EAI Endorsed Transactions
on Security and Safety

Research Article EAI.EU

Manipulating Users' Trust on Amazon Echo: Compromising Smart Home from Outside

Yuxuan Chen[1,*], Xuejing Yuan[2,3,*], Aohui Wang[2,3], Kai Chen[2,3,†], Shengzhi Zhang[4], Heqing Huang[5]

[1]Department of Computer Engineering and Sciences, Florida Institute of Technology, USA
[2]SKLOIS, Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
[3]School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
[4]Department of Computer Science, Metropolitan College, Boston University, USA
[5]Bytedance AI lab, USA

Abstract

Nowadays, voice control becomes a popular application that allows people to communicate with their devices more conveniently. Amazon Echo, designed around Alexa, is capable of controlling devices, e.g., smart lights, etc. Moreover, with the help of IFTTT (if-this-then-that) service, Amazon Echo's skill set gets improved significantly. However, people who are enjoying these conveniences may not take security into account. Hence, it becomes important to carefully scrutinize the Echo's voice control attack surface and the corresponding impacts. In this paper, we proposed MUTAE (Manipulating Users' Trust on Amazon Echo) attack to remotely compromise Echo's voice control interface. We also conducted security analysis and performed taxonomy based on different consequences considering the level of trust that users have placed on Echo. Finally, we also proposed mitigation techniques that protect Echo from MUTAE attack.

Received on 29 March 2020; accepted on 02 April 2020; published on 07 April 2020

Keywords: Internet of Things (IoT) security, Mobile and wireless security, Security of cyber-physical systems

doi:10.4108/eai.13-7-2018.163924

1. Introduction

Nowadays intelligent voice control is becoming more and more portable and convenient, thus consistently influencing people's daily life such as shopping, working, banking, education, etc. In particular, Amazon's intelligent voice recognition and natural language interpretation service (Alexa Voice Service) harbors a full range of features/skills, enabling customers to create a more personalized voice control experience. People can command Amazon Echo, a voice control console with Alexa service, to play music, make phone calls, send and receive text messages, get information such as news, sports, weather and so on. Moreover, it can be used for smart home control, such as lights, fans, switches, thermostats, garage doors, sprinklers, locks, etc., with compatible connected devices from Wemo, Philips Hue, Samsung SmartThings, Nest, ecobee and others.

Currently, there are more than 50,000 Alexa skills (little programs to enable new functionalities) which includes smart home control, streaming music, etc., and the skills are added by companies like Starbucks, Uber, Capital One as well as innovative designers and developers. Furthermore, Alexa also supports IFTTT, a third-party service that enables IoT devices, applications and websites to interact with each other through Applets. Applets are available for a large number of websites and smart home products, allowing developers as well as users themselves to create new functionalities by connecting various social networks and IoT devices.

In order to use IFTTT applets or Alexa skills offered by the other platforms, users have to bind their Alexa account with the IFTTT on the website or their accounts on the corresponding platforms. These applets and skills can be activated after users approve the authorization requests. Then users can interact with

*These two authors contribute equally to the work.
†Corresponding author. Email: chenkai@iie.ac.cn

Alexa to trigger these IFTTT Applets or the skills. Implicitly, users have placed "root" trust on Alexa, allowing it to perform various operations on various platforms. However, the problems are two-fold. On one hand, Amazon Echo, the device integrating Alexa service, has been fully investigated and vulnerabilities have been reported by researchers, e.g., [27], [12], and [28]. On the other hand, voice based identity authentication has not been available on Alexa, so anyone, other than the user, can talk to Amazon Echo, pretend to be the legitimate user, and manipulate the trust that has been placed by a user on Echo to compromise his/her security. Even worse, Alexa appears not to be capable of distinguishing the voices directly from human or played by speakers.

Based on the other observations, researchers recently designed novel attacks against speech recognition systems like Alexa, that is, either uninterpretable [7, 16, 56] or inaudible [57] voice commands to human, but recognizable by speech recognition systems. Though seminal, these works typically require a speaker or an ultrasound transducer be close enough to the target, e.g., Echo, to broadcast the obfuscated voice. The more practical attack against smart home environments is to transmit such obfuscated voice from the outside of the victim's home, to compromise the central device–Echo.

To the best of our knowledge, there lacks an intensive study on how Echo could be remotely compromised and the consequential impacts considering the "root" trust on it placed by users. Since Echo is always listening and ready to be triggered, it is dangerous if an attacker by somehow generates legitimate commands to wake it up. In addition, IFTTT Applets and Echo skills can connect social networks and IoT devices together, which may allow an attacker to leverage the compromised Echo as a bridge to hack into the victim's "digital" life. ***Hence, it is still an open question that whether an attacker could manipulate the victim's Amazon Echo from outside, and if so, to what degree of the damages or losses such attack could cause, especially after the user trusts and authorizes wide privileges to his/her Echo?***

In this paper, we proposed MUTAE (Manipulating Users' Trust on Amazon Echo) attack, a novel approach to remotely compromise COTS Echo's voice control through other sound-playable devices at home, which could be leveraged by attackers to further control lots of smart devices, online services, etc., based on the levels of trust that users have placed on their Echo. In addition, our attack can solve the problem of how to broadcast the hidden voice commands [16] and CommanderSong [56] remotely. We summarize our contributions as follows.

- We proposed a novel MUTAE attack that can remotely compromise COTS Echo's voice control via various sounds-playable devices at home, e.g., TV, radio, camera, speaker, etc. MUTAE attack is the first approach that enables a straightforward and stable attack channel against Echo's voice control. We implemented MUTAE attack with affordable expense, i.e., HackRF One, open source GNURadio toolkit, and VeCOAX MOD-2 modular.

- We looked into 50,000 Alexa skills and 600 IFTTT Applets related to Alexa. Totally 100 of them can be leveraged by hackers, causing privacy leakage, fraud message spreading, and even threatening the safety of people's lives. A taxonomy of catastrophic consequences has been provided given users place various levels of trust on their Echo. Such study helps the security professionals fully understand the security challenges when building a complicated smart-home like environment.

- Aware of the root cause of MUTAE attack, we discussed solutions e.g., fine-grained authentication and two-factor authentication, that can help improve the security of Echo's voice control. Such approaches not only defend Echo from MUTAE attack, but also defeat most existing attacks against voice control systems.

The rest of this paper is organized as follows. Section 2 introduces the background about exiting IoT vulnerabilities, attacks against Echo, software defined radio, etc. Section 2.3 details the MUTAE attack, exploring various channels to compromise Echo. The impacts of a compromised Echo in the virtual world and the physical world are summarized in Section 4. The potential solutions that could be utilized to defeat the proposed MUTAE attack are discussed in Section 5. Finally, we present discussion in Section 6, related work in Section 7 and conclude in Section 8.

