory conditions, approaching someone else's smart speaker with an ultrasonic speaker (and even more so with 60 speakers at once) is difficult, and embedding commands to other audio recordings takes time and effort that hardly pays off.

2.8. DolphinAtack: Inaudible Voice Commands

To break into the hack, the Zhejiang University team used a smartphone with a homemade USD 3 add-on—a microphone plus an amplifier. By converting the voice command to ultrasound, they were able to tell the other phone to dial a number [30].

The human ear cannot hear ultrasound, but these devices recognize it perfectly. The authors in [15] called their hacking method, "DolphinAttack", which can also be used to order the gadget to go to a site infected with a virus, unlock the electronic lock, and even reprogram the car navigator. The test used 16 devices—from Apple, Google, Amazon, Microsoft, Samsung and Huawei—that is, all the market leaders.

In some cases, the attack is effective from a distance of one and a half to two meters—this applies to the Apple Watch, among others—so it is necessary to get close. Att To hack an Amazon Echo speaker that controls the "smart home", for example, you must first get into the house. However, hacking an iPhone, theoretically, presents no problems—any person in the crowd can be a hacker [31].

However, in practice, everything will not be so smooth. This is a rather interesting, but not very dangerous vulnerability. The owner, under certain conditions, will not notice anything in the case of, for example, they have disabled the sound on the same Siri or on some other voice assistant. Further, we must understand that if we are talking about an

office or an apartment, an intruder must somehow get into this office or apartment, and in other cases they will be hindered by the noises around them in order to find the phone in a bag, where various other devices are, which will inevitably also interfere with this goal.

The manufacturers have not yet responded to the researchers' statement. If the problem is confirmed, then the vulnerability will most likely be fixed, but the question is how exactly. According to experts, it hardly makes sense to reduce the sensitivity of the sensors—this can cause problems with speech recognition. In addition, some smartphones use high sound frequencies to communicate with other gadgets.

Users themselves can easily protect themselves by simply ceasing to use the function all the time; then, however, it will lose its meaning. This particular vulnerability is not the problem. The Internet of Things has yet to be streamlined. It does not matter what is used for hacking. Today, the problem with many devices is that, as a rule, data transmission between the sensor and the head unit is not encrypted. It can be assumed that an attacker can easily intercept control. Today, the entire industry, without exception, is moving towards encrypting protocols, creating this layer of security, when it becomes clear that it is impossible to intercept, and even if it is possible, then significant resources must be used.

Still, what no encryption can cope with is the existing conditions for the use of gadgets. Absolutely everyone who buys software or electronic devices, first of all, accepts the license agreement. No one reads it, but it usually says that the responsibility in the event of such failures lies with the owner. So, if someone does suffer from the "DolphinAttack", it will be their personal problem.

2.9. LightCommands or Laser Attack

Researchers from the University of Michigan and Tokyo Tele-communications University found that lasers can interfere with some voice assistants found in smart speakers and smartphones [32,33]. These devices are susceptible to malicious attacks in certain situations because their microphones interpret light signals from the outside world as speech commands. By shaking the sensors in MEMS microphones included in the Apple HomePod, Google Home, Amazon Echo, Apple iPhone, and other devices, lasers potentially compromise voice assistants. Even from a distance of a few tens of meters, the microphone recognizes the light from a low-power laser pointer as the sound of a command. The light is modified by a laser driver and an audio amplifier. The majority of commands may be authenticated without a PIN, thus there is no need to get around any additional security measures. Any smart speaker that is in plain sight can be hacked. Researchers have looked into a range of potential orders that could fool a voice assistant by being transmitted through a laser attack. In these devices, user authentication is frequently absent or disabled, allowing an attacker to utilize light voice commands to:

- Make purchases online at the device owner's expense;
- Unlock the facility's entrance doors or unlock garage doors if they are managed by a smart speaker;
- Find, open, and launch various automobiles (such Tesla and Ford) linked to the facility's Google account.
- Amazon and Google are looking into the problem to resolve it.

Defense against the laser assault. However, relatively easy steps can shield devices from this kind of vulnerability: adding a voice PIN for the voice assistant's primary commands (opening doors, etc.) and, if at all possible, mute the device's microphone when not in use (this button is installed on most smart speakers). Attackers will not be able to remotely send the command if the microphone is off; you can configure your phone to display notifications for ongoing voice assistant sessions. Beginning a session while you are away could be a hint that the device is being used without your permission.

This study is quite intriguing, even if only for the discovery of a brand-new vulnerability in cellphones and smart speakers. To create defenses, researchers like us are collaborating with the businesses that make the devices. As voice technologies advance

and gain popularity, light attacks may become more prevalent in the future; thus, it makes sense to develop protections now.

2.10. Adversarial Attacks against ASR Systems via Psychoacoustic Hiding

German researchers at Ruhr University have found that orders that are imperceptible to the human ear can be inserted into audio files to hack voice assistants [14,34]. Artificial intelligence uses speech recognition technology, which has this weakness built in. The authors refer to this hacking technique as "psychoacoustic hiding." With its assistance, hackers can conceal themselves in a variety of audio files that contain music or even birdsong, along with messages and commands that only a machine can hear. The voice assistant will be able to identify something different even when the listener will just hear the typical chirping [35]. Using an app in an advertisement, hackers can play a concealed message. As a result, they have the ability to shop on behalf of others or steal sensitive data. According to the researchers, "in the worst-case scenario, an attacker may seize control of the complete smart home system, including cameras and alarms." Attackers may take advantage of the "masking effect of sound," which occurs when your brain is processing loud sounds of a particular frequency for a brief period of time and causes you to lose awareness of quieter sounds of the same frequency. The team has been discovered hiding commands to compromise any speech recognition system, including Kaldi, the brains behind Amazon's Alexa voice assistant.

The ability to compress MP3 files is based on a similar idea: an algorithm identifies what sounds are audible and eliminates anything else to make the audio file smaller. In contrast, hackers add the necessary sounds in place of the undetectable ones. Alexa is an example of artificial intelligence that can hear and process every sound, unlike humans. It was trained so that it could obey any sound command whether or not anyone could hear it. Authorities from the Amazon corporation stated that they take security very seriously and will undoubtedly implement patches. Nonetheless, the threat is still present right now. Now, workers from numerous corporations, including Apple, are collaborating to find a solution. Users are advised to disable the content display of notifications on the locked screens of iPhones and iPads in the interim.

2.11. SurfingAttack

In a world that is increasingly moving towards more digital technology around us, internet safety and security is becoming an increasingly visible concern, but now even sounds are unsafe, as researchers have managed to hack smartphones using ultrasonic waves that pass through the table on which the smartphone lay. This was achieved by a team from the University of Washington in St. Louis. The project participants sent ultrasonic waves to trick the voice assistant into performing various tasks. For those who do not know, smartphones are capable of detecting such sounds/vibrations at frequencies inaudible to ordinary people. Thanks to this, the team was able to make calls, take pictures, adjust the volume, and even receive passwords via text messages.

The most notable aspect of the entire process was that the device was not accessible through a public network or physical connection. Rather, they just kept it on the table. So, even keeping it on a flat surface is now a dangerous thing, including flat surfaces made of glass, wood, and metal. The researchers achieved this by attaching a microphone and piezoelectric transducer to the bottom of a table. This allowed them to generate a signal near the device to reproduce the corresponding signals. The software used was dubbed as surfing attack software, and the researchers found that 15 out of 17 common smartphones were susceptible to this exploit. The phones were from various brands and included Google, Motorola, Samsung, Xiaomi, and Apple models. Thus, the researchers concluded by pointing out methods to protect your devices from such attacks, including: an advanced solution, such as using a thicker body to cover the smartphone, which makes it harder for a hack to happen in the first place; a lid such as a tablecloth on a table or flat

surface also works against burglary; and disabling screen lock personal results on Android devices helps to prevent this.

2.12. Defending against Voice Impersonation Attacks on Smartphones

Researchers at the State University of New York at Buffalo have developed a software algorithm that protects a smartphone with a voice assistant from being hacked. According to the university, the new algorithm uses a digital compass built into the smartphone as one of the steps to protect against voice hacking [36]. More smartphones today are getting voice assistants, including Apple's Siri and Microsoft's Cortana. In addition, some communication applications, for example, WeChat, have voice control. To securely hack smartphones with voice assistants and control their functions, attackers can use records of the owner's voice commands. Modern systems are capable of distinguishing a synthesized voice. They can also distinguish commands pronounced by a person capable of imitating someone else's voice. However, there is no protection against playback of voice commands recorded on the Dictaphone. The new algorithm, thanks to the compass of the smartphone, is able to find out whether the command is pronounced live or played through the speaker. In the latter case, the speaker causes slight fluctuations in the magnetic field that the digital compass can pick up. In order for the recognition of magnetic field fluctuations to be effective, the algorithm requires compliance with several conditions. First, the smartphone must be close to the sound source. Compliance with this condition is controlled by the accelerometer data—before uttering the command, the smartphone must be brought to the lips. Secondly, the smartphone needs to be wiggled a little near the lips while pronouncing the command. Such wiggles cause additional fluctuations in the magnetic field, which are already guaranteed to be captured by the compass. Compliance with this condition is also monitored using data from the accelerometer. The developers are currently testing the system. After their completion, the algorithm is planned to be finalized, and then a smartphone application will be created on its basis, available for download. It is not specified which operating systems the new application will run under.

2.13. Towards Evaluating the Robustness of Neural Networks

Researchers have learned how to create new recordings from speech recordings that have subtle differences in the form of noise, but are interpreted by speech recognition systems in a completely different way [37]. This method can be used to attack voice assistants or to protect speech from being recognized by computers and smartphones. Many modern speech recognition systems are based on neural networks. With a sufficiently large and well-formed set of training data, the accuracy of machine learning systems is no longer inferior to specialists in any field—for example, in speech recognition or pneumonia diagnosis. However, in addition to the complexity of training, such algorithms have another serious drawback. They can be vulnerable to adversarial examples—data in which changes are invisible to humans, but strongly affect pattern recognition systems. For example, researchers have learned how to create realistic-looking 3D-printed models of turtles, which neural networks mistake for a rifle, or glasses with unusual patterns to trick face recognition systems.

In [38], the authors proposed introducing small changes in audio recordings of voice that completely change the content of the recording for recognition systems. To do this, they added the desired phrase to the original recording and, using the loss function and gradient descent, in a few minutes brought the output recording closer to the desired one.

The researchers tested the created adversarial examples on a free implementation of the DeepSpeech recognition system, created by specialists from Mozilla. Testing has shown that this method has one hundred percent accuracy: in all cases, DeepSpeech recognized the one introduced by the researchers, and not the original phrase. The authors of the work presented several examples.

In this paper, we focused on the following research questions. Why are there many attacks, but no solutions yet? What are the characteristics of those many attacks? What

threat models do users have in relation to AI speakers? To answer these exploration questions, we have run our experiments to show and demonstrate the correctness of all the work. In our tests, a typical Android or iOS smartphone served as a model for a computer system with a top-notch speaker setup. In order to launch a side channel assault, we will repurpose the speaker. Certain gadgets create an inaudible signal, and when processed, those signals are delivered to the microphones in smart speakers, changing the safe phone system into a weak one. With this method, an attacker who takes over a phone's microphone can first muffle the sound before carrying out various actions, such as making a purchase. We will especially show the difference between experiments and how they are used, and how cheap they are for an attacker, as well as how AI speakers process and fulfill their requests. In this paper, we showed and proved the vulnerability of well-known voice assistance systems such as Siri and Google Assistant to laser-based audio attacks. After our experiments, we can easily say that the laser power of 5 mW is the same laser pointer, enough to take control of many well-known smart home devices and some voice-activated phones, and about 60 mW is enough to control almost all smart devices (Table 1).

Table 1. Laser power information.

Laser Power	Taking Control Over	Brand	Voice Assistant
5 mW	Many well-known smart home devices	Amazon Echo Google Home Apple HomePod	Alexa Google assistant Siri
60 mW	Almost all smart devices (smartphones and tablets)	Samsung smartphone iPhone	Google assistant Siri

Unfortunately, many recent results have shown that cybercriminals may not even need to use malicious tools to carry out an attack. Apparently, a hacker could get access to a smart assistant with a laser pointer, an audio amplifier, and a few other things. Using this vulnerability, we give our own commands to voice assistants using light at a short distance, which will assume that they are dealing with a normal voice command. It would seem that the microphones in these gadgets could have the said weakness, and this could be exploited when using things such as a laser pointer, an audio amplifier, an audio cable, and a laser current driver. By combining these things, we created a tool that would allow you to enter a recorded command into a voice assistant. In other words, attackers can communicate with audio assistants by converting light into sound and entering recorded commands through the target device's microphone. The test rig cost us a handful of dollars and managed to hack several speaker models at a distance of 87 m.

The number of public charging at airports, bus stops, metro stations and other public places has been growing rapidly in recent years. However, using such USB inputs is not safe: through them, attackers can gain access to data stored on the phone using a voice assistant. We show that the absence of a 3.5 mm jack in smartphones can be a serious hazard when a data cable is used to charge the smartphone, which is a common practice. In terms of this form of identity theft, we clearly demonstrated the stolen data obtained using the charging cord. However, despite the risks, many users continue to use public chargers. An experiment proving this was organized in our laboratory, which took place inside our university, where many smartphone users did not even notice that their electronic friend was attacked with a new attack called "Cable Attack". We assembled a charging station at our booth and offered cords for their devices to the participants of the event. About 80 percent of visitors took advantage of the offer without even asking if it was safe. They were in a lab where security issues were discussed, and probably where they were supposed to understand such things.

3. Attack on AI Smart Speakers with a Laser Beam

As security researchers, we are aware of a means to utilize a laser beam to break into an Amazon Alexa device and take over a user's account. We investigated whether sound-command-based sensory systems, such as digital home assistants, can be tricked into performing tasks using light. We have shown that a variety of digital assistants, in this example cellphones, can be controlled by light. A digital assistant could serve as a doorway for burglars to manipulate other devices in the house. When gadgets are linked to other smart home features such as door locks, garage doors, laptops, and even cars, attacks can become even more hazardous.