2. Background

There are several parts in the IoT eco-system [39]. For example, the storage, processor, network, sensor nodes, and human. Therefore, every part influence the security of IoT usage. In this section, we first overview existing vulnerabilities of IoT devices, including Echo. Based on our observation, the identification of bugs in protocols, code, logic, etc., always requires physical access to the corresponding devices, and typically demands a large number of manual efforts. Such observation motivates us to compromise the voice control channel of Echo, and then manipulate the connected IoT devices indirectly. Hence, we then provide background of the software defined radio technique, which will be used when introducing our attack. Finally, we briefly introduce IFTTT Applets which are related to Echo.

Table 1. Vulnerabilities of IoT devices.

Devices	Vulnerabilities	Attack difficulty	Impact
Samsung, Windows, Google, Apple devices [32]	BlueBorne	hard	control devices
August lock [50]	plaintext BLE	hard	replay attack
Wemo switch [9]	port services available	moderate	root shell
Wemo devices [49][58]	authentication bypass	moderate	root shell
Ring's smart doorbell [33]	plaintext credentials	moderate	leave Wi-Fi
Netvue HD camera [5][35][37]	stack buffer overflow	hard	remotely control
Lifx light, TP-Link camera, Nest thermostat, Linksys router, Sonos speaker, etc. [53]	WPA2 logic vulnerability	hard	manipulate data
Sony Android TV [54]	install unknown sources	easy	spy on users
Samsung SmartTV [18]	hacking tool (Weeping Angel)	complex	take over TV
Samsung SmartTV [43]	function vulnerabilities	hard	take over TV
LG SmartThinQ [11, 42]	authentication logic vulnerability	easy	take over devices
Wink/Insteon Hub [19]	plaintext credentials	moderate	root
smart home App [36]	over-privilege	moderate	remotely control
BMW [38]	Vehicle Identification Number and cross-site scripting vulnerability	hard	configure infotainment settings
BMW, Mercedes-Benz, Chrysler [45]	internet-connected vulnerability	hard	full control
BMW [6]	Bluetooth stack vulnerabilities	easy	unavailable resource

2.1. Hacking IoT devices

We investigate the vulnerabilities of IoT devices over the recent years and classify them based on the complexity and difficulty levels. We consider approaches such as protocol or code analysis to find vulnerabilities as hard, e.g., [32], due to the time consuming and manual efforts, while collecting data by a malicious application as easy, e.g., [11]. Table 1 shows the summary of publicly available IoT devices vulnerabilities, which are identified by the method of reverse engineering, code analysis, protocol analysis, functions analysis, hardware security testing, ports scan, brute force attack, etc. Most of these vulnerabilities are coding bugs, either because of careless developers [42] or intended malicious code [18], typically specific to some individual product or brand. In contrast, the logic and protocol flaws generally lead to huge and wide impacts [32, 40, 53].

Based on our observation, exploiting the vulnerabilities in Table 1 to manipulate IoT devices generally involve a considerable amount of efforts, and sometimes quite complicated. From another perspective, since most of such IoT devices support Echo, it will be straightforward to control such devices if Echo is compromised. Thanks to the lack of voice authentication, we found that compromising Echo can be much "easier" compared with directly hacking the IoT devices via the vulnerabilities exploitation. For instance, broadcasting the Echo commands by TV or radio signals can trigger Echo to operate on the commands, which will be detailed in Section 2.3. Based on our experiments and analysis, we found compromising Echo can also lead to the same effects of exploiting vulnerabilities of IoT

devices, and even more unpredictable consequences as in Section 4.

Vulnerabilities of Echo Device. Various studies have been conducted to analyze Echo via physical access. For example, Ike Clinton et al. [27] reversed the pins located at Echo's bottom and debugged the device through the pins [1]. Finally, they extracted the file system used by Echo and got the root privilege. Utilizing such root privilege on Echo's file system, Mark Barnes installed a rogue software on Echo, then he created a root shell to access over the network, so that he transferred microphone audio from the hacked Echo to his own server [12]. Certainly, such work can help understand Echo's internal myth or control one's own Echo in multiple ways, but cannot directly attack others' Echo remotely. Researchers from ISACA [28] made a theoretical analysis on a various attack surfaces of Echo, including network traffic encryption, firmware update, skill security, Alexa Voice API security, etc. No obvious vulnerabilities have been found till now, which implicitly indicates the necessity of compromising voice control channel of Echo.

2.2. Attacks against Echo's Voice Control System

Amazon Echo is designed to recognize a legitimate command based on the pre-defined pattern, i.e., wake word + command. The wake word, by default *Echo*, *Alexa*, or *Amazon*, is used to wake up Echo from sleeping mode. Certainly, users can freely revise the wake word with their preference. The command is just neutral language, related to either shopping, media control, smart home, etc. However, the fundamental problem of Echo, or other similar products, is that the voice control does not come with the voice based authentication. Due to the lack of voice based authentication, any clear and loud enough sound containing Echo's wake word can trigger Echo and put it into the command waiting state, no matter the sound is from human being or speakers. People have reported that Echo could be falsely triggered when the TV plays Amazon Echo advertisement.

Hence, a straightforward idea is to record the desired commands and play them towards Echo. For instance, malware or logic bombs on smartphone can be triggered to open a web browser and play a video on Youtube, or automatically download video/audio files from an online server and play it using a local media player. As long as the attacker-crafted video/audio files are played and valid voice commands are over the air, voice control systems, like Echo could be compromised. Researchers proposed to build obfuscated commands [16] or broadcast inaudible commands using ultrasonic carrier [57], which can be utilized to compromise Echo's functionality even stealthily. The proposed MUTAE attack in this paper can be viewed as a generic approach that can transmit their obfuscated or inaudible commands remotely to Echo, thus fulfilling the corresponding attacks even unnoticeable.

2.3. Software Defined Radio and HackRF One

Software Defined Radio (SDR) technology is to use modern software tools to control and manipulate the traditional "pure hardware circuit" wireless communication. The key idea of SDR is keeping the Analog-to-Digital and Digital-to-Analog converter as close to the antenna as possible, and letting software implements as many radio functions as possible. GNU Radio software is a free software development toolkit which provides several wireless communication blocks to implement SDR. It can be used to attack smart home systems, vehicles, and launch side channel attacks to crack encryption algorithms. Various hardware can be utilized to build SDR platform, e.g, HackRF One, bladeRF, ASR-2300. In particular, HackRF One is an open source hardware platform that provides both reception and transmission functionalities, capable of transmitting or receiving radio signals from 1 MHz to 6 GHz, with the maximum bandwidth of 20 MHz. A special Linux distribution, Pentoo, with full support for GNU Radio can be installed on HackRF one. In this paper, we use HackRF One and its default antenna ANT500 to broadcast radio and TV signals.

2.4. IFTTT

IFTTT is a free web-based service allowing developers and users to construct simple conditional logics. Such service, named as Applet, follows the pattern of "trigger" plus "action". The "trigger" can be happened in web services like Gmail, Facebook, SmartThings, Alexa Voice Service, while "action" is the operation desired on other web services or IoT devices. Even though the categories and number of Alexa skills are growing quickly, there are still a ton of services and devices that Alexa can not work nicely with. With the help of IFTTT however, users can bridge the gap between devices that do not officially work together, and thus control countless online services as well as third-party devices. There are more than three hundred thousand Applets available, among which more than 600 Applets are related to Alexa. They connect Echo with phones, IoT devices and social platforms. However, the various of IFTTT applets may have already caught people's security awareness.

[1] Amazon patched such vulnerability for Echo sold from 2017.

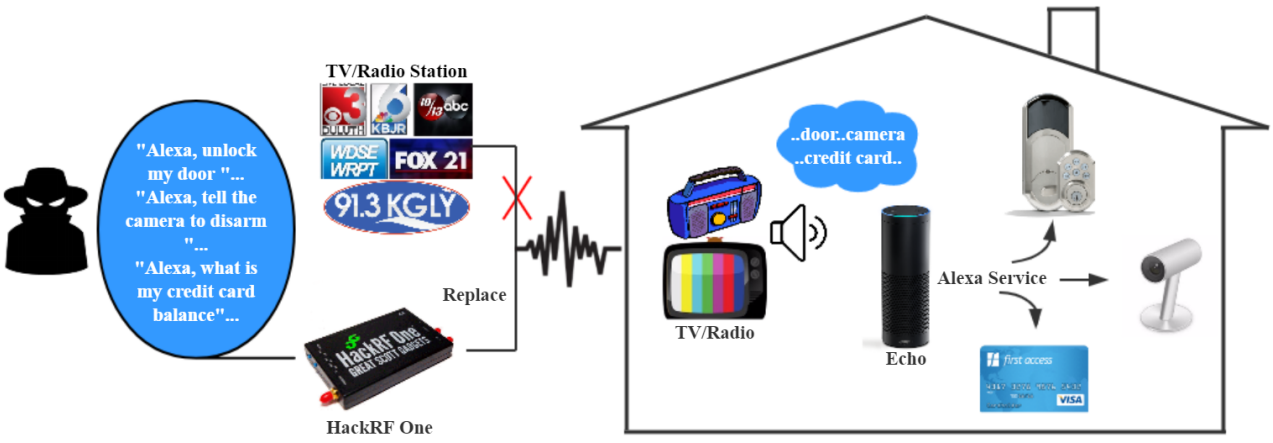

Figure 1. MUTAE Attack Diagram.

3. MUTAEattack

As discussed in Section 2.2, Echo can be falsely triggered by commands played by any speaker. Hence, a straightforward idea is if attackers can remotely control one of the speaker-equipped devices at one's home, e.g., TV, radio, speaker, etc, they would be able to compromise lots of Echo's skills. We demonstrate the potential attack scenario of Echo in Figure 1. It can be seen that attackers can try different channels to compromise the speaker-equipped devices. As Figure 1 shows, firstly, we record our command signal (*wav* for radio signal and *mp4* for TV signal). Then we transmit it by HackRF One or HDMI. The signal transfers through the air and cable to radio and TV. Naturally, our commands are broadcast by the speaker of radio and TV. Finally, Echo interacts with the IoT devices and social network. In this section, we present our work in spoofing radio and TV to transmit the commands, and investigate other sources like speaker and camera that can be used to control Echo remotely. The attack demo is uploaded on the website (`https://sites.google.com/view/mutae-demo`).

3.1. Radio Signal Injection

Radio is one of the most widely used wireless signal receivers in families around the world, which receives the radio signal and extracts information carried. According to the report [34], over 95% of American families have at least one radio receiver at home. Even though probably the percentage of people who use it daily is much lower, cars are usually equipped with radios. Meanwhile, according to the report from Morgan Stanley, millions of Amazon Echo devices have been sold worldwide [3]. And Echo has kinds of skills to control some functions of cars. Therefore, we assume lots of families could own both a radio receiver and Amazon Echo, and they are placed close enough, typically in the same living room or in the car. As the car is running with a high speed, the victim maybe cannot escape from the attack in time. In this situation, attackers can generate a stronger fake radio signal to replace the original legitimate one, and the radio receiver will decode the fake signal. If the fake signal carries the command "Alexa, disarm camera" for example, Echo should follow the command and turn off the associated camera.

The attackers can use HackRF One to produce fake radio signal. The payload to be transmitted should contain the voice commands that can wake up and control Echo. Depending on the owner's appetite, the wake up word can be set as *Alexa*, *Echo*, *Amazon* or *Computer*. Next, attackers have to identify the channel that the radio is operating in the victim's home. There can be various ways for attackers to get this information. A typical way is to obtain all available radio channels by searching around, and launch a brute force attack by transmitting the prepared attacking payload across all the available channels. Note that the searching process will not be long, i.e., typically 10 minutes, considering there are usually 30~40 channels in a city. Furthermore, if the attackers know the hobbies of the victim (e.g., through social media like Twitter or Facebook), they may be able to infer which channel the victim usually listens to. The first demo on the website (`https://sites.google.com/view/mutae-demo`) shows our attack of injecting commands into the radio. Firstly, we recorded the command, i.e., "Echo, turn on the light". We assumed that the current radio channel was unknown. Hence, we started to search for all the available channels as discussed previously. The HackRF One was located in the open air and configured to broadcast the signal. Finally, we found that when the FM signal was at 103.4MHz, our manipulated voice

command could be received by Echo, which in turn operated the command successfully.

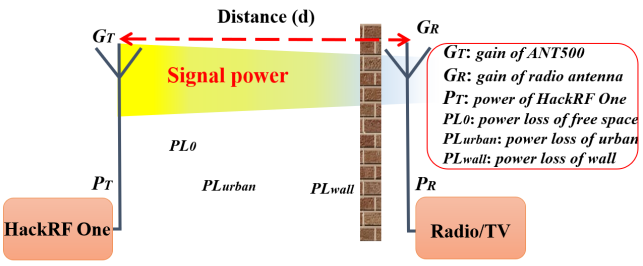

Figure 2. Power gain and loss through the transmission channel.