3.1. Directing the Signal by Converting Sound into a Laser

At first look, it seems impossible to "recode" the movement of light into sound waves because, physics-wise, they do not have anything in common. The intensity of the sound wave is encoded as the intensity of the light beam in order to transform sound waves into laser light, with louder sounds causing bigger changes in light intensity and weaker sounds corresponding to smaller changes. The laser driver is then used to modify the laser diode current based on the audio file being played into the driver's input port because, as we already know, the strength of the light beam created by the laser diode is proportional to the direction of the applied current. As a result, the laser's output light's intensity was directly derived from the sound wave.

$$I_t = I_{DC} + \frac{I_{PP}}{2} \, Sin \, (2\pi f t) \tag{1}$$

In particular, an amplitude modulator was utilized to modulate a sine wave on top of the I_t diode current (AM). Here, I_{PP} is the peak-to-peak amplitude, I_{DC} is a DC bias, and f is the frequency according to time t.

3.2. Experimental Setup

In our tests, we demonstrated how to use a regular laser pointer to circumvent voice assistant algorithms. Hence, before discussing our experimental setup, we will describe the voice commands we chose and what constitutes a successful experiment. As described in Table 2 [39], we initially chose a few speech commands (such as prompts) that voice-controlled computers can use to perform routine tasks. The most commonly used commands are:

- Hey Siri or OK, Google—wake word (normalized to adapt the general loudness to pick up the microphone, no device specific calibration);
- Do you hear me? —the foundational level of our experiments. This was done to ensure that everything functions properly and replies;
- What Time Is It? —Our experiments will be based on using this command because it only needs the device to correctly detect it and be connected to the Internet in order to restore the current time;
- Set the volume to down . . . —This voice command is crucial and dangerous since it will be used by an attacker as their first attempt to avoid garnering the attention of the target's legitimate owner;
- Purchasing . . . —By using this command, it will be possible to demonstrate the purpose of a hacker who will order a variety of items on the owner of voice assistants' dime. As a result, a potential attacker may easily wait for delivery close to the delivery address and take the purchased items.

Using the audio recording system, we have created audio recordings for each order. The recordings were then adjusted for overall volume and normalized to a constant value in accordance with the corresponding testing device before wake-up words were added to each audio command (for instance, Hey Siri, or OK, Google). Following some preliminary work, we were given a dozen complete commands. We then injected complete commands

into the microphone of the device—in our case, smartphones—using the configuration we will cover in more detail later, and we saw the response of the device. In addition, we would like to point out that no machine learning algorithms or device calibration was performed or used during the re-rendering of audio commands containing voice recordings. In the course of our studies, all of the examined devices were controlled using the prepared voice commands without any further modification. We constructed a complex enabling remote interception of control of various smart devices using the photoacoustic effect produced by low-power laser radiation using many laser installations and other electronic parts. A 650 nm red laser diode linked to a laser driver was all that was needed to carry out the experiment successfully. The driver diode's constant current is raised until it can continuously output a 60 mW-laser beam. Then, the beam is pointed at the remote sensing port, which is supported by a tripod and a smartphone microphone. Lastly, using a Tektronix TDS 2012B oscilloscope, the diode current and microphone output are read and recorded.

Table 2. Voice commands setup.

Commands	Explanation	Attack Possibility	Verification
OK, Google; Hey Siri	Word that serves as a trigger or wake-up signal for a VA	✓	Optional
Do you hear me? What Time Is It?	Device preparation asking few simple questions	✓	Optional
Set the volume to up or down . . .	Receiving both audible or inaudible response from the device	✓	Optional

We direct the laser beam to the microphone ports of the devices (smartphones) stated in Table 3, starting from a distance of around 30 cm to 25 m to another part of our laboratory, using the same laser and laser driver that we mentioned earlier (because the maximum length of our laboratory is 25 m). As mentioned previously, we used the modulation port of the laser driver to convert sound to laser current without the need for any specific algorithms or equipment. Subsequently, in order to test it in a typical setting, our full installation was put up, repeating our experiment for feedback every meter, from one side of the wall to the other. The oscilloscope and other instruments were mounted on a computer system unit or a neighboring cabinet that is conveniently accessible throughout the laboratory, while the laser itself and the laser diode are situated on one primary base side.

Table 3. Test outcomes.

Device Type	Operating System	Voice Assistant	Min/Max Distance	Verification	Laser Power	Successful Attack Power	Achieved Distances	
iPhone 6	IOS	Siri	30 sm~ ~87 m	✔	60 (mW)	22 mW	25 m	Farther, the devices do not recognize the laser
Galaxy J6	Android	Google Assistant		✔		60 mW	25 m	
iPhone 8 Plus	IOS	Siri		✔		21 mW	87 m	Maximum space that we managed to stretch
Galaxy A7	Android	Google Assistant		✔		59 mW	87 m	

The tests were performed in our lab under controlled circumstances with close-up, regular mirrors. Here, we would want to discuss the outcomes of our assaults at a range with more realistic targeting parameters that could have an impact on the experiment's outcome. In general, the key issue is laser focusing and aiming to the attack point. Nevertheless, because the attack is replicated at great distances (in our instance, up to 87 m), a

large diameter lens is required in place of the one included in the manufacturing kit for accurate laser focusing (e.g., every conventional laser has its own lens inside the laser).

Lastly, laser pointing and focusing were done by hand, and a target was also placed on a different tripod in order to recreate realistic attack conditions for an attacker. The laser strength, which is equal to 60 mW, is sufficient to successfully take control of all the targets we assaulted. In addition, as we previously mentioned, we frequently encounter laser pointers with measured output powers that are far higher than the legal limit of 5 mW, at 60 mW or higher (although they are nevertheless available for purchase and use). As a result, we carry out our experimental attacks using a 60 mW laser, which is sufficient to carry out a successful attack on all tested devices, emulating the purpose of an attacker who will likely not follow the laser security standards for consumer products that we use on a daily basis (on all devices that we had available).

We also employed a number of inexpensive mirrors to precisely focus and alter the laser beam in any direction so as to strike the device's microphone port in the middle, which was necessary for a successful attack on all targets. Therefore, in order to prevent the owner from detecting any traces of an attack, we manually focused the laser before each experiment such that the size of the laser spot hitting the microphone was as small as possible. Table 4 contains an overview of our distance comparison findings. We were able to successfully send voice commands to every device we tested from a distance of 30 cm to 25 m while using a laser power of 60 mW.

Table 4. Distance covered.

Device Brand	**Samsung**		**iPhone**	
Device Name	Galaxy J6	Galaxy A7	6s	8 Plus
Device Type	Smartphone	Smartphone	Smartphone	Smartphone
Laser Power	60 mW	60 mW	60 mW	60 mW
Successful Attack Power	60 mW	59 mW	22 mW	21 mW
Distance/Meter	**Attack Accuracy**			
1~2 (m)	100%—All devices successfully attacked			
3 (m)	80–90%	100%	100%	100%
4 (m)	50–60%	100%	80–90%	100%
5 (m)	40–50%	50–60%	60–70%	50–60%
10 (m)	Attack success is only 10–30%			
25 (m)	Attack success is only 5–10%			
87 (m)	failed	0–1%	failed	0–1%

We primarily employed a 60 mW laser because, as previously indicated for successful command injection, some well-known smart speakers, in our case Samsung and Apple iPhones, only require 60 mW of laser power. This was done in an effort to reduce the cost of device installations. Smart speakers such as Google Home and Eco Plus 1st and 2nd generation are sensitive and can be attacked even with a 5 mW laser (for instance, a typical laser pointer) at tens of meters away, according to the team of Lightcommads [33].

Despite the inevitable assembly differences in the devices evaluated in this experiment, we did not notice any significant differences in how different recipients responded to laser injection. All microphones had a consistent level of behavior, responding to light as though it were sound without any receiver-specific adjustments. This evidence also supports the adaptability of our attack because, once the laser was focused and pointed, all devices responded to the commands entered without the need for individual device adjustments.

In the attacks described, all tested devices successfully recognized the entered orders after achieving a suitable point and concentration. However, as shown in Table 4, certain devices stop detecting commands after traveling a certain distance.

Table 4 provides a summary of our findings, where each order was entered into the devices indicated there. It should be clear that up to 5 m infusion assaults are frequently successful with only a single error in the understanding of certain terms, but, after 10 m, the success rate drastically decreases, and at 25 m, fruitful infusions become extremely challenging because concentration becomes a problem. These results suggest that although some groups are slightly more difficult to execute than others, the unexpected decline in execution of 10–25 m shows that the likelihood of our assault being successful does not seem to be influenced by the group's phonemes. Intriguingly, it appears that non-group factors such as the interior design of the mouthpiece, the presence of tissue covering the receiver ports, the force thickness of light hitting the device's amplifier ports, intense attention, arrangement, and climate, commotion level, AI calculations, and so on, determine the probability of success. You may have observed that speaker authentication is enabled by default on the smartphone devices that are the main focus of our experiment because of their powerful processing capabilities and single-owner usage. The tablet or phone adapts to the owner's vocal tests by articulating a few lines, then persistently listens to the microphone and collects a number of speech samples. The device's secure speech recognition system then makes use of the gathered sound to identify the owner of the device when they speak particular wake-up words like "Hey Siri" or "OK, Google." The phone or tablet will eventually start to carry out the voice order after the owner's voice successfully matches the voice.

Despite the fact that the focus of our work is not on house smart speakers, we want to draw attention to the fact that the LightCommands [33] team notes in their article that many voice recognition features in smart home speakers are by default, deactivated.

Smart speakers are designed to be used by many clients, they note, whether or not they are enabled by considerate users. As a result, they tend to treat unidentified voices as visitors and limit their speaker recognition capabilities to content customization rather than validation. They discovered that smart speakers such as Google Home and Alexa prevent voice purchases from being made by unrecognized voices (likely because they are unsure of which account the purchase should be charged on), but still allow already unheard voices to carry out voice commands that are important for health, such as opening a door. In addition, they point out that voice validation (rather than customization) is not currently available for smart speakers, which are common household smart speakers.

By using authentic voice recordings of the device's legitimate owner speaking the proper speech commands, an attacker can instinctively get around the speaker's confirmation. DolphinAttack [15] advises employing discourse mix tactics, such as joining and coordinating with phonemes from other proprietor's voice accounts to assemble orders, if such accounts are not available.

However, testing has showed that Google and Apple, which are employed to recognize speech, just evaluate the wake-up word and not the entire order. For instance, An-droid and iOS smartphones set up to recognize a female voice can successfully carry out commands where only the wake-up word is pronounced by a female voice and the rest is spoken by a male voice. An attacker only has to record the device's wake-up word in the owner's voice to get around voice verification; this can be done by recording whatever commands the owner gives.

3.3. Attack Range

As demonstrated in Figure 4, by precisely concentrating the laser, we demonstrate the first command injection attack against voice assistant systems, which may travel up to 87 m (the maximum distance permitted by our building). We set up the smartphone for long-range attacks so that it is as far away as possible and that the laser beam is pointed squarely at it. Our training facility, which is located on the second floor of one of Dongseo

University's instructional buildings and is around 87 m long, was an appropriate location for this. As the attacker will not have the chance to be close to the attack device, we execute an attack at a more remote distance where the attacker can aim from a distance.

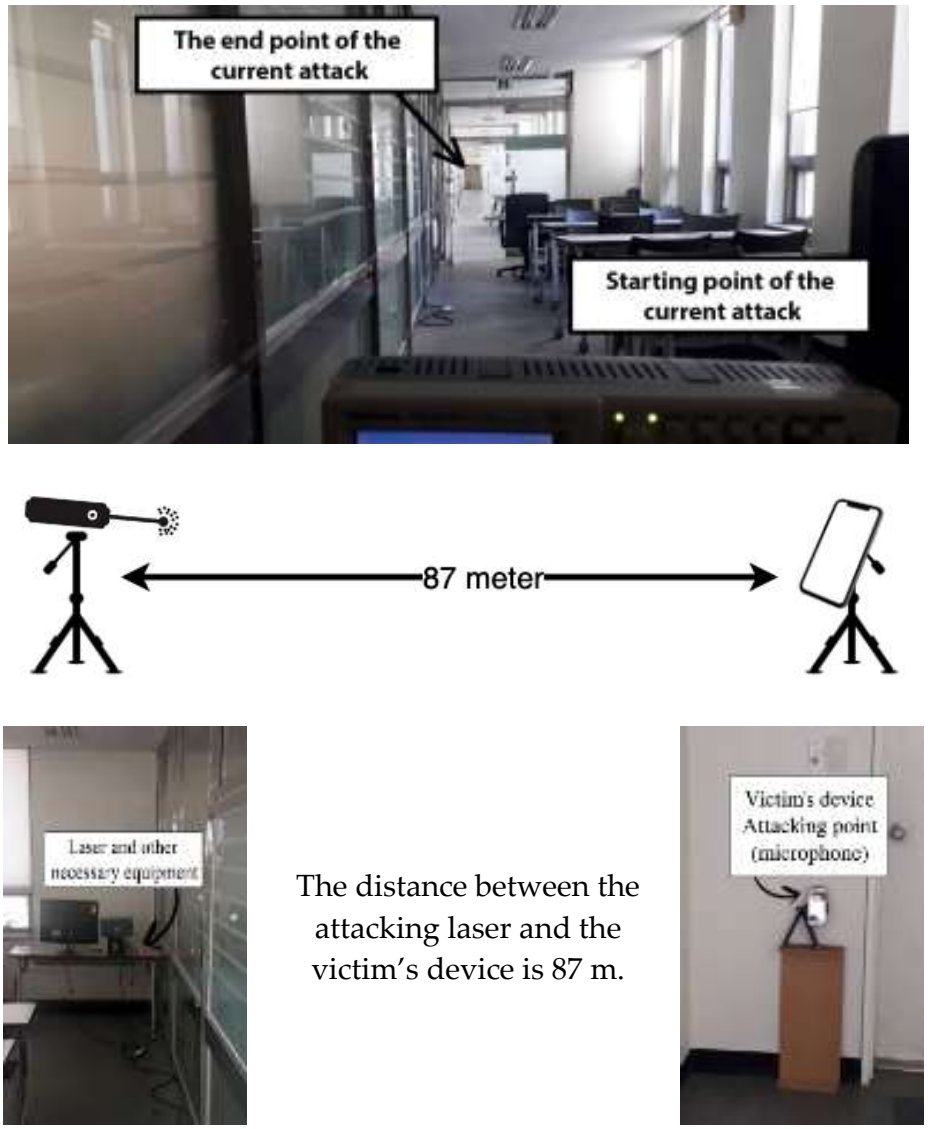


Figure 4. Experimental configuration for investigating attack range.

The laser is on the table, along with other necessary tools. We aim the laser at the target on the victim's device, which is located at the other end of the corridor, which is 87 m long. A victim's device—in this case, a smartphone—with a laser point on a sight mounted on a tripod.