The attack range closely depends on the receiver performance and received power. As the performance relies on multiple factors which are out of our control, we summarize the main factors impacting the received power in Fig. 2 [41]. It can be seen that the received power P_R can be enhanced by antennas. Besides, we can enhance the power by using amplifiers. For our experiment, the signal power change through the broadcast channel is calculated as Eq (1).

$$P_R = P_T + G_T + G_R - PL_0 - PL_{urban} - PL_{wall} \quad (1)$$

where P_T is the output power of HackRF One; G_T is the gain of ANT500 antenna; G_R is the gain of receiver antenna; PL_0 is the loss of free space; PL_{urban} is the loss of urban areas; and PL_{wall} is the diffraction loss by a wall. For the calculation of PL_0, we adopt FSPL (Free-Space Path Loss) introduced in [41]:

$$PL_0 = -27.55 + 20lgd(m) + 20lgf(MHz) \quad (2)$$

where d is the distance between HackRF One and radio; f is the frequency of the carrier. Note that d and f are in meters and megahertz in Eq 2. For the calculation of the urban loss PL_urban, we adopt COST-231 Hata model [8], one of the most widely used empirical propagation model in signal processing, as below:

$$\begin{aligned} PL_{urban} = {} & 46.3 + 33.9lgf(MHz) - 13.82lgh_b(m) \\ & -a(h_m) + (44.9 - 6.55lgh_b)lgd(m) + C_m \end{aligned} \quad (3)$$

where h_b is the height of ANT500 antenna; h_m is the height of the radio antenna; $a(h_m)$ is the correction factor of the radio antenna; and C_m is the correction factor for the urban environment. PL_{wall} is strongly influenced by the frequency, structure and material of the wall, as a result, it is difficult to calculate by a formula. According to a field measurement of the $400MHz$ signal, $40cm$ concrete wall attenuates $30dB$; $84cm$ double cement board with a mid-layer concrete wall attenuates $33dB$. Since the US Federal Communication Commission regulates that low powered devices are limited to an effective service range of approximately 200 feet (61 meters) on FM frequencies, the attacker should avoid creating too strong signal to be detected.

Generally, there are scenarios manipulating radio signals to control Echo. Firstly, the target victim is at home listening to a radio program. To avoid attracting the victim's attention, we suggest integrating the attack payloads into the original radio program, so the victim may not realize the attack. However, this also depends on what the attackers ask Echo to do. For instance, if the attackers command Echo to read Washington Post News, the victim will definitely notice something getting wrong. In particular, we can first receive the radio program that the victim is listening to. For example, we record advertisement the channel always broadcast. Therefore, we can synthesize it with the attacking commands as an attack *wav*, and then transmit the manipulated program immediately. Obviously, the success rate would be lower because there will be kinds of noise accompanying the payloads. Secondly, if no one is at home, usually the radio is turned off but mostly Echo is still on. In this scenario, the success of this attack depends on whether the radio is powered through a smart switch like Wemo Switch. If so, attackers can leverage existing vulnerabilities (e.g., [4]) in those switches to remotely turn on the radio, and transmit the attack payloads on either the speculative channel or all available channels.

Sometimes it is difficult to know whether the attack is successful or not, since the radio and Echo are inside the home, while the attackers are typically outside. Attackers can actually rely on some noticeable phenomenon to convey if the attack succeeds or not. For example, they can change the light illumination a little periodically to indicate a successful attack (still by injecting voice commands to Echo), without attracting victim's attention. This can also be used to identify the channel that the victim's radio is operating on.

3.2. TV Signal Injection

In order to implement similar attacks on TV, we need to first understand the signal transmission of TV systems. In North America, TV networks (like ABC, NBC, Fox and CBS) use Advanced Television Systems Committee (ATSC) standard to broadcast digital TV signals over the air, coaxial cable and satellite networks. Once the TV receives the signal via either corresponding antenna or cable, the TV tuner can convert the signal into audio and video contents that can be displayed on the TV screen. Below, we describe the ways we hack both antenna TV and cable TV.

Antenna TV. Firstly we try to compromise the TV signal transmitted over the air. Similar to the radio attack, the goal of our TV signal injection attack is to completely replace the video contents of one specific channel on victim's TV with our video stream containing several valid Echo commands. The second demo on the website (`https://sites.google.com/view/mutae-demo`) shows our antenna TV signal injection attack. For regulation concerns, our experiments are restricted within the range of our lab, without interfering other legitimate TV watchers at their home. Similar to the radio signal injection approach, we still use HackRF One attached with ANT500 antenna to broadcast a fake TV signal. We setup *LG 43UF6400* TV connected with a 50-mile range antenna to receive the TV signal. We configure the volume of the TV at 20%, and place the Echo six meters away from the TV. The HackRF One is placed around 4 meters away from the Echo in the room. We record the attacking voice command (i.e., "Alexa, turn the bedroom light on") in *mp4* format and then convert it to Transport Stream (*ts*) format used by HackRF one for over-the-air ATSC broadcast.

Then we have to figure out the victim's favorite TV channel, and inject our signal to that specific channel to replace the original TV program. In our experiment, we assume that before the TV was turned off, it operated at the Channel 48 (operated by HBO, provided the victim is the fan of Game of Thrones, etc.), whose frequency range is 674 MHz - 680 MHz (6 MHz is normal bandwidth for ATSC standards). For the situation that the TV channel is unknown, similar to brute force approach used in radio signal injection can be reused here. Hence, in our experiment, we directly set our central frequency to be 677 MHz and bandwidth to be 6 MHz in the GNU Radio software to broadcast the fake program and replace the original HBO program.

Finally, we start the HackRF One to transmit our recorded program on the channel used by HBO. Once the TV is turned on (preset at the HBO channel), it will play the program we are broadcasting. So the recorded commands are played by the TV and recognized by Echo to operate. The success rate for such attack and range highly depend on the hardware used.

Coaxial Cable TV. Besides hacking Echo with TV signal over the air, physical injection of coaxial signal towards TV is also feasible. In particular, the attackers can cut off the dedicated coaxial cable outside the victim's home and inject their manipulated signal over the end to victim's TV. The last demo on the website (`https://sites.google.com/view/mutae-demo`) shows the coaxial signal injection attack. Due to the regulation concerns, we simplify our experiments without cutting the cables. Instead, we assume one end of the coaxial cable is just connected directly to the TV inside one's home, and the other end is outside. We record the attacking voice command (i.e., "Alexa, turn the bedroom light on") in *mp4* format.

In order to transmit such video content over coaxial cable, we use the Pro Video Instruments VeCOAX MOD-2 (VeCOAX), an HDMI to coax modulator, which can convert HDMI output to coaxial signal. Therefore, we used a laptop to play the recorded video, and sent it to the VeCOAX through an HDMI. The VeCOAX can be configured to convert the incoming HDMI signal into commonly used digital TV standards, such as ATSC, DVBT, DVBC, etc. We set the VeCOAX to ASTC mode, which is used in the United States, and start playing the recorded commands on the laptop. Then the TV plays the command immediately. As a result, Amazon Echo was waked up successfully to operate on the command. Compared to TV signal injection over the air, coaxial signal injection attack didn't suffer that much from the signal loss over distance. Hence, such an attack could be launched much further away from the victim's home.