3.4. Experimental Conditions

We employ a laser diode that, regrettably, lacks a telephoto lens and is quite challenging to aim, as well as a laser driver that operates a diode (which is equivalent to a laser) with the same modulation settings that we previously discussed. Instead, we utilize standard glasses and manually adjust everything to position ourselves in front of the laser, as shown in Figure 5. Later, in our university building's second level, we set the smartphone device close to a wall to increase our distance. The laser of the assailant was set up at the opposite end of the hallway. There were 87 m between the assailant and the laser in all.

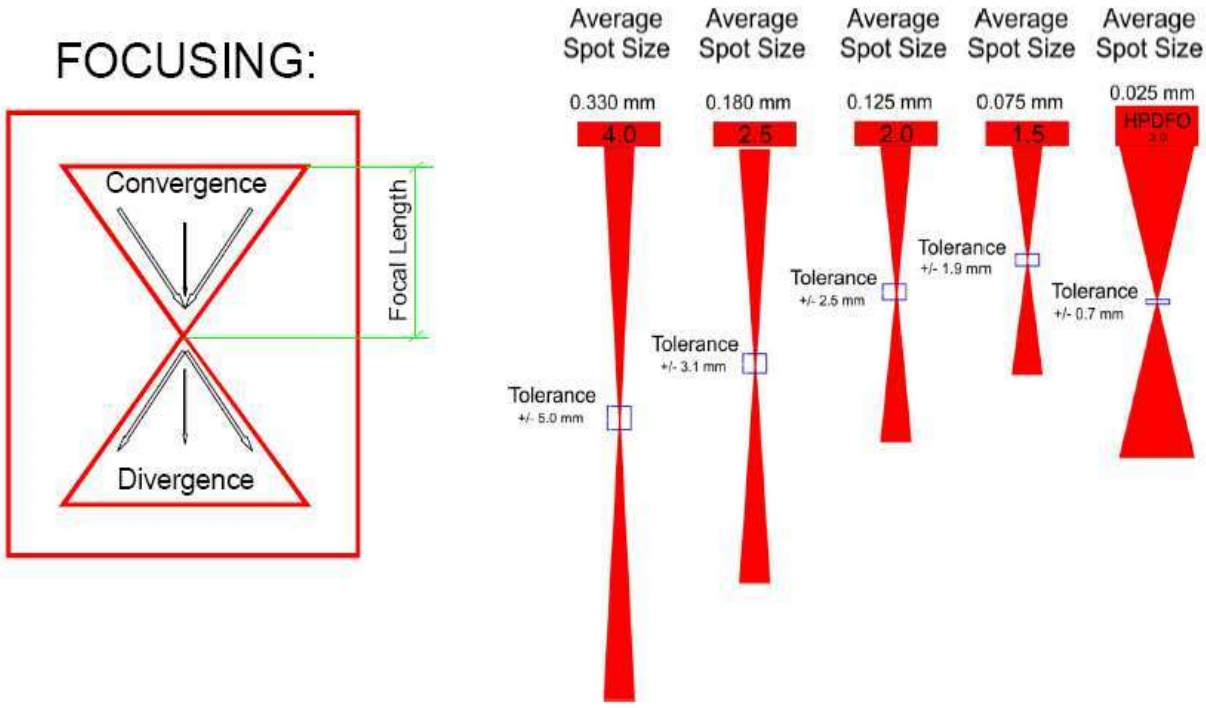


Figure 5. The lens is manually installed in front of the laser [40].

3.5. Attacking Authentication

The identical requirements, as previously outlined, have effectively been implanted into the smartphone's microphone. Additionally, we observe that the laser beam successfully entered the spoken command over a considerable distance, despite several circumstances (which resulted in some beam wavering due to the movement of the laser). Little modifications made no difference in either case; the laser beam entered the smartphone's upper microphone, which is close to the front camera, and successfully input the command without the need for any further apparatus [41]. As a result, it stands to reason that under actual circumstances, the introduction of a laser command over great distances is plausible. Early evening was the time of the experiment shown in Figure 4. Several of the modern AI speakers try to prevent the execution of private orders without permission by adding an additional client validation phase. When using Siri or Google Assistant on smartphones or tablets, the user must first open the device in order to carry out some commands (such as opening the front door or disabling the home alert system) [41].

3.6. The PIN Code Can Be Caught, Too

The client's vocally spoken PIN is inherently vulnerable to eavesdropping attempts, which can be made remotely with a laser mouthpiece (estimating the acoustic vibration of a glass window via laser reflection [42]) or through conventional listening techniques. A similar PIN is also used in the application to confirm multiple important orders (such as "open the vehicle" and "turn on the motor"), and customers frequently reuse PINs across different applications. In any case, a rise in the number of orders with PIN assurance amusingly increases the risk of PIN-snaring attacks.

3.7. Study of Hidden Attacks

Three approaches exist for the target voice assistance host to recognize the aforementioned attacks. As soon as an assault is successful, the client can first see the indicator light on the target device. The user can hear the device confirm the instruction was executed in the second step. Thirdly, the attacker can attempt to aim the laser at the objective receiver port, but the owner can see the place. The attack is being constrained by this issue (and for almost any intent and purpose of any attack).

3.8. Acoustic Secrecy

An attacker can end an attack by asking the target device to turn down the speaker level in order to overcome the problem of the user of the target device hearing the goal device confirm a voice order or ask for a PIN while the attack is in progress. The volume can be reduced on some devices to zero, while it can also be lowered to a barely audible level on other devices. Moreover, an attacker may misuse the capabilities of the device to achieve the same result. When "Don't Disturb" mode is activated for Google Assistant, updates, broadcast messages, and other voice notices are muted.

3.9. Reducing the Attack Costs

One of our objectives is to lower costs, which makes it clear that the cheapest equipment is what is needed. Yet, we only invested a relatively small sum of money (USD 5 for a laser pointer, USD 25 for a laser driver, and USD 35 for a Neoteck NTK059 audio amplifier) to do the experiment.

A device for playing audio commands is shown in Figure 6. Figure 6a,b shows an audio amplifier for amplification of audio. Figure 6c shows a laser current driver to clearly control the optical output of a laser diode. Figure 6d shows a power source for controlling the necessary electricity. Figure 6e shows a laser for injection on command. Figure 6f shows the victim's tool.

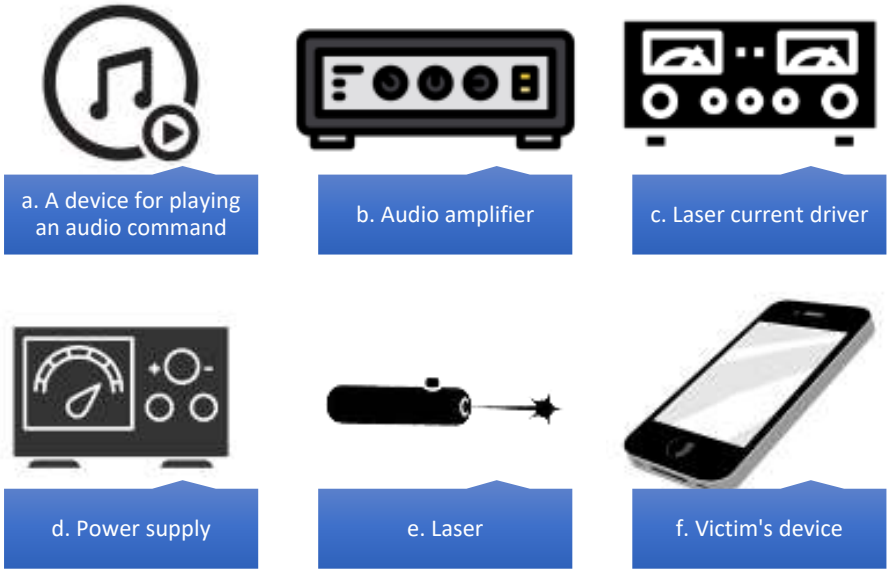


Figure 6. Equipment needed.

3.10. Laser Diode and Optics

A simple way to obtain a laser source using collimated optics is to adjust common laser pointers. Small laser pointers in particular frequently lack current controllers, and their anodes and cathodes are directly connected to batteries. Then, without a doubt, we may connect the pointer's battery connectors to the ebb and flow driver using alligator clips. Figure 6 depicts a low-cost laser pointer setup that may be purchased online for under USD 10.

3.11. Laser Driver

The most specialized component of our experiment setup is the modulation port laser current driver. As we were unable to afford the approximately USD 1500 scientific-grade laser drivers, we employed less expensive laser drivers in order to demonstrate our experiment and keep costs down. Yet, the items we utilized cost between USD 30 and USD 40 US and are easily found at any online retailer, including Amazon and Coupang. After experimenting with the setup on my old iPhone 6s, which was a 30 cm distance, it was

discovered that the laser focusing optics and an artificially constrained power budget of 60 mW for security purposes were the main factors limiting the range. Ultimately, using a low-cost installation, we have obtained a range of 87 m.

3.12. Sound Source and Test Results

Next, the attacker needs a plan to replay the recorded sound commands. We started with the commands listed below in Table 5 [39], which shows each command with its individual functions, from setting the voice assistant to make them inaudible to the owner, to trying to purchase something in online stores. To do that we used a typical smartphone that was accompanied by a Neoteck NTK059 headphone amplifier (it costs only USD 30 online). Figure 6 depicts a complete, low-cost setup that only requires wires and does not require any unique components or further programming.

Table 5. Each command's ability to accurately measure distance.

Voice Commands		Attack Possibility	Distance Attack Accuracy		
			0.5 m	5 m	10 m
Voice assistant activate word	OK, Google; Hey Siri	✔	All devices successfully attacked	40–70%	10–30%
Device preparation	Do you hear me? What time Is It?	✔			
Making device silent	Set the volume to up or down . . .	✔			
Malicious commands	Purchase attempts	✗			

4. Cable Attack, Attacking with the Help of Charging Cable

One of the items most appreciated by many mobile phone users is the presence of a 3.5 mm Mini-Jack headphone jack. For several years now, high-end mobile phones have lost the headphone jack. See Figure 7. This connector has been with us for a long time and allowed us to connect a set of headphones to our smartphone in half a second. Of course, we can still use wireless headphones, but for those who travel often and spend a lot of time on the road, this option is not very suitable. Although there are limitations, it is still possible to connect a wired headset to a mobile phone.

Figure 7. A 3.5 mm high-end mobile phone headphone jack.

We can solve the problem with the purchase of one of the accessories that smartphone manufacturers and some companies offer for additional fees, including one that is called a multifunctional splitter adapter (see Figure 8). These multifunctional splitters are designed to simultaneously connect headphones with a standard 3.5 mm jack and charge the device.

(**A**) (**B**)

Figure 8. Samples of multifunctional splitter adapters for simultaneous connection of headphones with a standard 3.5 mm jack and charging the device. (**A**) For IOS operating systems. (**B**) For Android operating systems.

The multifunctional adapter for a smartphone allows us to simultaneously support the charge of smartphones, listen to music on headphones, answer a call and talk through a microphone, and also activates Siri and Google smart voice assistants in IOS and Android operating systems, providing fast charging speed without distorting the quality of transmitted information. However, those devices can really be dangerous for people with bad intention. By this kind of new threat, we show the possibility of remote access and taking control over the voice assistants of smartphones, without letting the owner and without any additional modifications, solely with the help of what these same companies offer, which decided to get rid of the 3.5 mm jack. As many exhibitions have shown, virtual assistants will be integrated into everything from cars to refrigerators. New opportunities for virtual assistants will certainly create new risks to the privacy, information, and even physical security of users. We will not even remind manufacturers that when developing such systems, safety aspects are the first priority. The following recommendations can be given to end users.

- Amazon Echo and Google Home speakers can be "deafened" by muting the microphone using the corresponding button on the device. The disadvantage of the method: we will always have to keep in mind the need to "neutralize" the assistant;
- Purchases through Echo can either be completely prohibited or password protected in the account settings;
- Antivirus protection of computers, tablets, and smartphones reduces the risk of any leaks, preventing intruders from hosting your device;
- Amazon Echo users who have the same name as Alexa should change the word to which the assistant responds. Otherwise, any conversation in the presence of an electronic assistant will turn into a real torment.

Now, we seemed to be safe: we have sealed the cameras on laptops, covered our smartphone with a pillow, put the Echo speaker in the box. We have turned off all audio headsets, removed all microphones in the house, and speak in a whisper. However, we have bad news again: we can still be listened to.

- Physically, headphones (and passive speakers) are microphones inside out: headphones plugged into a computer input may well pick up sound;
- Some sound chips allow you to programmatically remap audio jacks. This function is not a secret at all, and is indicated in the specifications of motherboards.

As a result, while smartphones lying peacefully on the table, connected to their "legitimate" charger outlet, but in public, can be the reason for access to voice assistants. Experiments conducted by researchers have shown that with the help of ordinary chargers, we can take control of a smartphone at a distance of several meters. According to statistics, today for every inhabitant of the metropolis, on average, there are two smartphones. With active use, the battery rarely lasts even a day. Millions of people daily face the problem of a

dead phone at the most inopportune moment. Sometimes there is no opportunity to make an extremely important call. In such cases, if we do not have a power bank with us, the only way to charge our device is at a public charging station. Mobile phone charging stations for public places are another way to communicate with our customers, guests or partners. Personalization of equipment helps in achieving marketing goals of brand promotion, strengthening loyalty and shaping the consumer's belief that we care about them.

Commonly existing public charging stations are divided into two categories/types:

- A public charging power bank is what offers a variety of charging cables connected to charging stations from a service provider. They can be both paid by scanning a QR code to make a payment or in other ways, and can also be completely free;
- A public charging port is a regular USB port that allows you to charge your smartphones using your own cable (lightning cable or USB-c).

While public charging stations are a great option if our phone runs out of battery, they can also be useful tools for hackers. The fact is that USB charging stations can be designed to inject malware and steal data from everyone who uses them.

4.1. The Danger of Charging Our Smartphones with This Way

The risk of charging our smartphones this way is theoretically equal to the risk of electric shock or fire. Against this backdrop, surprising is the news that in 2019 the Los Angeles County Attorney's Office (USA) were notified of the possibility of attackers using public USB chargers to steal data, while noting that tourists should be careful when using such stations in public places [43]. Please note that the possibility of attack by external chargers was demonstrated years before this notice. For example, at the BlackHat security conference in 2016 [44], the report "MACTANS: Injecting Malware into iOS Devices Using Malicious Chargers" was presented. Even earlier, in 2011, at the DEF CON conference, researchers from Aires Security demonstrated a charging kiosk and an attack through it.