From the victim's perspective, it is easy to notice the original TV program is replaced if he or she is watching the TV with concentration. Then the victim would cancel all the operations that Echo will operate if possible. Therefore, the TV signal injection attack makes more sense when the victim is out of the home with TV turned off. Similar to the radio attack, we assume the TV is powered through Wemo Switch with exploitable vulnerabilities as in [4]. Hence, we can start broadcasting our recorded program on one dedicated channel, and then leverage the Wemo Switch to turn on the TV automatically. If the TV happens to be on the channel we are broadcasting with a proper volume, our voice command will be recognized and operated by Echo. Otherwise, we can repeat the above procedures by switching to another channel until success.

3.3. Wireless Speakers

Besides radio and TV, there are many other devices in one's home that can be controlled to generate voice commands. These include, but not limited to, speaker, smart camera, smartphone, computer, etc. Below we will discuss how attackers can also take over these devices and wake up Echo by voice commands.

Modern speakers are usually equipped with the capability of wireless connectivity, e.g., Bluetooth, NFC, Wifi. Once remotely connected, such speakers can play the audio over the air without wiring to the audio device like smartphone or Ipod. Unlike Radio or TV, wireless technology like Wifi, Bluetooth or NFC works effectively within a limited transmission range. For instance, Wifi typically works great around 10 meters, Bluetooth within several meters, and NFC around 10 centimeters. It is clear that the speakers only supporting NFC are not ideal for remote hacking, since it requires

the attackers be in the home close enough to the speaker. Regarding speakers supporting Bluetooth or Wifi, once attackers could retain a short distance with them outside the home, the speakers will be visible to attackers' audio devices (either Bluetooth or Wifi capable).

Table 2. Bluetooth speakers.

Brand	Model	Remote
Anker	SoundCore mini	✓
Willnorn	SoundPlus	✓
Samsung	HW-MS650	✓
Klipsch	ProMedia 2.1	✓
Insignia	NS-PSB4721	✓
VIZIO	S2121w-D0	✓
LG	SH7B/SJ7/SJ8	✓
LG	SJ9	✗

For most wireless speakers, there are two ways to pair them with audio devices. One is direct paring and the other is to input the passcode printed on the product package or manual. For the first situation, pairing is straightforward for attackers without any other additional actions. For the second situation, the attackers have to crack the passcode for victim's speaker. Usually, the passcode for wireless speaker is 4-digit number which is vulnerable to brute force attack. To demonstrate the feasibility of the above scenario, we conducted our experiment using several speakers of different brands which support Bluetooth connectivity. In Tabel 2, we summarized the remote pairing results of the tested Bluetooth speakers. In particular, we own Anker SoundCore mini, Willnorn SoundPlus and VIZIO S2121w-D0 speakers, so we use our iPhone 7 plus to pair them and play the recorded command to trigger Echo successfully. The distance between the phone and the paired speaker can be up to 4 meters across the wall and some furniture, which should be enough for the attackers to be outside the victim's home. Then, we went to Bestbuy display area for speakers, and turn on Bluetooth connectivity on our iPhone 7 plus. We found pairing most speakers is straightforward without requiring any passcode input. However, we did have difficulty pairing LG SJ9 speaker, with the error message of "Parking unsuccessful, make sure LG SJ9 (2D:2D) is turned on, and is ready to pair". We check manuals on LG official website, and find that some LG sound bars require function button be pressed to enter pairing mode. Hence, the LG SJ9 in display probably is not in pairing mode when we tried to connect.

In addition, nowadays camera manufaturers have integrated many additional user-friendly features to their products, such as sharing pictures to social networks, backing up pictures to cloud platforms, printing pictures remotely, connecting to smartphone, etc. However, such rich features also enlarge the attack surface against the cameras themselves. For example, recently people have seen the compromise of Samsung NX300 smart camera with root privilege [1]. Once the attackers successfully hack the smart camera, they can direct the camera to download video files containing their voice commands and let the camera play it loudly to control Echo to operate on the commands.

4. Evaluation

If we set the output power of HackRF One with full power, and the gain of ANT500 antenna at the max gain, the radio can clearly receive the injected fake signal up to 20 meters away from HackRF one. The commands played by the radio can control Echo up to 7 meters away. The demo in Section 3.2 shows that we can use the HackRF One to control the TV outside about 4 meters. In fact, our TV antenna and HackRF One could be as far as 15 meters across one wall, and the Echo can be as far as 8 meters away from the LG TV. The recorded commands can be played clearly when we set the volume of the TV to be 20%. Furthermore, using a high-end amplifier could extend the effective broadcast range a lot. Therefore, we can inject any commands in the signal and direct Echo to manipulate the virtual and physical world. According to the influence relates to human life, property, privacy, experience and so on, we assess the impact level as fatal, moderate and low, respectively.

4.1. Attack Evaluation in the Virtual World

Alexa knows what music you listen to, your shopping list, and the connected smart-home products information. Both the always-listening aspect and the data-collection tendencies raise privacy concerns. Although various certificate authentication, protection protocol and encryption algorithm could protect the data, an adversary is still able to apply IFTTT Applets and Alexa skills to retrieve sensitive information like victim's to do list, cell number, bank account and Facebook account. In addition, the attackers can send fraud, extort SMS/email. They can also spread rumors, reactionary statement and advertisement by the victim's social ID. Many users prefer to bind Twitter ID with Echo, to automatically or rapidly send short Twitter messages in a fixed format (e.g., Tweet the song the user is listening to on Echo through IFTTT Applets, synchronize the Shopping list and To-do list to the user's Gmail and Twitter). We analyze the threats as follows and list the representative attacks in Table 3. We use web crawler to obtain more than 50,000 Alexa skills and 600 IFTTT Applets related with Alexa, and conduct a key word searching method to find at least 100 of them can be taken advantage to attack. For example, if we want to look for related skills or applets which can

Table 3. Impact of attacking Echo in the virtual world.

Skill	Attack	Impact level
Mastermind	information leakage, fraud	fatal
To-do/Shopping list	information leakage, fraud	fatal
Facebook, Twitter, Sina Weibo	information leakage, fraud	fatal
Credit card	information leakage	fatal
Gmail	fraud	Moderate
Call or send messages	fraud	moderate
AmazonCloudDrive, DropBox, GoogleDrive, OneDrive	fraud	moderate
Google Photo, iOS Photo	fraud	moderate
OneNote, EverNote	fraud	moderate
Linkedin	fraud	low
Skype	fraud	low
Github	fraud	low
Online shopping	manipulate cart	low

be used as bank information stealing, we would search for key words like bank or credit card. We explore how to trigger them to conduct the attacks as below, and determine the impact level according to the actual effect.

Leakage of Sensitive Information . In order to serve the user conveniently, Amazon Echo may ask for access to user's private or sensitive information, such as her Twitter ID, Shopping list, bank account, etc. When an adversary could operate the malicious voice commands, such information could be exposed.