4.2. An Overview of the Connection Configuration

4.2.1. Charging Cable

Unfortunately, smartphone chargers today do not have a single standard. How convenient it would be if they were all the same. However, for a successful attack, when choosing charging cables, we have to delve into completely uninteresting technical subtleties on the principle of "whatever happens." During an attack, in order for the charger to fully perform all the functions assigned to it, it should be original. Currently, the market is flooded with Chinese accessories, fakes of different quality, often quite difficult to outwardly distinguish from the original. On the other hand, we deliberately try to save money by buying analogues, thereby exposing our smartphones and tablets to a fairly high risk of damage. Very often, not only the battery of a mobile device suffers, but also its microcircuits, such as a charge controller. Chinese manufacturers have become so carried away with counterfeiting accessories and components of modern smartphones that they have already begun to counterfeit goods from their own manufacturers. In fact, to the delight of mobile device users, today it is much less common to choose from a variety of bad exotics, as the connectors in most models are reduced to a small number of standards:

- USB Type-C is a progressive standard with a bunch of goodies: the ability to turn on a connecting cable at either end, a high power of transmitted energy, and a high data exchange rate. The most common and at the same time the most promising is the USB Type-C connector for mobile equipment, which was designed for use with gadgets with the Android operating system, but manufacturers of other mobile equipment also use it;
- Lighting is a special standard used by Apple in its gadgets. This connector is used in Apple mobile devices. It replaced the bulky 30-pin terminal and is the standard for gadgets manufactured by this company, although there is information about the company's plans to switch to USB Type-C. The connector is two-sided—you can

connect either one or the other side. This solution was used for the first time in the development of the connector, and at that time it was a breakthrough.

Cheap chargers and cables from unknown manufacturers in Southeast Asia may be of poor quality. First of all, this refers to the use of alloys of unknown composition in conductors instead of copper. This may cause overheating or limit the charging current. The same effect is caused by an underestimated cross-section of wires and poor-quality soldering (or even a crimp connection) of connectors. Simplified circuitry can lead to sub-optimal charging modes, which reduces battery life. Finally, the simplest, but most unpleasant concern is that the life of such devices, as a rule, is short, and such an adapter can fail at any time.

Sometimes, when connecting a charger to a smartphone, you can see a notification that this accessory is not supported. Such a message can be an unpleasant surprise during an attack. See Figure 9. To avoid this kind of surprise, we need to choose a charger for mobile gadget consciously. Otherwise, it not only disappointments, but also technical problems cannot be avoided, and in the worst case, there may even be financial losses.

Figure 9. The error message that the current phone does not support the accessory if the cable is of poor quality.

4.2.2. USB Type-C

Modern embedded devices use a huge number of different connectors, such as USB Type-B, miniUSB, microUSB, and so on. All of them differ in form factor, maximum throughput, and other various characteristics. The most correct solution in this situation would be to minimize the number of connectors used and stop at one, "single" connector for most developments. The most promising is the use of the Type-C connector. It combines incredible bandwidth with high power. Manufacturers such as Apple, Huawei, and Sony are already introducing the Type-C connector into their designs, gradually abandoning the use of "old" connectors.

When contact is made, the configuration process starts. It occurs on the control pins (CC1 and CC2) and consists of several stages, including:

- power source and consumer determining;
- plug orientation determining;
- host and device roles defining;
- USB Power Delivery (USB PD) protocol communication;
- power supply profile determining;
- an alternative operating mode setting (if necessary).

The communication protocol is USB PD, which is responsible for alternative modes and the correct choice of the power supply scheme for devices. If you ask how not to burn the smartphone we are attacking, the answer is that it resolves the same USB PD protocol by switching power profiles. After the connection is established, the devices try to agree on who can do what and who needs how much. For safety, the "conversation" will start with a voltage of 5 V.

4.2.3. Lighting Cable

Lightning is a cable that replaced the previous 30-pin connector in 2012. Since then, most Apple wearables have been connected via an 8-pin USB Lightning cable. Lightning is an all-digital connector that does not carry an analog signal. A practical bonus: the cable has become symmetrical—it can be connected in either direction—and it is comfortable. Even if we skip the moment that the material of the Lightning wire is much denser than that of USB-C, it is necessary to consider their functional components and which chips regulate the operation of each of the devices.

- Inside Lightning is a full-fledged microcomputer that controls the charging process of the device. It analyzes the battery level at the current moment and manages the charging process;
- Several chips are involved in data transfer using the function of a cable to connect to a computer;
- Two chips are responsible for converting the incoming electric current signal into a state that is maximally adapted for the battery installed in the smartphone;
- Chip Apple technology—the plug can be connected to either side, the built-in microprocessor analyzes the position of the wire and commands the necessary contacts in the direction of voltage.

Apple devices most often come with a native, original Lightning–USB cable marked MD818Z/MA. Buying a Lightning–USB MD818Z/MA cable is both the easiest and most controversial choice because with frequent use, the original Lightning cable can be damaged near the connector. What happens when the iOS on your iPhone, iPod or iPad detects that the chip inside the USB-to-Lightning cable is not working properly? We will receive an error message. This may be a warning about an unsupported accessory or a lack of certification (Figure 9). To use such cables for a successful attack is possible only with great risk. Dealing with all the nuances of USB cables is not easy, but that is the future of the USB standard, and it is already here. If we use its best features and check the characteristics and compatibility of connected devices and cables, pay special attention to the latter and then the attack will definitely be successful. In modern smartphones, connecting our headset via the TRRS connector, which is also an analog 3.5 mm headset jack, is less common, but there is a Type-C connector.

4.2.4. Headset

Headsets allow us to make our attack through which you can listen to sounds or other audio content in which we transmitted sound commands to smart voice assistants, without distracting other people from their business. For a successful attack, we need to choose a suitable headset.

Today, there are many options for headsets on the market, including wireless ones, but wired options are still popular. For our attack, we need wired ones, since wired headsets are connected through an adapter directly to a smartphone.

A microphone—used with headsets, of course, as everything is a thousand times better than wireless headsets, the reason for which is clear as day—refers to the presence of a microphone on the wire. It is very convenient to talk on it: you do not need to hold anything, just put on the headphones, and you are done. Sending a quality voice command during an attack is very important.

It is clear that it would be naive to demand the same level of headphones for USD 14 and USD 100, but for their category, the presented models work out every penny invested

in them and already have a solid segment of fans. At the same time, some of them are often noted as a more profitable option compared to solutions several times more expensive. However, all these non-original headphones can stop working at a critical moment during an attack. Therefore, it should be noted that for a successful attack, original headphones are needed to avoid unpleasant and unexpected moments.

4.2.5. Headset Button for Smartphone

Usually, we buy a smartphone and they come with a wired headset that has buttons to control some operations. What does the button on the headset that was bundled with the smartphone do? How is our attack related to the buttons? How do we make the attack inaudible and at the same time successful?

Initially, this button is intended for receiving and rejecting a call, since we are talking about a headset. Additionally, it is often "hung" on it programmatically and performing other actions, such as, for example, when playing music in the phone player. It all depends on the imagination of the headset manufacturer. Many headsets support the following actions.

- One short press—the playing track is paused;
- Two fast short ones—turn on the next track;
- Three quick presses—previous song.
- One long press brings up the voice assistant from Google or Siri. By the way, it is very convenient.

When we listen to music, in this way we can ask the voice assistant for the time, weather, or something else, without taking our phones out of our pockets, as well as, of course, answer a call and hang up a call.

4.2.6. Working Process of the Headset Button

Buttons on the headset is so simple that almost any modern smartphone is compatible with any one-button headset. The remote control of the one-button headset contains a microphone, a capacitor, and a short circuit button. All of them are soldered parallel to each other and brought to contacts 3 and 4.

Typically, the buttons in headsets are located in the same way as shown in Figure 10, and they work in much the same way. Button No 1 is optimized for Siri and Google Assistant. Button No 2 increases the volume. Button No 3 answers or rejects an incoming call and finally Button No 4 decreases the volume.

When the button is pressed, the microphone is shunted and the resistance between pins 3–4 of the plug drops to zero. On this basis, the smartphone understands that the button has been pressed. The capacitor, on the other hand, serves to smooth out the click that occurs when the button is pressed. In addition, it is by the presence of a capacitor that some smartphones determine that a headset is connected to them.

The voice assistant is called by holding the button until a characteristic signal appears—Google or Siri "beep" (Figure 11). When playing sound or video, the button performs the pause function, as well as when recording to the voice recorder.

4.2.7. Adapters

Dongle Adapter

In this section, we are going to review a Chinese sound card, or rather a Lightning adapter, to a 3.5 mm jack. We show it to compare difference between dongle and splitter adapter.

Digital Part of the Adapter

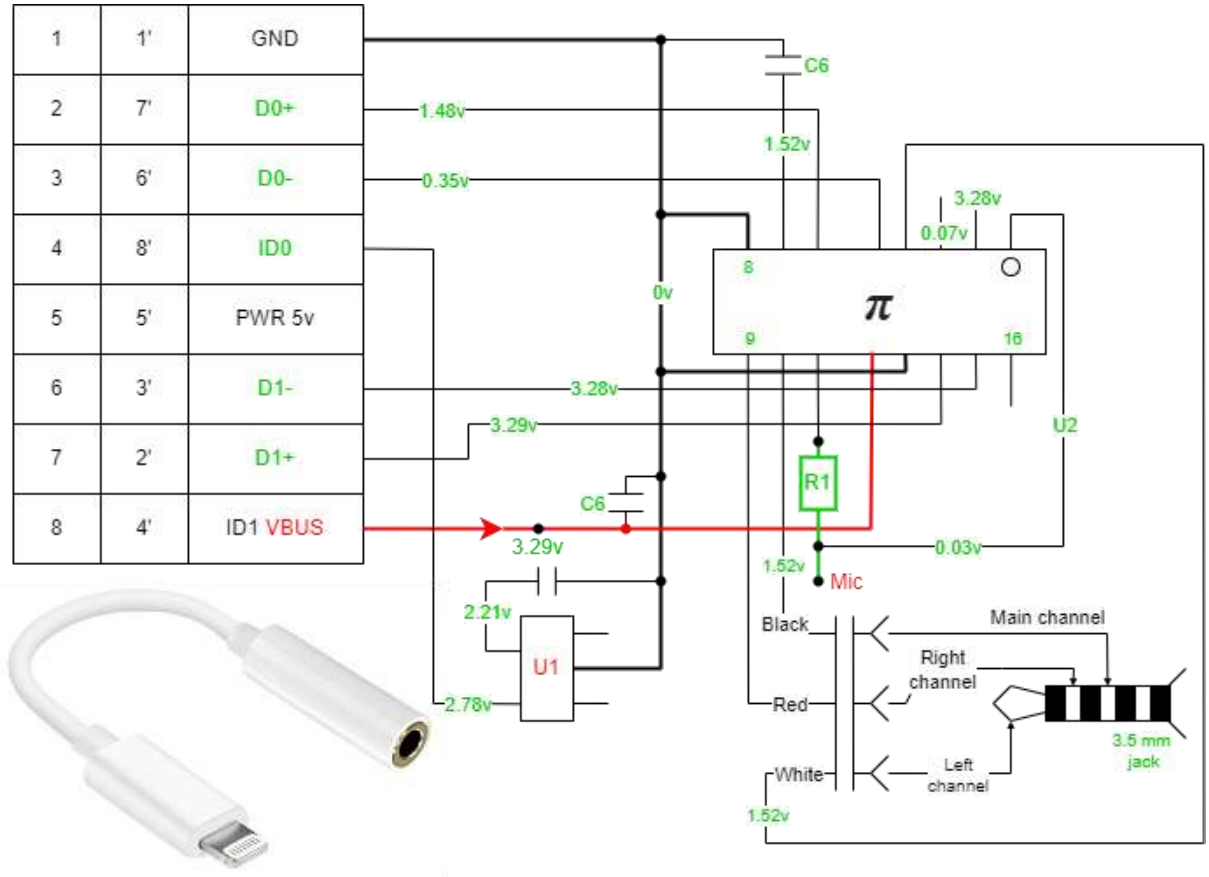
The circuit is located in the plug housing where one side is a 16-pin microcircuit, and on the other side a 6-pin microcircuit.

- The 16-pin chip is responsible for the interface and the DAC (digital-to-analog converter);
- The 6-pin chips are responsible for identification.

Figure 10. Typical headset button controller.

Figure 11. Differences between required devices and actions during an inaudible attack on voice assistants.

Scheme 1 shows that it is connected to the lightning interface using 4 signal pins. The first is a serial port and the second is another port. On the second port, as on this and other adapters, the voltage is supported by 3.3 volts and 3.0 volts. For the microcircuit, 3.3 volts are given out and for the phone itself, 3 volts are given out. VBUS is powered by 3.3 volts, just like the phone itself. After it is connected to the ID0 line, the power transfer process begins, where the phone determines that such an accessory is connected to the phone. To do this, the phone uses an identification chip.



Scheme 1. The connection of the lightning interface using 4 signal pins.

Analog Part of the Adapter

It connects to a soldered stereo headphone jack. It also has a free contact and a special place for the sealing resistor R1. This is for the functionality of the microphone, that is, for the full connection of the headset.

Splitter Adapter

Now, in this section, we will consider a splitter adapter circuit designed to connect an iPhone charger and headphones at the same time.

To begin with, it is extremely difficult to physically open the adapter, and opening it, intending to then assemble it back, is impossible in principle. We carefully unsolder the plastic, after which you gradually release the electronic circuit of the adapter from an unrealistic amount of metal and foil. Some protective plates are held on by latches, others are soldered, and others are planted on solid glue. It should be noted that if the design were simpler, there would certainly be those who wanted to solder something into the circuit: for example, you can add a separate audio output, an undertaking without hope of success—firstly, because of the design of the case, and secondly, due to the fact that even an experienced master will not understand the scheme (and, yes, there are chips, several chips)—unlike a USB cable, where there is only one. It is likely that the chips work in pair with the Lightning controller on the device and are also responsible for remapping contacts.

Lightning is "all-digital," says Apple, but guarantees analog audio output through an adapter. This means that a digital-to-analog converter is mounted in the adapter. The adapter can be purchased in two versions, original and Chinese, but we will talk about these in more detail below. Inside is a board with several parts that respond to several parts. As shown in Scheme 2, the "Lightning" part connects to the phone, the charger part connects to the bottom as shown, and the headset connects to the headset part on the top right of the diagram. The charging part, as shown, is nothing special, as is usually

understood where the two ID lines are combined and sent to the ID chip, which in turn relays information about the chargers to the phone and tells the charger that voltage can be applied. As you can see, there are only two contacts. One of them is responsible for identifying where it connects to the chip via ID1. This is done so that the identification process takes place on both sides, regardless of which side was connected. The second connector, as we noted, is for a headset. It first produces the same identification for the chip and then goes to the phone. However, we found that in some non-ordinary adapters, identification is on line 2 from the bottom, and voltage is on line 1. If you look at it like that, then there is nothing supernatural if they combine the connector with the Lightning adapter, and it seemed to us that the phone would issue this USB on line 2 to D1. We assume that they are identified independently of each other, because it includes additional wires D−1 and D+1, where one of them is responsible for data and the other for identification. However, in any case, this device with this adapter will work intermittently, because the Chinese devices are different from the original device and such devices are not used for our attack, as the original headset uses a different coding scheme, and it will work. From a non-original device, only 5 volts is supplied to the phone from the charger. Data lines go accordingly to connect computers for data transmission, and so the data transfer will not work with a non-original device.

Another device we tested, unlike the Chinese adapter, has data lines—that is, you can transfer data for synchronization—but as you can see in the diagram, it has other difficulties. This device has pin 7 in the headphone jack connected to a common wire. This suggests that there is a special chip to identify this device, and it seems to us that this is just a marriage, since all the devices were different from each other (Scheme 3).

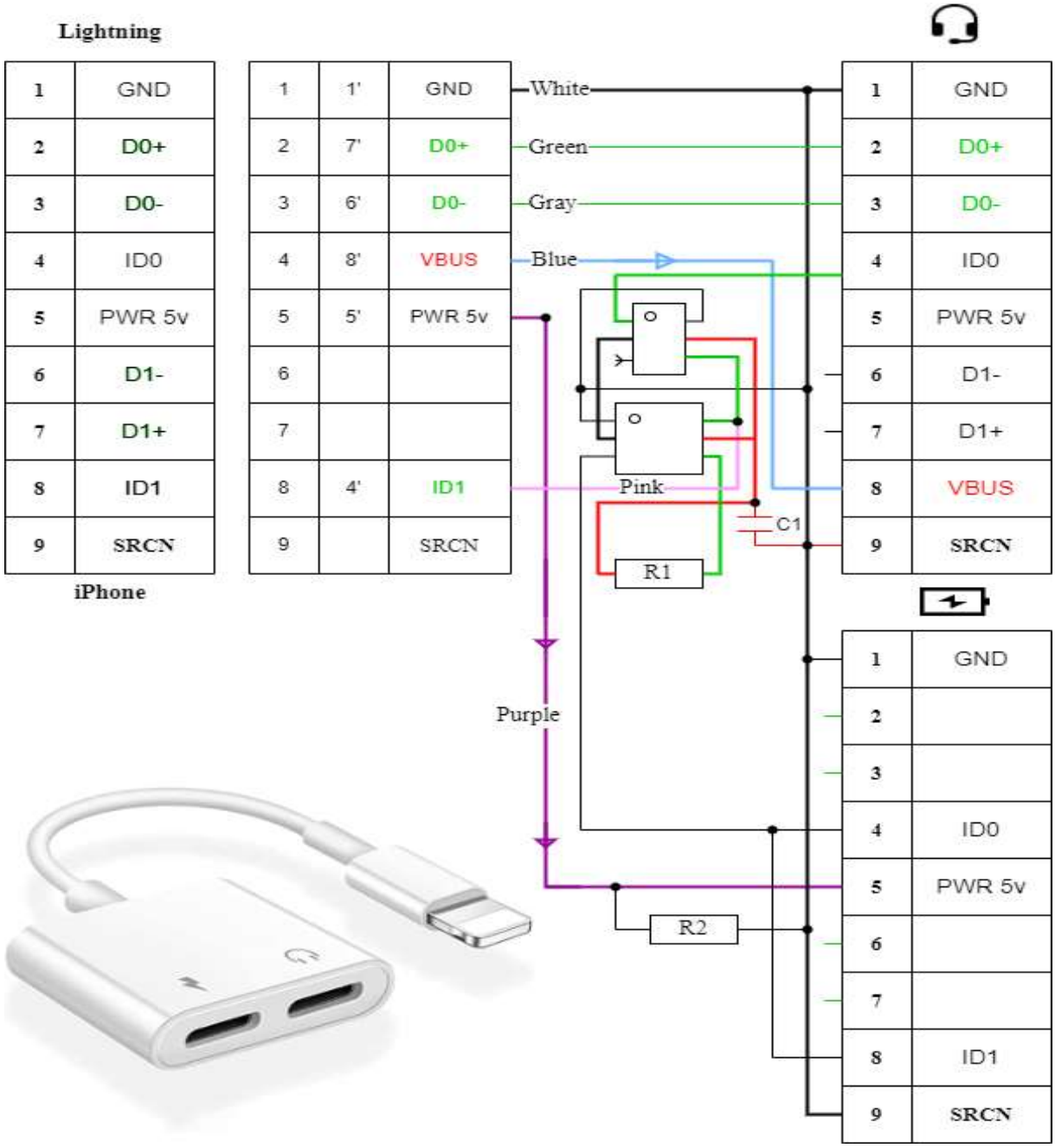
As an adapter for charging, it works without a doubt, but unfortunately, such an adapter is of little use, because ID1 of this device goes through EX and is powered by D0.

4.3. Attack Motivation

To inject inaudible sounds to voice assistants, many attacks use one or another device. See Figure 12 below. For example, the attack with ultrasonic signals [15,45], exploits the non-linear nature of MEMS microphone circuits in the attack. To do this, they used a piezoelectric transducer, which was attached to the surface of the table, and the audio signal is also converted to the corresponding ultrasonic waves. This is far from the first such attack, and not the first-time specialists have used ultrasound in this way. The most striking example of this kind is DolphinAttack [15], which was also aimed at deceiving voice assistants through "inaudible" ultrasonic commands. We can also recall similar studies of BackDoor [19] and LipRead [21], as well as a similar attack technique, Light Commands [33,46], which also allows us to silently issue commands to smart devices, but does not use ultrasound, and instead a laser for this. However, unfortunately, all these attacks require certain devices and work to be carried out on them. To reduce all these inconveniences, we decided to demonstrate our attack, which costs almost nothing and requires almost no laborious work.

The idea behind this method is to interact with the virtual assistant using normal voice commands and transmitting them through the usual everyday equipment that we use every day.

In the hands of attackers, any device, even a harmless earphone [47], can become a tool for stealing personal data stored in the victim's device, hacking smart home devices, etc. It is worth noting that the complexity of this method minimizes the possibility of it being used by someone in practice. However, the fact that the voice assistant can be controlled without obvious voice commands is a reason to reconsider the settings of our gadget. The easiest way to protect ourselves is to prevent the device from constantly listening.



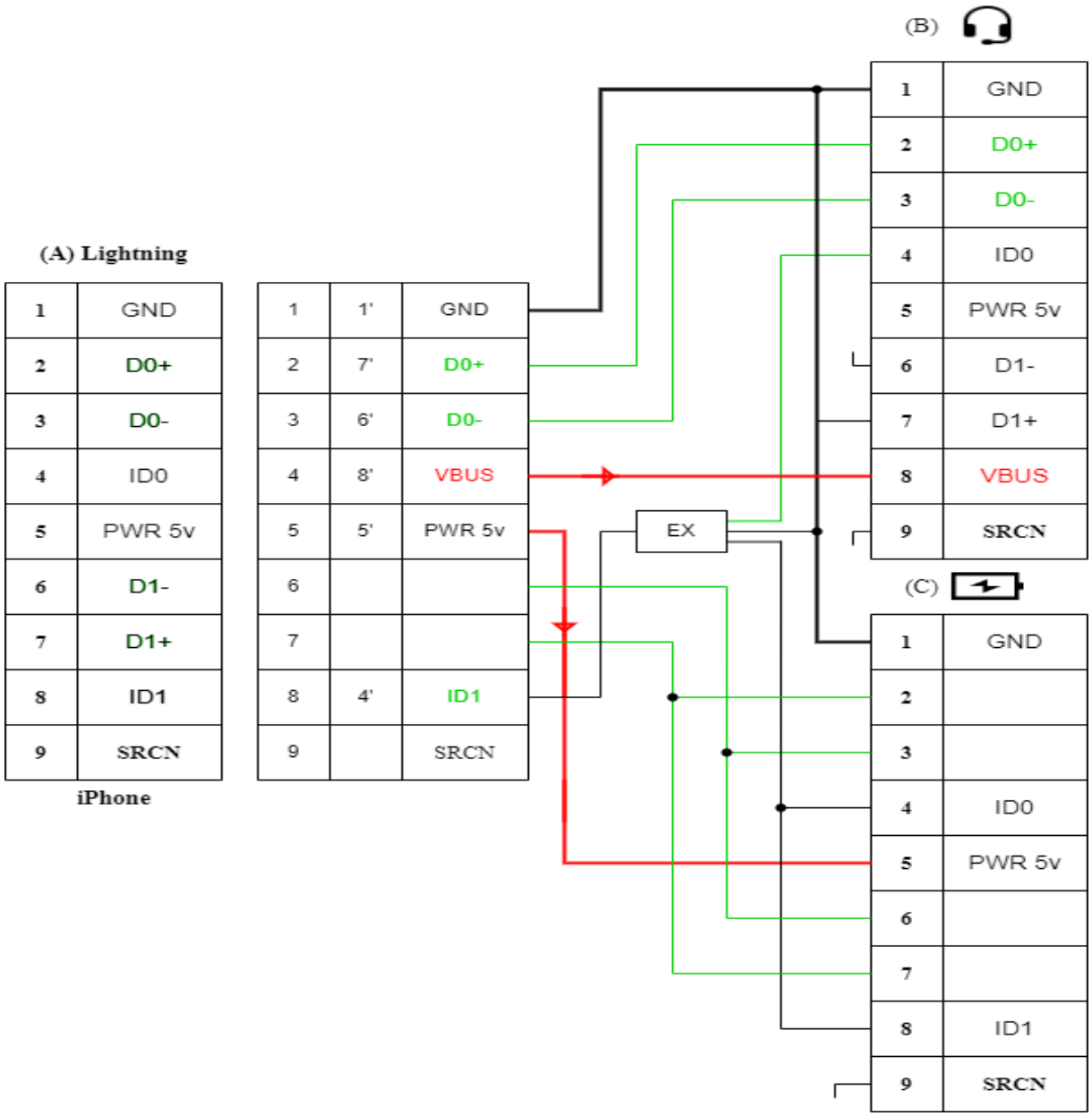
Scheme 2. Inside Scheme of the Lighting multifunctional splitter.

4.4. Attack Preparation

To implement an attack, an attacker only needs to buy multifunctional splitter adapters and connect chargers and a headset so that they support audio signal transmission. It will not be difficult for attackers to buy and connect and also hide in the right place.

Figure 13 (1) shows the connected multifunctional splitter adapter (see number 1 in Figure 13) cable for audio signals and to charge the connected smartphones. Number 2 and 3 cables, as shown, charge the smartphone and also makes it possible to connect to the headset, which is used to transmit the audio signal (see the left side in Figure 13). Number 4 in Figure 13 shows the extension of a USB charging cable until the smartphone of the victim where the other side can be connected to the smartphone. Figure 13 (2) shows the flip side of the prepared equipment which can be stick to the public charging station to make a successful attack. Attackers can hack into publicly available chargers and replace

standard cables with specially designed ones, as well as carry out attacks while smartphones are charging.



Scheme 3. Device has pin 7 in the headphone jack connected to a common wire.

4.5. Working Process

In order to carry out this attack, a splitter adapter is used for the headset and charging the phone. Then, through the headsets a voice command is sent to the smartphone, and since all commands will be sent through the headphones, the smartphone does not reproduce any external sound, so these signals are not perceived by the human ear. See Figure 14.

Figure 12. Differences between required devices and actions during an inaudible attack on voice assistants.

(**1**) (**2**)

Figure 13. A connected multifunctional splitter adapter for transmitting audio through the headset and charging connected smartphones.

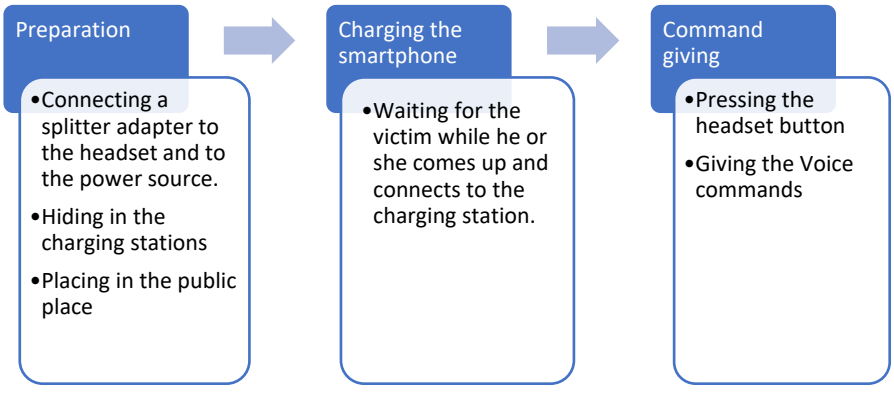


Figure 14. The attack flow.

To reach the voice assistant, the attacker would first have to visit a public charger location, be it an airport or a mall, to plug in and hide the pre-installed devices, and simply get close enough to the victim.

Our attack has been tested on popular voice assistants, including Siri and Google Assistance. By issuing a series of inaudible commands, we were able, for example, to

activate Siri and ask her to start a video call on an iPhone using the FaceTime service. We were also able to convince Google Assistance to switch the smartphone to Airplane mode.

The success of an attack in no way depends on how noisy the room is where it is carried out, since the attack performs through the headset. As an example, we gave a command to Siri to put the iPhone into "Airplane" mode. Then, in the laboratory of our educational institution, the assistant performed it in 100% of cases; this can be said with certainty that there are no problems.

An attacker can use charging stations and order commands in order to get any benefit from a USB (or similar) device connected to them. The fact is that the owner, having connected their smartphone via USB, can move away a little and do other things, such as reading a book, talking with friends, or just falling asleep. On the other side of the USB cable, an attacker with a splitter adapter splits the cable into a charger and a headset. One side of the cable, which is responsible for charging, charges the smartphone and the other part, which is responsible for transmitting data through the headset, does its job to play the audio command (Figure 15).