- By applying the skill "Mastermind", the owner could send and read SMS messages, make and answer phone calls, get the caller's name, get notifications and launch apps on his mobile device. To launch the information stealing attack, the adversary could ask Echo to call his phone number, the victim's phone number will be shown as a caller ID. This way the victim's phone number could be retrieved by the attacker.

- Usually, the Facebook ID, Twitter ID and etc, as private information, cannot be retrieved directly. However, IFTTT Applets are able to post a Twitter with the user's account if Echo updates a new song. Therefore, the victim's Twitter can be posted by changing the playing music. The attacker first lets Echo post a Twitter through the command "Alexa, play the song My Heart Will Go On". Then he looks for the most recently published Twitter messages from the homepage of Twitter. By looking for the same Twitter ID which sends the two messages at the specific time (controlled by the attacker), he is able to retrieve the ID. In other words, the Twitter ID can be obtained by the attacker.

- Compared with users' information in social networks, what is even more sensitive and more dangerous is the users' business and finance information. There are some skills handling such information, e.g., checking credit card bills. Once these skills are enabled, the attacker can directly lets Echo answer the sensitive questions. Taking Capital One as an example, users can inquire about the information about credit card, savings, home loan and so on. Example questions include "What are my recent transactions?", "What's my car loan principal balance?". We find that one option of enabling the skills is to equip an extra authentication such as a personal key to secure the sensitive information. Unfortunately, very few users really activate this authentication.

Spread of Malicious information. A user could easily send her own messages an email address or cell phone and post some contents to her social account like Twitter or Facebook, simply by talking to Echo with corresponding commands. While such contents will be shown in victim's account not attacker's, it could help the attacker to spread malicious messages conveniently. Examples are as described below.

- The skill "Tweet it" and "Twitter bot" are both unofficial Twitter client built to support the simple use case of tweeting. Once it is added

to Alexa account, it can help the owner use Echo to tweet by saying "Alexa, tell Twitter Bot/Tweet it to tweet···" Therefore, the attacker can spread malicious messages on the victim's account. Also, by applying IFTTT applet which could link to Sina Weibo, one of the most popular social platforms in China, an adversary could also conduct a similar attack.

- Malicious links and photos are also the main spreading routes in the virtual world. IFTTT enables the messages to be sent to Gmail, Linkedin, Google Photo, iOS Photo, OneDrive, Github, etc. Therefore, if the victim has enabled these Applets, the attacker can spread malicious links and photos to these platforms.
- Additionally, with IFTTT applets, an attacker can even command Echo to download the file at a given URL and add it to popular cloud drive service such as Google Drive or Dropbox, which could further lead to the malicious software attack and cause a severe loss for the victim.
- Attackers can add some advertisement on the victim's list with Echo by saying "Alexa, add···on my To-do list", and then ask "Alexa, what is on my To-do list?", so that the information can be sent to his email or phone. Since the email and SMS are sent by Alexa, they cannot be filtered out as junk mail or message. More pernicious, attackers can spread even violent, pornographic, reactionary and other illegal information. In addition, if the victim has enabled the Applet that prints the list with his wireless printer, the malicious information can be printed as well.

Calling or Sending SMS. The "Call and Message" skill makes Echo a free phone to call the owner's friends and family. However, the called subscribers can be their names or phone numbers, and people sometimes save the contacts as their appellations. Thus, after the attacker getting through the contacts by Echo, he can broadcast a help, threaten, extort signal by radio or TV.

To add a person to the contact list, what Echo needs is only the friend's phone number, which enables the attacker to call or message the user's friend on behalf of the user. Making things worse is to attach malicious contents with the message (e.g., a link to download a malicious application). This is very confusing since the message is sent to the friend using the user's account. Once the malicious contents are read by the friend, his machine may be compromised. Sometimes, even a figure (in the content) can trigger a severe vulnerability [22], allowing the device to be fully controlled by the attacker.

More recently, Alexa added "Phone.com Audio Interface" skill into the list. By using this skill, an attacker can easily talk to Echo like "Alexa tell phone dot com to call 888-280-4331" to call one US phone number without any authentication. If the called number is not toll free, the victim may face financial problems later.

Online Purchasing. Through voice commands, a user can ask Echo to order selected Prime-eligible products from the prime catalog or from her order history. After the order is placed, Echo uses the default payment method and shipping address in 1-Click setting to finish the order. Such an order can even be a service such as Uber. People can make a payment, get Amex Offers, check the balance with a 4-digit PIN.

From an attacker's point of view, besides shopping using the user's credit cards, he can also directly steal money from the user. For example, the attacker can pretend to be a Uber driver, stop around the user's house, command Echo to order Uber and, in the end, receive the Uber order. By default, voice purchasing is activated once the user registers his Echo. Further configurations can be operated in the Alexa app, such as turning off voice purchasing or requiring a confirmation code before every order. For security, a user can set a 4-digit code in Alexa app which Echo will ask for when the user is placing an order from Amazon. Unfortunately, this option is not mandatory. So the attacker can place items in the name of the user if he does not set 4-digit security code.

As the maturity and pervasiveness of Amazon Echo device and Alexa platform, more and more third-party shopping services like Best Buy have launched nowadays. Hence, an attacker can also operate related commands towards victim's Echo and make an order from Best Buy, causing much inconvenience and even money loss for the victim.

4.2. Attack Evaluation in the Physical World

As discussed above, hacking Echo can make a great attack in the virtual world. Similarly, people not only enjoy the convenience of IoT devices, but also take the risks of their vulnerability at the same time. Despite the vulnerabilities listed in Table 1. Unquestionably, even some of them are safe, we still can use Echo to attack them. Amazon Echo, as a hub for controlling IoT devices at home, can naturally send commands to the connected IoT devices, including smart locks, switches, thermostats, doors of garages, security cameras, etc. Such commands, once manipulated by the attacker and successfully interpreted by Echo, can further be executed by the corresponding IoT devices, which could bring serious threats to home. Table 4 shows the main attack in the physical world.

The attacker, after crafting voices of commands to the Echo, is apparently able to do whatever a legitimate user can do at home. Unfortunately for the attacker,

Table 4. Impact of attacking Echo in the physical world.

Target	Attack	Impact level
Car	remotely control vehicle	fatal
Garage	unlock and lock	fatal
Camera	disarm for further actions	moderate
Android Devices	remotely manipulate Bluetooth/Wifi	moderate
Router	wifi on/off	moderate
Switch	power on/off	moderate
Thermostat	control temperature	moderate
Oven	roast, preheat	light
Shower	turn on/off	light
Washer/Dryer	turn on/off	light
Air Purifier	change the air quality settings	light
Light	control light brightness and colour	light

we found that Echo still needs extra authentication for sensitive operations such as requiring a passcode to unlock the door. Below we made several experiments to understand what the attacker can really gain. To our surprise, we found that the attacker can still control most of the IoT devices.