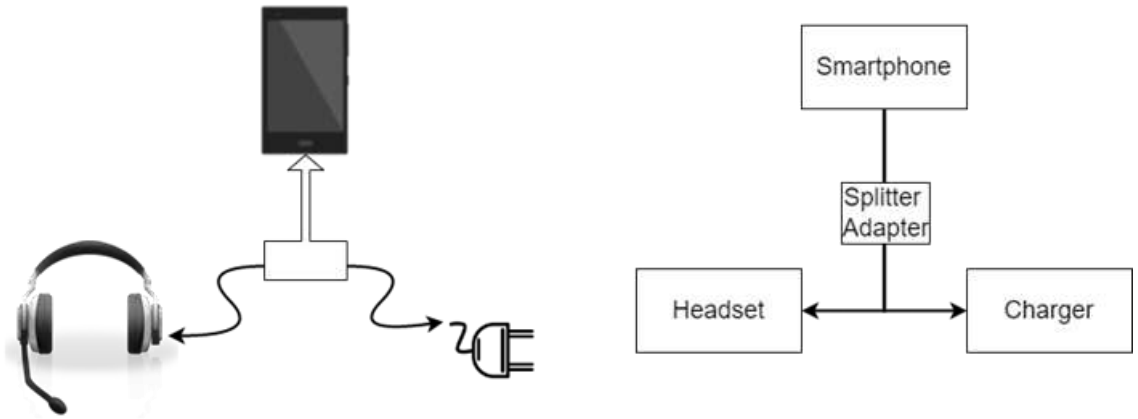


Figure 15. Scheme of how ordinary headphones are used with a splitter adapter and to charge the phone at the same time.

4.6. Attack Implementation

One of the options for a successful stealth attack is to modify the station from the inside. This can be achieved by connecting an additional connector to separate the charging cable from the charger to the headset. In addition to infiltrating officially established charging stations, attackers can prepare and install (in public places, for example) their own station under the guise of a public one. The miniaturization of electronics allows a person not only to place the desired controller directly in the charging station, but also to hide it inside ordinary cables, which will look no different from those that come with our device. No one would suspect that a hacking device is hidden in an ordinary charging station. See Figure 16.

Perhaps now such attacks can still be called exotic, but they can be implemented if the attackers have the necessary resources and desire. Especially when it comes to charging stations, let us say in a park, where there are no guards around them and the visitors themselves hardly think about the reliability of the station, and the same charging cable can be quietly replaced with a version with a surprise or successfully leave it at the charging terminal, where it will wait for the next victim. The further success of the hack will depend on various factors: the presence of vulnerabilities in the device software, the OS version installed on it, changes to the security settings by the user, etc.

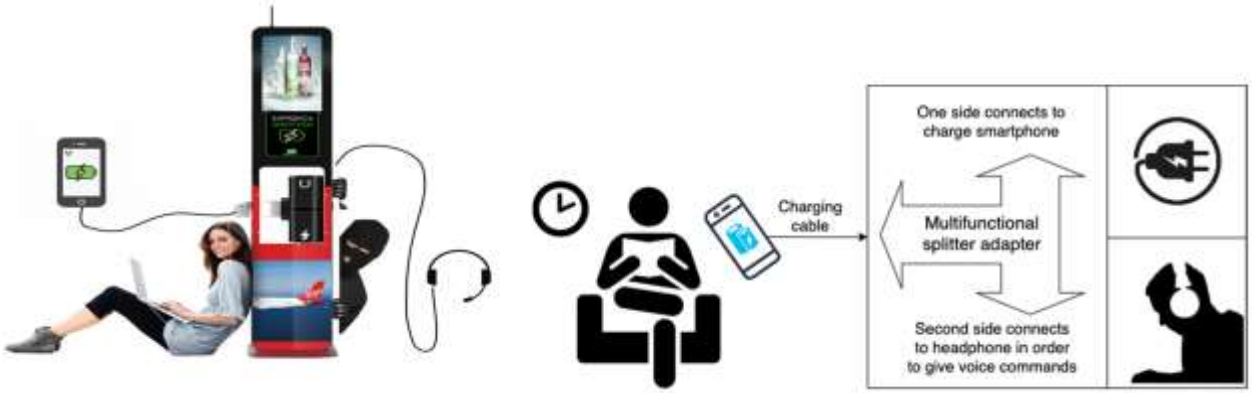


Figure 16. The Scheme of the attack.

4.7. Experiment

An experiment proving this was organized in our laboratory and study place areas, which took place inside our university, where many smartphone users did not even notice that their electronic friend was attacked by a new attack. We assembled a charging station in our university (see Figure 17) and invited others to try charging their devices. About 80 percent took advantage of the offer without even asking if it was safe. They were in a lab where security issues were discussed, and probably where they were supposed to understand such things.

Figure 17. Attack implementation.

Experiment setup: As part of our experiments, we tested a cable attack on 10 (ten) types of smartphones with IOS (Apple) and Android (Samsung) mobile operating systems. See Table 6. The experiment setup places are shown in Figure 17. We have placed hidden

equipment (multifunctional splitter adapter + headset + power supply) in the study areas of our university.

Table 6. Devices for the experiment: operating systems and device model. CableAttack components, including activation of the voice assistant (Activation), response from the device—listening to the request (Response).

No.	Smartphone	Model	OS	Voice Assistants	ASR	Activation	Response
1	Apple	iPhone 13 pro	iOS	Siri	100%	✔	✔
2	Samsung	Galaxy Note 20	Android	Google Assistant	100%	✔	✔
3	Samsung	Galaxy A33	Android	Google Assistant	100%	✔	✔
4	Apple	iPhone 10	iOS	Siri	100%	✔	✔
5	Apple	iPhone 12 Pro	iOS	Siri	100%	✔	✔
6	Huawei	Honor 10	Android	Google Assistant	100%	✔	✔
7	Apple	iPhone 8	iOS	Siri	100%	✔	✔
8	Apple	iPhone X	iOS	Siri	100%	✔	✔
9	Xiaomi	MI 8 Lite	Android	Google Assistant	100%	✔	✔
10	Samsung	Galaxy S9	Android	Google Assistant	100%	✔	✔

Attack performance: To evaluate the performance of our attack, we activate the voice assistant (Siri on IOS systems and Google Assistant on Android systems) by pressing the button of the connected headset to the smartphone. After activating the voice assistant, we enter voice commands on each victim's smartphone and repeat the experiment 2–3 times while the owner of the smartphone is busy with their own business.

Experimental setup: A CableAttack attack is being evaluated on the smartphones listed in Table 6. Victims' smartphones are powered with DC power and are charged with standard Lightning or USB-C cables.

Experiment results: The results of our experiments are shown in Table 6. In the last two columns of Table 6, we list the successful penetration (ASR) resulting from our attack. Fortunately—or unfortunately—all victim smartphones are vulnerable to our attack, as no hardware constructs are required. With a CableAttack attack, voice assistants on every smartphone of a victim can be compromised with a 100% success rate, surpassing most silent commands in use today.

We recorded all smartphone victims and warned the owners that we were experimenting on their electronic friends. We also warned that charging in public places is very dangerous. It is worth noting that the attacker can further improve this attack and may come up with more dangerous ideas.

Getting a response: To evaluate the eavesdropping attack on our response to our request in our attack, we issue live voice commands through a connected headset, not a computer. At this time, the victim smartphone accepts the issued voice command as its owner and at the same time gives its answer to our request. We issue these verbal commands while close to the charging station and the victim's smartphone. We then eavesdrop on our response through a headset connected to the station. For all victims' smartphones, CableAttack and eavesdropping have the same value because they have the same connection system, regardless of different smartphones. However, in order to avoid direct contact with the owner of the smartphone, it will be possible to prepare recordings of voice sound commands and play them directly from the device.

5. Discussion

5.1. Laser Attack

5.1.1. Programmatic Approach

As noted above, an extra layer of authentication can be effective in mitigating the attack somewhat. Alternatively, in the event that an attacker cannot overhear the device's response (for example, because the device is far behind a closed window), the VC system asking the user a simple random question before executing the command can be an effective way to prevent the attacker from successfully completing the command. Note, however, that adding an extra layer of interaction often leads to a decrease in usability, limiting user acceptance.

Manufacturers may then try to use sensor fusion techniques [48] in the hopes of detecting light-based commands. In particular, voice assistants often have multiple microphones that must receive the same signals due to the omnidirectional nature of sound propagation. Meanwhile, when an attacker uses one laser, only one microphone receives the signal, and the others receive nothing. As such, manufacturers can try to mitigate the attack presented in this article by comparing signals from multiple microphones, ignoring the input commands using a single laser. However, attackers can try to circumvent such comparison countermeasures by simultaneously shining light at all microphones of the device using multiple lasers or wide beams. We leave this task of implementing such protections and researching their security properties for the future.

Finally, light commands are very different from normal sound commands. For touch devices such as phones and tablets, touch-based intrusion detection techniques [49] can potentially be used to identify and then block such irregular injection of commands. We leave further study of this area for the future.

5.1.2. Hardware Approach

It is feasible to diminish the measure of light hitting the microphone diaphragm by diffracting the membrane or using a barrier that genuinely prevents direct light beams, permitting sound waves to sidestep them. In evaluating some of the literature on proposed amplifier plans, the LightCommands group [33] is highlighting several suggestions to protect microphones against some sudden pressure spikes.

In addition, it will also be possible to install a light barrier at the level of the device, which reduces the amount of light falling on the microphone where the location of the opaque cover on top of the microphone is opening.

5.1.3. Hardware Limitations

Laser attacks are a light attack, and naturally all the limitations of physics associated with light. In instance, the threat model of laser attacks lies in Line of Sight (LoS) and certainly cannot overcome opaque obstacles that can be penetrated by sound. The team, LightCommands [33], have shown that even if it is sometimes possible to attack devices covered with fabric, it can be assumed that for microphone ports covered with fabric, the thickness of the cover can prevent successful attacks. See Table 7.

Table 7. Protection way against laser attack.

The Way	Definition	Materials
The Light Diffraction	It changes direction of light	Hologram Light entering a dark room; Crepuscular Rays
Light Absorption	Light is absorbed	Coal Black paint; Black perfect
Light Reflection	A beam of light reflects off a smooth polished surface	Glass Mirror; Acrylic Mirror; Can Lids
Light Barrier	Light stops passing	Plastic Metal
Fabric	Light stops passing	Net Polyester Silk

Moreover, in contrast to sound attacks, careful aiming and LoS is necessary for laser attacks. We show in our experiments that in using an ordinary cheap lens it is possible to attack distant equipment. However, if the attacker has a telescope available, then the attack will be much easier than without it.

Eventually, while AI smart speakers visible through windows are often directly accessible, the situation is very different for mobile devices such as tablets and smartphones. This is due to the fact that, unlike AI smart speakers, these devices are mostly mobile and move from place to place, so an attacker must aim and enter commands as soon as possible. Consequently, for such an attack on devices, laser attacks can be particularly challenging with picking up higher power lasers and accurate targeting.

In conclusion, a laser attack is an attack that uses light and a laser to enter commands into voice-controlled systems from long distances. We have shown that with the help of a laser it is possible to carry out an attack on many well-known available systems with AI voice control, where the assistants are Siri, Google Assistant, and Alexa, receiving successful command injection over long distances of more than 87 m, despite the weather and noise conditions.

5.1.4. MEMS Microphones Respond to Light

Laser attack is possible due to the function of the device microphones in gadgets. Most modern microphones built into smart electronics are microelectromechanical systems (abbreviated as MEMS or MEMS). These are miniature devices in which electronic and mechanical components are combined into one intricate design.

MEMS devices are mass-produced using the same technologies as computer chips, mainly from the same material, silicon, and with the same degree of miniaturization: the sizes of individual devices are measured in micrometers or even nanometers. These devices work at the intersection of electronics and the physical world.

The main sensitive part of a MEMS microphone is the thinnest membrane, about a hundred times thinner than a human hair. This membrane is vibrated by sound waves. As a result, the space between it and the stationary part of the sensor increases and decreases. In this case, the membrane and the fixed base of the sensor together form a capacitor, so that when the distance between them changes, the capacitance changes. These changes are easy to measure and record, and then convert to sound recording.

The light beam can also create waves that vibrate the sensitive membrane. The so-called photoacoustic effect has been known since the end of the 19th century. Then the Scottish scientist Alexander Graham Bell (the one who patented the telephone) invented the photophone—a device that allows you to exchange sound messages using a beam of light at a distance of several hundred meters.

Most often, the photoacoustic effect arises from the fact that light heats up what it hits. When heated, objects expand and become larger, and when they cool down, they shrink to their original size. That is, under the influence of a flickering laser beam, they will change in size. You probably will not notice, but the MEMS sensor is microscopic and susceptible even to microscopic influences. It will feel and honestly transform the sound recording to be recognized as a voice command.

5.1.5. Cable Attack with Help the of Charging Cable

Without charging, modern laptops and smartphones do not work as long as we would like. Promised days of operation after some time of use turn into a matter of hours, and the user often finds us in search of a power source. Charging racks, to which we can connect a USB cable and charge the device, are now available at airports, railway stations, in modern subway cars, etc. Many, having seen the coveted connector, are in a hurry to use it: after all, the device can be discharged at the most unexpected moment.

Any USB port in a public place is a potential source of danger not only for the gadget, but for its owner [50]. If attackers upload a virus to it, they will get payment data, personal photos, documents and the content of correspondence in instant messengers. The virus can

lock the device (and hackers demand a ransom to unlock it), export passwords, or delete important data. Infection occurs imperceptibly, and the user does not even realize that third parties have gained access to the gadget.

Our attack is a new type of threat. Some people have never discovered the real facts of smartphone attacks through chargers in public places. However, we recommend people to be vigilant and, as far as possible, use their "own" chargers or external batteries. The latter are especially relevant: on average, their capacity is enough for two or three recharges of most smartphones on the market.

5.1.6. Protection Our Smartphones from This Particular Attack

- Fully charge our gadgets before leaving the house: the higher the capacity of the battery and the more economically you use the charge, the less likely you will have to use chargers outside the home;
- Be careful about charging in public places—cafes, airports, train stations, rental cars—and minimize their use if possible;
- Do not use donated charging cables. Promoters can give them out in high traffic areas ostensibly as gifts from well-known brands. Guess what you can pay for such a free gift?
- If possible, use an AC outlet for recharging, not charging stations, and carry our own cable (it takes up minimal space in a backpack or bag);
- If recharging in a public place cannot be avoided, turn off the gadget in advance and only then connect it to the charging station. Power over the cable can only be transmitted in one direction, while data transmission is possible in two directions at once. The owners of smartphones on a Windows Phone are in the most disadvantageous position—their devices turn on automatically after you connect the charger to them.