Control smart cars. Vehicles are becoming more and more "smart". For example, a voice responsive door lock system is provided to further automate the open and close operations of doors. In this way, users do not need to stand by the vehicle opening the door. Instead, before coming to the vehicle, he could voice-command the door to open and walk into the vehicle. Many vehicles such as Tesla, BMW and Automatic support this kind of voice operations. Further making the vehicles "smarter" is the connection with "smart hub" like Echo. More functionalities can be supported beyond simply opening and closing the doors within a short distance, such as remotely getting the vehicle's location, which greatly extends the distance that an attacker could reach.

Once Echo is controlled by manipulated voices, an attacker can locate the vehicle no matter where it is. In most cases, such sensitive and even dangerous operations need extra authentications such as supporting a PIN code to Echo. An attacker without knowing such code cannot operate on the vehicle. An example is Genesis, a vehicle model of Hyundai, which permits a user to remotely start/stop/lock/unlock the vehicle with the PIN code. However, the openness of the platform of smart vehicles and Echo allows third-party developers to build their own skills for operating on the vehicles through Echo. Without considering strict safety and security policies, the developers of these unofficial skills may let attackers easily control the vehicles with no supply of any extra authentication, further exposing the legitimate users to dangers. We found such an app called "My Tesla". Once a manipulated voice command "Alexa, tell my car to flash lights/honk the horn" is sent to Echo, the attacker can remotely control the flash lights and honk of the vehicle, start or stop the charging system, set the temperature inside, etc. Therefore, the attacker gets full control of the vehicle.

Control smart locks/thermostats. Smart locks are one of the most favorite IoT devices that attackers like to control. Sending the voice command "Alexa, ask August to unlock my door" can unlock the door of the home which allows the attacker to walk in. Usually, extra passcode is needed for authentication before opening the door. However, we did find that some smart devices controlling the locks have no such authentication (e.g., Nexx Garage and Garadget products controlling doors of a garage). Once the voice command is sent to Amazon Echo, the smart lock connecting to Echo will let the door open, which allows an attacker outside the door to enter freely.

Another IoT device related to door opening is smart thermostats, which are originally designed to control the temperature at home through the voice commands from Echo. For example, after receiving the voice command "Alexa, set the downstairs temperature to 72", the thermostat will set the home temperature to 72 Fahrenheit if the unit was chosen as Fahrenheit. Also supported is the increase of the temperature at home. The interesting thing is that the high temperature will let some smart windows open itself to lower the temperature, as reported by Jack Jia, etc. [30]. As a result, even if the attacker cannot directly control the lock, he could still enter the home by setting up a high temperature to let the window open.

Control smart camera. Besides the IoT devices related to locks, attackers also care about the security cameras at home. Many of them connect to the Internet, allowing the owner to check the statuses at home anytime and anywhere. As a result, to avoid being found, the attacker should let the smart cameras not be able to work. For example, the attacker could use Echo to control the smart Homeboy cameras by simply crafting a voice command "Alexa ask Homeboy to disarm". Then the camera will stop working.

Until now, there are not so many cameras working with Echo, even though Echo possesses skills to arm or disarm Homeboy camera, but there is nobody enables them yet. However, once the owner enables them, the attacker will have the ability to turn off the camera when the owner is not in home so that owner cannot monitor the house, or turn it on at night and monitor the privacy inside.

Control devices' communication. Alexa skills such as "Find My Phone" can help the user find their phone call somebody. Specifically, "Alexa, ask Find My Phone to add another number" can add and delete the contacts on the address book. In addition, IFTTT can enable Echo to control the communication models of the phone. For example, The Applet "Tell Alexa to turn on your phone's wifi". If the attacker sets up a malicious WiFi hotspot with the same name and password as the victim linked before. The phone can automatically be linked to the malicious WiFi. It can turn on Bluetooth as well. As Table 2 shows that some Bluetooth devices need not require any passcode to pair. Therefore, attacker can monitor and manipulate the data.

Control smart router/switch/oven/light etc. The capability of commands manipulation to Echo can further be extended by other "smart hubs" such as Samsung SmartThings and other third party IoT platforms. These smart hubs, similar to Echo, connect hundreds of smart sensors, lights, locks, cameras, and even more to monitor and control home. In this way, if a smart device at home is not directly operated by Echo, it can be controlled by one of the smart hubs which connect to Echo. In other words, the voice command manipulated by the attacker can finally control the smart device through Echo.

- Echo can control the ASUS Router to pause the Internet, so that the IoT devices are offline, if the victim uses his camera to monitor his house, the video would stay at the last frame, which the victim may not realize that his house has been attacked.

- Another brute way for the attacker is to control the smart switches, again through Echo. Once an "Alexa, turn off my switch" command is sent to Echo, it will let the smart switch shut down by itself, and further all the devices connected to the switch will lose power to run.

- Echo works with Douch oven, Barbecue master. So an attacker can ask oven to roast, heat or stove, which can lead to fire if there is no person in the house.

- Even though people think the smart light is unconsidered for the thread, people can control the light brightness and color to transmit special signals [24].

5. Defense

Usually, researchers use signal processing and machine learning to defense the replay attack [13, 14, 16, 57]. In addition, voice print authentication is believed as an effective method. However, our test results on Echo, Google Assistant and Apple Siri are not very effective, as somebody or recorded audio can control them, which indicates that the root cause that enables the MUTAE attack is lack or weak of 1) user authentication, 2) user awareness and 3) fine-grained authorization for different (security sensitive) services. Therefore, we propose several defense solutions from three aspects.

First, the lack of user authentication in current voice control devices, like Amazon Echo, opens the initial loophole for the MUTAE attack. Therefore, it is critical to provide authentication. One strong and nature approach is to authenticate a user based on his/her *voice pattern*. That is, only a voice control command from an authenticated user can be executed. This kind of check needs to build a model to characterize users' voice. However, this approach also has a problem. It cannot prevent replay attack. The adversary can record the voice of a registered user and replay his/her command accordingly. To prevent the replay attack, we proposed a defense mechanism on the base of voice pattern authentication. We name it as two-factor authentication over the voice channel. That is, besides using the *voice pattern* for authentication, the Amazon Echo will act like a chatbot and ask questions on the fly. The questions can be based on user historical profile that was registered previously. The questions can also be simple questions, like "who is the current president of U.S.?", to test the intelligence and presence of a real user in front of the voice control device. The user must answer the question directly through the device. This type of two-factor authentication must be performed whenever a security critical voice command is received by Echo. In this way, not only the *voice pattern* of the user is matched, the presence and the human user identity will be checked. Hence, it prevents the MUTAE attack and other potential replayed MUTAE attack. Besides above method, Echo could also enable user's location check inside it. If the user's cell phone is not

in house wifi range, Echo will consider the user is out of safe range and will not response any further voice commands.