One of the main advantages of this study was to show that there is no need for a lot of extra work. Many previous attacks have consisted of acquiring appropriate equipment for further processing of voice commands, including voice assistants, in order to develop in the calculation, the perception of a normal human voice, then filtering out noise that does not correspond to it in frequency, including ultrasound. See Figure 18.

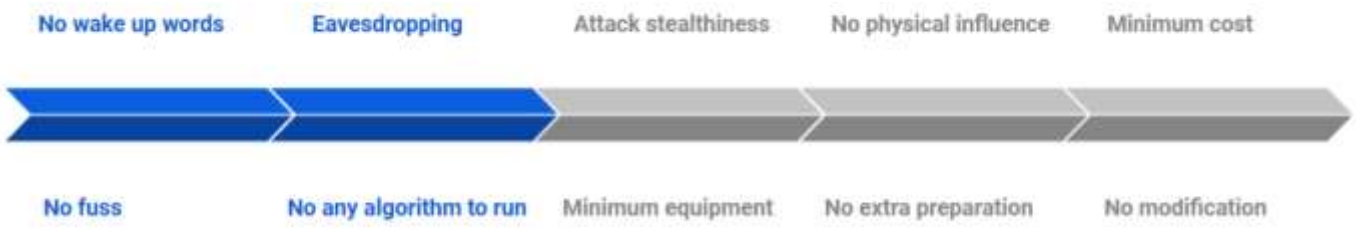


Figure 18. The advantage of this attack over previous attacks.

The fact is that "smart" speakers constantly listen to speech around and the "wake up" activation phrases from it. For example: "OK, Google" or "Hey Siri". After that, the voice assistant starts to recognize words and is transmitted to the servers.

Since our attack requires a headset connection, all outgoing sounds are transmitted through the headset accordingly.

Many previous attacks are due to the fact that they use a smartphone speaker; thus, our attack remain at an advantage.

5.2. Open Issues/Problems for the Researchers

5.2.1. Disadvantages of Modern Voice Assistants

We must say that some functions are now successfully imitated, for example, face recognition, objects, and so on. However, natural language conversation and other linguistic functions are not imitated very well yet. In general, in applied linguistics, that is, in natural

language processing, there are two peaks, machine translation and communication, in natural language.

Machine translation is finally working a little bit. Until today, it did not work well, despite the fact that both Google and Siri have been doing it for the past 10 years. The use of neural networks has now improved it, and it has become much smoother.

It does not work that way with natural language dialogue. Let us take Alexa as an example.

5.2.2. Standard Templates

If we look closely, we will see that there are a number of templates that are manually entered, or programmed. This concerns Alexa itself: when it answers, how old it is, who it is, and so on. There are a number of patterns for responding to swearing, or insults. We still have to do such things with our hands, and we cannot get away from this, because the reaction must be predetermined.

Everything else, it seems, the developers expect to do with the help of training a neural network, using examples of dialogues. What do we see when we chat with Alexa? This, in fact, can be said about Siri, and about Google, or a voice speaker from Amazon.

In addition, there is no model for voice assistants as far as we can understand.

5.2.3. Everything Will Be but Not Immediately

All of the above does not mean that modern AI-based voice assistants are futile and useless. Already, on the basis of AI, it is possible to create more or less convenient scenarios for the work of voice assistants, combining them with visual ones. It was in this logic that the presented smart speakers equipped with a display were developed—for example, we ask the voice assistant for a recipe for a dish, and it displays it on the screen.

A lot of such scenarios can already be invented, especially those related to pattern recognition—this is exactly the area where AI has made the greatest progress.

Periodically, various researchers conduct AI tests, finding out which of them is smarter, but here it is not even absolute numbers that are important, but the fact that the IQ of artificial intelligence doubles approximately every two years.

Such a progression shows that sooner or later AI will nevertheless reach the human level—but this is still a long way off. According to leading analysts, voice assistants will become the hallmark of premium consumer electronics in the next year or two, and then move to the mainstream and budget segment.

This market is just emerging and is in the early adopters' stage, but it is absolutely certain that intelligent voice assistants and voice interfaces are a promising technology that will eventually enter every home. The speed of its penetration into the market will depend on the prices of new equipment and the usefulness of those services that support voice interfaces, but in the next two years, one should not count on the massive penetration of intelligent voice assistants into consumer electronics.

The massive introduction of voice assistants into everyday life will not happen in the near future. Voice assistants certainly have bright prospects, but we are talking about the future, about a new quality level, unattainable today. Many people need, for example, a personal "social" robot, which they would like to use as an interlocutor, companion and friend. In a situation with zero physical interface, it can be just a voice from a smartphone.

However, in order to act as an interlocutor and, moreover, as a companion, we need a qualitative leap in the development of technology to see a significant improvement in cognitive function.

The voice assistant that everyone would like to have should be indistinguishable in terms of voice interaction with a human assistant. When technology reaches this level, the demand for such companion assistants will be cosmic.

Nevertheless, the usefulness of a voice assistant is primarily determined not by its intelligence, but by the quality of speech recognition and integration into various information networks, Internet services, smart city and smart home systems.

5.2.4. Tech Companies Defend against "Voice" Threats

Amazon, Google and Apple are always working to improve their voice assistants, although they usually do not get too technical. A paper presented by researchers in Zhejiang Province [15] recommends redesigning the device's microphones to limit the input signal to the ultrasonic range that humans cannot hear, or to block inaudible commands by detecting and canceling a specific signal.

The authors also suggested using machine learning capabilities to recognize frequencies likely used in fraudulent attacks and to study the differences between inaudible and audible commands.

In addition to these options for rapid improvement in algorithms of work, it will be necessary to address the safety and effectiveness of voice recognition technology at the legislative level. Currently, most countries do not have a national legal or regulatory framework for voice data and privacy rights.

Nevertheless, the first steps have already been taken. California was the first state to pass legislation restricting the acquisition and sale of consumer voice data, but it only applies to voice data collected by AI TVs.

As the number of voice use cases grows, so does the risk of fraud. Making suitable audio files is much easier and faster than cloning a credit card or copying someone's fingerprint with silicone. This means that voice data can be targeted by criminals.

5.2.5. Voice Device Protection

Now, the risk of intercepting voice commands seems to be something very hypothetical and practically unrealizable, but there are prerequisites for such a fraud, and they are very weighty. In addition, the above methods of gaining access to devices and experiments have shown us that scammers are quick to adapt to new technologies.

It is wise to follow security guidelines that will help protect your devices from being hacked. Here are some tips:

- By using strong, unique passwords for your voice devices;
- By not leaving your smartphone unlocked if you are not using it;
- By protecting voice assistant tasks related to your security, personal data, finances or medical records with a PIN. Better yet, do not associate this information with voice control devices.

Research on "inaudible" voice commands and the associated risks is still "new". Nevertheless, tech companies have realized that every new improvement gives fraudsters more and more opportunities. As more researchers shed light on the weaknesses of AI in speech recognition, the industry has an opportunity to make its products safer. If we keep track of the types of information shared with Alexa, Siri, Cortana and other voice assistants, then our data will be safe.

By studying our experiments, we can provide gadget manufacturers with methods to prevent such attacks. For example, if voice assistants receive information not from one, but from several microphones at the same time, then it will be much more difficult to deceive them with a laser. Another option is to change the design of the microphones by adding a special reflective layer.

5.2.6. No Interlocutor Model

That is, we need to understand that a virtual assistant of such a scale, which is so advertised, is a media project. It needs a screenwriter, producer, and so on. It looks like these people were not hired.

The voice assistants, for example, do not have a character. It cannot be seen. It has neither character nor any peculiarities.

Moreover, of course, the database of conversations that these neural networks were taught from seemed to be small. Perhaps the developers expected that they should first distribute this virtual interlocutor, and then they will learn from the dialogues that people will conduct with them. In fact, we, as researchers of such systems, have serious doubts

that neural networks can allow us to seriously learn dialogues. There is a huge gap in the amount of information between what a person can ask and what is in real dialogues. There will always be this gap, and therefore, in general, it is impossible to learn future questions; that is, it means that dialog scripts must also be programmed.

5.2.7. The Dialogue Depth Is Equal to One

For voice assistants, as you yourself may have noticed, the depth of dialogue is on average one, that is, it responds with a replica to a replica. It does not remember what happened, does not use parameters from previous replicas in order to form the next replica, and so on.

We think this is, in part, because developers find it shameful; they believe that neural networks will do everything for them.

Doing something with our hands in voice assistants is generally considered shameful. We know that they do not manually regulate the processing of requests and so on. Everything has to be done by artificial intelligence. It is the same story here.

The overall effect of this is that it turns out to be a rather "stupid" creature that, in general, does not answer most of the questions, but has a number of functions.

5.2.8. The Main Problem with Voice Assistants

Why does it happen? First, as we have mentioned, this project had to be conducted as a media project, but this is not the main problem. The fact is that there are two different problem statements for the information assistant.

We can do exactly what the assistant that helps us to perform some functions on the computer does. This means that we still have to program these functions, and the neural network will not help us to do this. For example, this assistant can organize something there on our computer or order tickets (as AI evangelists usually tell us), and so on.

The problem is that, firstly, there cannot be very many such functions, and secondly, it is not known whether people need them.

In principle, a person does not need an assistant on the screen. Very few people want to look at the screen and talk to it: it is inconvenient to do it in public rooms or in the office; it is strange to do it even in a family. In addition, there are buttons on the screen, and pressing them is much faster than talking.

That is, if we think about where information assistants are generally applicable, then the choice will be limited: these are the places where a person's hands are busy and it is possible to seize the voice channel, for example, in a car, or, for example, when making a phone call to a call center . . . Where it is already much easier to speak with a voice than vice versa, to press buttons (press one, press two, and so on).

That is, the applications of the voice assistant themselves turn out to be rather narrow, and this usually shows the time that this very voice assistant holds attention. In fact, little is written about this, but Siri, and any such applications from private players, from third-party companies, hold attention for about a day.

5.2.9. Not an Assistant, but a Companion

It seems that people in terms of voice communication with a computer (and text, that is, in natural language) need a personal companion with whom we can talk a lot about. Most likely, not everyone needs it. It will be useful for children (relatively speaking, from 4–5 to 13–14 years old) and the elderly. Its character must be very smart, as it must be able to maintain a conversation on fairly broad topics; it does not have to be just a search engine connector.

It is much more difficult to make such an assistant. That is why not a single one has appeared yet. Now, let us say, a talking column from Amazon Alexa has many functions, but, of course, its purpose is to help in shopping on Amazon. For this, it was released. It can perform many different functions, but it cannot maintain a free dialogue. At the same time, we personally have experience in recent years: people really want to communicate,

even if they know that a robot is talking to them. The euphoria that we are now making such assistants will most likely end up with a few companies left on the market that have invested very seriously in this, and the rest will switch to doing something else.

5.2.10. Applications on Phones Will Die, No One Will Use Them

Speech recognition will be built into almost any device that is able to receive and process this speech, where there is a narrow subject area (kettle, washing machine, microwave, and so on). Most people do not read the manual, and usually use a few percent of the functionality of their devices. If you integrate a voice assistant there, the device can be used more efficiently. Everything will work fine there, because speech recognition is already working, and in a narrow subject area, understanding can be programmed quite completely. Voice assistants will be available at ATMs, in shopping centers, and so on.

Voice assistants will be available wherever there is a mass service. Nanosemantics sells its information to large companies to provide technical support or take orders, and this will continue and develop. Soon, at the entrance to very large services, you will be greeted by a mechanical woman who will ask you what you need and walk you through the same trees of choice, which are now being used for IVR.

In the end, of course, a home companion will be made a companion of the child, who is about its level in the development of intelligence, but much more well-read. It will have access to Wikipedia, search engines and courses, it will be able to teach English, and so on. Most importantly, it will be able to maintain a natural free dialogue in ordinary language. This will not be easy.

6. Conclusions

Voice assistants are spiraling out of control. From year to year, researchers discover vulnerability in Siri, Alexa and Google Assistant programs. As it is clear from this work, hackers can send them hidden sound commands and control other people's gadgets. For example, they can order a smartphone to transfer money or dial a number, and a smart speaker to open a door.

The device for hacking voice assistants' cost is less than expected. All it takes to hack is a smart speaker within sight. Researchers have studied many potential commands that can be transmitted through a laser attack in order to deceive a voice assistant. User authentication on these devices is often missing or not enabled, allowing an attacker to use light voice commands:

- To read our recent SMS messages and make fraud calls;
- To identity theft that can be manipulated and blackmailed;
- To purchase via the Internet at the expense of the owner of the device.

This study is very interesting, if only because it is a new type of vulnerability in smart speakers and smartphones, finding work with some behind the need to develop countermeasures. Light attacks may become more real in the future as voice technologies become more identified and more important, so it makes sense to develop defenses now.

Increasingly, in public places, visitors are being asked to charge their phones, and many are happy to do so. However, we have found that using these USB inputs can be unsafe, with the threat of access to data, downloading, malware and other unpleasant consequences for the owner. We show a possible malicious impact when a regular charging cable can be used against the owner of a smartphone. When a gadget is plugged in to charge, it is essentially plugged into some other device. Thus, attackers can infect the device and get access to it. In this case, malicious chargers work in the same way as when you connect your phone to a computer. If there are any vulnerabilities in the phone's USB software, hackers can take full control of the connected phone. In order to avoid this, we advise you to avoid public charging stations or use your own USB cable for this, in which the data wires would be removed so that the cable is used only for charging the device.

New technologies including voice control systems without exception become threats day by day. Cybersecurity researchers are constantly looking for them so that device

manufacturers can secure their creations before potential threats turn into very real attacks. The purpose of this article is to demonstrate the feasibility of implementing voice attacks and prove their effectiveness at the same time. We have compared our work with many attacks to highlight the distinguishing feature. And as shown in Table 8, many attacks require costly hardware along with the same time-consuming work. Table 8 clearly shows the differences between attacks from each other. On the one hand, many attacks are as effective as others. But on the other hand, we show that with one small adapter sold in official smartphone stores of the same model, audio signals can be injected through the charging cable, which allows new interactive attack scenarios. When a gadget is plugged in to charge, it is essentially plugged into some other device. Thus, attackers can infect the device and get access to it. In this case, malicious chargers work in the same way as when you connect your phone to a computer. If there are any vulnerabilities in the phone's USB software, hackers can take full control of the connected phone. In order to avoid this, we advise you to avoid public charging stations or use your own USB cable for this, in which the data wires would be removed so that the cable is used only for charging the device.

Table 8. Common characteristics of the most attacks.

No.	Attack Name	Required Equipment	Required Actions	Inaudible	Response
1	DolphinAttack	Ultrasound speakers Function generator Power supply Power amplifier	Converting Generating Powering Measuring	✔	✔
2	Psychoacoustic hiding attack		Preprocessing audio input decoding	-	✔
3	SurfingAttack	Piezoelectrick disk Function generator Power supply	Converting Generating Powering Measuring	✔	✔
4	LightCommands	Audio amplifier Laser current driver Power supply Laser pointer Photo lens	Converting Generating Powering Measuring	✔	✔
5	Ghostalk	Cable modification Power supply Battery more than 95%	Converting Generating Powering Measuring	✔	✔
6	Laser Attack	Audio amplifier Laser current driver Power supply Laser pointer Photo lens	Converting Generating Powering Measuring	✔	✔
7	Cable Attack	Splitter adapter	Connecting	✔	

Author Contributions: S.S.A. and M.A.A.-A. contributed to the main idea and the methodology of the research; S.S.A. and M.A.A.-A. designed the experiment, and wrote the original manuscript; A.A.A.-A. contributed significantly to improving the technical and grammatical contents of the manuscript; M.A.A.-A., A.A.A.-A. and H.J.L. reviewed the manuscript and provided valuable suggestions to further refine the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (grant number: NRF2016R1D1A1B01011908).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Brannon, D. Attacking Private Networks from the Internet with DNS Rebinding. Available online: https://medium.com/@brannondorsey/attacking-private-networks-from-the-internet-with-dnsrebinding-ea7098a2d325 (accessed on 20 June 2022).
2. Diao, W.; Liu, X.; Zhou, Z.; Zhang, K. Your voice assistant is mine: How to abuse speakers to steal information and control your phone. In Proceedings of the 4th ACM Workshop on Security and Privacy in Smartphones & Mobile Devices, Scottsdale, AZ, USA, 7 November 2014; pp. 63–74.
3. Xiao, X.; Kim, S. A study on the user experience of smart speaker in China-focused on Tmall Genie and Mi AI speaker. *J. Digit. Converg.* **2018**, *16*, 409–414.
4. Kumar, D.; Paccagnella, R.; Murley, P.; Hennenfent, E.; Mason, J.; Bates, A.; Bailey, M. Skill squatting attacks on Amazon Alexa. In Proceedings of the 27th USENIX, Santa Clara, CA, USA, 14–16 August 2018; pp. 33–47.
5. Clinton, I.; Cook, L.; Banik, S. *A survey of Various Methods for Analyzing the Amazon Echo*; Citadel, Military College South Carolina: Charleston, SC, USA, 2016; Available online: https://vanderpot.com/2016/06/amazon-echo-rooting-part-1/ (accessed on 8 March 2023).
6. Zhang, N.; Mi, X.; Feng, X.; Wang, X.; Tian, Y.; Qian, F. Dangerous skills: Understanding and mitigating security risks of voice-controlled third-party functions on virtual personal assistant systems. In Proceedings of the IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 20–22 May 2019; pp. 1381–1396.
7. Lau, J.; Benjamin, Z.; Florian, S. Alexa, are you listening: Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proc. ACM Hum.-Comput. Interact.* **2018**, *2018*, 102. [CrossRef]
8. Gruber, J.; Hargittai, E.; Karaoglu, G.; Brombach, L. Algorithm Awareness as an Important Internet Skill: The Case of Voice Assistants. *Int. J. Commun.* **2021**, *15*, 1770–1788.
9. Masina, F.; Orso, V.; Pluchino, P.; Dainese, G.; Volpato, S.; Nelini, C.; Mapelli, D.; Spagnolli, A.; Gamberini, L. Investigating the Accessibility of Voice Assistants with Impaired Users: Mixed Methods Study. *J. Med. Internet Res.* **2020**, *22*, e18431. [CrossRef] [PubMed]
10. Vincent, J. Lyrebird claims it can recreate any voice using just one minute of sample audio. *Verge* **2017**, *24*, 65.
11. Wah, B.W.; Lin, D. Transformation-based reconstruction for audio transmissions over the Internet. In Proceedings of the 17th IEEE Symposium on Reliable Distributed Systems, West Lafayette, IN, USA, 20–23 October 1998; pp. 211–217.
12. Huart, P.H.; Surazski, L.K. Method and Apparatus for Reconstructing Voice Information. U.S. Patent 7,013,267, 14 March 2006.
13. Carlini, N.; Mishra, P.; Vaidya, T.; Zhang, Y.; Sherr, M.; Shields, C.; Wagner, D.; Zhou, W. Hidden Voice Commands. In Proceedings of the 25th USENIX Conference on Security Symposium, SEC'16, Austin, TX, USA, 10–12 August 2016.
14. Schönherr, L.; Kohls, K.; Zeiler, S.; Holz, T.; Kolossa, D. Adversarial Attacks Against ASR Systems via Psychoacoustic Hiding. *Netw. Distrib. Syst. Secur. Symp.* **2019**. [CrossRef]
15. Zhang, G.; Yan, C.; Ji, X.; Zhang, T.; Zhang, T.; Xu, W. DolphinAtack: Inaudible Voice Commands. In Proceedings of the ACM Conference on Computer and Communications Security (CCS), Dallas, TX, USA, 30 October 2017.
16. Jang, Y.; Song, C.; Chung, S.P.; Wang, T.; Lee, W. A11y attacks: Exploiting accessibility in operating systems. In Proceedings of the ACM Conference on Computer and Communications Security (CCS), Scottsdale, AZ, USA, 3–7 November 2014.
17. Vaidya, T.; Zhang, Y.; Sherr, M.; Shields, C. Cocaine noodles: Exploiting the gap between human and machine speech recognition. In Proceedings of the USENIX Workshop on Offensive Technologies (WOOT), Washington, DC, USA, 10–11 August 2015.
18. Cisse, M.M.; Adi, Y.; Neverova, N.; Keshet, J. Houdini: Fooling deep structured visual and speech recognition models with adversarial examples. In *Advances in Neural Information Processing Systems*; The MIT Press: Cambridge, MA, USA, 2017.
19. Roy, N.; Hassanieh, H.; Choudhury, R.R. Backdoor: Making microphones hear inaudible sounds. In Proceedings of the ACM International Conference on Mobile Systems (MobiSys), New York, NY, USA, 19–23 June 2017.
20. Song, L.; Mittal, P. Inaudible voice commands. *arXiv* **2017**, arXiv:1708.07238.
21. Roy, N.; Shen, S.; Hassanieh, H.; Choudhury, R.R. Inaudible voice commands: The long-range attack and defense. In Proceedings of the USENIX Symposium on Networked Systems Design and Implementation (NSDI), Renton, WA, USA, 9–11 April 2018.
22. Seiderer, A.; Ritschel, H.; André, E. Development of a privacy-by-design speech assistant providing nutrient information for German seniors. In Proceedings of the 6th EAI International Conference on Smart Objects and Technologies for Social Good, Online, 14–16 September 2020; pp. 114–119.
23. Jesús-Azabal, M.; Medina-Rodríguez, J.A.; Durán-García, J.; García-Pérez, D. Remembranza Pills: Using Alexa to Remind the Daily Medicine Doses to Elderly. In Proceedings of the Gerontechnology: Second International Workshop, IWoG 2019, Cáceres, Spain, 4–5 September 2019.
24. Conde-Caballero, D.; Rivero-Jiménez, B.; Cipriano-Crespo, C.; Jesus-Azabal, M.; Garcia-Alonso, J.; Mariano-Juárez, L. Treatment Adherence in Chronic Conditions during Ageing: Uses, Functionalities, and Cultural Adaptation of the Assistant on Care and Health Offline (ACHO) in Rural Areas. *J. Personal. Med.* **2021**, *11*, 173. [CrossRef] [PubMed]
25. Zhang, N.; Mi, X.; Feng, X.; Wang, X.; Tian, Y.; Qian, F. Understanding and mitigating the security risks of voicecontrolled third-party skills on amazon alexa and google home. *arXiv* **2018**, arXiv:1805.01525.
26. Saparmammedovich, S.A.; Al-Absi, M.A.; Koni, Y.J.; Lee, H.J. Voice Attacks to AI Voice Assistant. In Proceedings of the Intelligent Human Computer Interaction, Daegu, Republic of Korea, 24–26 November 2020.

27. Kuskibiki, J.; Akashi, N.; Sannomiya, T.; Chubachi, N.; Dunn, F. VHF/UHF range bioultrasonic spectroscopy system and method. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control.* **1995**, *42*, 1028–1039. [CrossRef]
28. Carlini, N. *Evaluation and Design of Robust Neural Network Defenses*; University of California: Berkeley, CA, USA, 2018.
29. Chen, G. Uncovering the Unheard: Researchers Reveal Inaudible Remote Cyber-Attacks on Voice Assistant Devices. UTSA Today University Strategic Communications, The University of Texas at San Antonio. Available online: https://www.utsa.edu/today/2023/03/story/chen-nuit-research.html (accessed on 20 March 2023).
30. Kobie, N. Siri and Alexa can be Turned against You by Ultrasound Whispers. NewScientist. Available online: https://www.newscientist.com/article/2146658-siri-and-alexa-can-be-turned-against-you-by-ultrasound-whispers/ (accessed on 7 September 2017).
31. Hammi, B.; Zeadally, S.; Khatoun, R.; Nebhen, J. Survey on smart homes: Vulnerabilities, risks, and countermeasures. *Comput. Secur.* **2022**, *117*, 102677. [CrossRef]
32. Schwartz, E.H. Lasers Can Hack Voice Assistants in Example Worthy of Mission Impossible But the Risk is Minimal for Consumers. Voicebot. Available online: https://voicebot.ai/2019/11/05/lasers-can-hack-voice-assistants-study/ (accessed on 5 November 2019).
33. Sugawara, T.; Cyr, B.; Rampazzi, S.; Genkin, D.; Fu, K. Light Commands: Laser-Based Audio Injection Attacks on Voice-Controllable Systems. In Proceedings of the 29th USENIX Conference on Security Symposium, Berkeley, CA, USA, 12–14 August 2020.
34. Albrecht, K. Amazon's Alexa Can Be Hacked Using the Sound of What? Government Technology. Available online: https://www.govtech.com/question-of-the-day/question-of-the-day-for-10042018.html (accessed on 3 October 2018).
35. Nekkalapu, E. The Backend of Inaudible Voice Hacking. Medium. Available online: https://medium.com/@Gentlemen_ESWAR/the-backend-of-inaudible-voice-hacking-60ae7641dec6 (accessed on 10 September 2022).
36. Potter, G. Study Describes How App, Soon to be Available, will Help Thwart Growing Cybersecurity Threat. University at Buffalo. Available online: https://www.buffalo.edu/news/releases/2017/06/007.html (accessed on 5 June 2017).
37. Carlini, N.; Wagner, D. Towards Evaluating the Robustness of Neural Networks. In Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP), San Jose, CA, USA, 22–24 May 2017; pp. 39–57.
38. Carlini, N.; Wagner, D. Audio Adversarial Examples: Targeted Attacks on Speech-to-Text. In Proceedings of the 2018 IEEE Security and Privacy Workshops (SPW), San Francisco, CA, USA, 24–24 May 2018; pp. 1–7.
39. Seyitmammet, A.; Al-Absi, M.A.; Al-Absi, A.A.; Lee, H.J. Attack on AI Smart Speakers with a Laser Beam. In *Proceedings of the 2nd International Conference on Smart Computing and Cyber Security. SMARTCYBER 2021. Lecture Notes in Networks and Systems, Kyungdong University, Global Campus, South Korea 16–17 June 2021*; Pattnaik, P.K., Sain, M., Al-Absi, A.A., Eds.; Springer: Singapore, 2021; Volume 395.
40. Laser Lenses, Optics, and Focus. Available online: https://lasergods.com/laser-lenses-optics-and-focus/ (accessed on 24 November 2022).
41. Mao, J.; Zhu, S.; Liu, J. An inaudible voice attack to context-based device authentication in smart IoT systems. *J. Syst. Arch.* **2020**, *104*, 101696. [CrossRef]
42. Melena, N. Covert IR-Laser Remote Listening Device. Bachelors Thesis, The University of Arizona, Tucson, AZ, USA, 2012.
43. Los Angeles County Attorney's Office (USA). USB Charger Scam. 2019. Available online: https://da.lacounty.gov/sites/default/files/pdf/110819_Fraud_Friday_USB_Charger_Scam_ENGLISH_FLIER.pdf (accessed on 8 January 2023).
44. Black Hat. MACTANS: Injecting Malware into iOS Devices Using Malicious Chargers. In Proceedings of the BlackHat Security Conference, Las Vegas, NV, USA, 3–4 August 2016.
45. Yan, Q.; Liu, K.; Zhou, Q.; Guo, H.; Zhang, N. SurfingAttack: Interactive Hidden Attack on Voice Assistants Using Ultrasonic Guided Waves. In Proceedings of the Network and Distributed System Security Symposium, San Diego, CA, USA, 23–26 February 2020.
46. Seyitmammet, A.; Al-Absi, M.A.; Al-Absi, A.A.; Lee, H.J. Attack on AI smart speakers with a laser beam. In Proceedings of the Proceedings of 2nd International Conference on Smart Computing and Cyber Security, Gangwon, Republic of Korea, 16–17 June 2021; pp. 250–261.
47. Andy Greenberg. Great. Now Even Your Headphones Can Spy on You. 2016. Available online: https://www.wired.com/ (accessed on 22 February 2023).
48. Davidson, D.; Wu, H.; Jellinek, R.; Singh, V.; Ristenpart, T. Controlling UAVs with sensor input spoofing attacks. In Proceedings of the 10th USENIX Conference on Offensive Technologies, Austin, TX, USA, 8–9 August 2016.
49. Zhang, X.; Huang, J.; Song, E.; Liu, H.; Li, B.; Yuan, X. Design of Small MEMS Microphone Array Systems for Direction Finding of Outdoors Moving Vehicles. *Sensors* **2014**, *14*, 4384–4398. [CrossRef] [PubMed]
50. Cimpanu, C. Officials Warn about the Dangers of Using Public USB Charging Stations. 2019. Available online: https://www.zdnet.com/ (accessed on 14 December 2022).