Second, for most of our proposed attacks, an victim's unawareness is necessary for the adversary, otherwise the victim could stop any dangerous and malicious actions caused by Echo immediately. So it is obvious that a natural defense method is to set user alerts for potential malicious actions. For example, if an adversary is commanding Echo to conduct potential dangerous actions like making a payment, the victim will receive SMS or email alert showing that actions and decide if he/she wants to continue. This way the attacker's action will be revealed to user and the user could stop it immediately.

Finally, the lack of fine-grained authorization for different users and under different contexts, also enlarges the attack surface of various attacks, which has been discussed in Section 3.2. For instance, the system can enforce the fine-grained policy that only authorized users (e.g., parents but not children) can purchase expensive items or media contents from Internet. Also, the system can enforce the fine-grained policy that only certain authenticated users are able to voice control the security critical operations (e.g., open the front door or windows). Furthermore, IFTTT enriches the skills of IoT and social activities, thus Echo can control IoT or online service based on registered IFTTT skills. Therefore, some fine-grained policy enforcement should be deployed with the application context. When MUTAE attacks trigger a set of Applets or leverage an Applet with low security sensitivity to trigger one with high security sensitivity, the fine-grained security policy should prevent these type of privilege escalation. For instance, if the user enables several *applets*: *applet 1*–"If motion is detected in my Homeboy location, turn my Philips Hue bulbs red"; *applet 2*–"If You say 'Alexa trigger switch off', then turn off Wemo switch." Considering that the Philips Hue bulbs are connected with the Wemo switch; *applet 3*–"Switch on Wemo if my Homeboy detects motion." Attackers can use Echo to turn off the switch, and then take actions, even if the Homeboy camera detects something. The bulbs will not turn red, and the owner will not discover it. The development of IFTTT Applets and how to use them should be scrutinized carefully. We also suggest that before a new skill or Applet is enabled, Alexa and IFTTT platform should provide a security vetting automatically based on the usage context.

6. Discussion

Comparison with other attacks. During last few years, different types of voice spoofing attacks have emerged towards intelligent voice controlled systems and devices. We hereby showed a overview plus comparison among MUTAE attack and other similar voice attacks. We define three metrics including effective distance, target systems and practicality. For attack distance, we consider 10 meters is the bar for long attack range which is enough for an attacker to be outside the room safely, with distance between 1 to 10 meters as medium and distance smaller than 1 meters as short. Target systems refer to the target of the attack, and practicality refers which type of the attack. For a practical attack in the real world, an over-the-air attack would be expected. **As we can see in Table 5, to the best of our knowledge, MUTAE Attack is the first long-range and practical attack which could compromised Amazon Echo devices.**

Limitations of our attack. Although our attack can control Amazon Echo in a long distance expectedly and may lead potential physical damage and financial loss for victims, there are two main limitations of our attack. First, in order to launch our attack in more aspects, we tried our best to do a comprehensive analysis for existing Alexa skills and IFTTT Applets for evaluation of the attacking consequences, and the results strongly indicate that our attack is promising for potential harmful issues in both virtual world and physical world. However, to successfully finish the whole attack, an adversary must ensure the victim has already enabled corresponding skills/applets, which means the adversary can only target one certain group of users. Second, despite the fact that our attack can be conducted remotely, the effective range is still not long enough to cover large amount of target devices and cause severe impact. Currently, our FM and TV signal injection attack can only be effective to 20 meters, which would only allow us initiate our attack for 2 to 3 houses normally. This distance range is highly related with our SDR device power limit, so we would believe a more powerful equipment can make us attack range much larger.

Future work. In this work, we explored Alexa skills and IFTTT Applets and revealed many potential security concerns if an adversary could conduct MUTAE attack and control victim's Amazon Echo. With the rapid development of AI and smart home technologies, Internet of Things have been increasingly equipped in our home and we could ask Echo to control more devices in the future. However, such communication channels not only remain between Echo and devices, but also among those smart devices. For example, an oven may be automatically turned on to prepare the dinner if the kitchen light is on. By now, we have little knowledge how such channels in smart home ecosystem work and whether vulnerabilities exist that an attacker could exploit to control those machines. Therefore, a potential future direction would be to develop a comprehensive and effective security vetting

Table 5. Comparison among voice spoofing attacks.

Attack Name	Effective distance	Target Devices/Systems	Practicality
Dolphin Attack [57]	Medium	iOS/Android Devices, Laptops, Amazon Echo, etc	over-the-air
IEMI Attack [31]	Medium	iOS/Android Smartphones	over-the-air
Hidden Voice Attack [16]	Not Given	CMU Sphinx (white-box attack) Google Speech API (black-box attack)	over-the-air (white-box attack) wav-to-API (black-box attack)
Practical Hidden Voice Attacks [7]	Not Given	Bing Speech API, Google Speech API, IBM Speech API	over-the-air
Carlini Attack [17]	Not Given	Mozilla Deepspeech	wav-to-API
Commandersong [56]	Short	Kaldi, iFlytek	over-the-air
MUTAE Attack	Long	Amazon Echo, Google Home, etc	over-the-air

system which could automatically evaluate the control flow and security issues in one smart home environment controlled by voice console like Echo.

7. Related Work

There are many research related to our topic, which can be summarized as four main categories: (1) voice command injection, (2) audio adversarial samples, (3) voice authentication for voice-based Internet of Things (IoT), (4) smart devices interacting with IFTTT.

Voice commands injection. Many researchers have demonstrated that it is feasible to inject voice commands remotely without raising victim's awareness. Kasmi et al. [31] introduced a new technique for remote silent voice command injection in smart phones based on smart IMEI. Diao et al. [23] and Jang et al. [29] proposed that malicious apps could play voice commands to control victim's cell phones. Zhang et al. [57] realized the inaudible attack on voice control systems by the carrier of ultrasonic. The inaudible attack could be interpreted as commands by voice-based devices. Our work differs with them: The previous works mainly targeted voice control system in a short range, but our attack can be performed in the long distance. Most similar to our work is that R.Martin [2] found Amazon Echo could be influenced by public radio stations while in our attack, we build a radio stations and extend the voice-generate equipment to more kinds of devices including TV, radio, speakers, etc.

We note that a shorter conference version of this paper appeared in [55]. In this manuscript, we proposed a new physical-world attack to inject coaxial signal towards TV. We further did a detailed analysis of Echo's voice control channel and the corresponding impacts if being compromised, in both physical and virtual world (e.g., social network). We also mentioned several feasible defense solutions to mitigate our attack, then users could further trust Echo to command other smart home devices or online services.

Audio adversarial samples. With the significant improvement of state-of-the-art deep learning [26] technologies, more current speech recognition systems are adopting neural network which could bring more accuracy. However, such deep learning technologies show vulnerabilities to adversarial sample [47], which is usually normal object added with small and unnoticeable perturbation but could be misclassified by machine as other target. Recently, researchers have proved such adversarial examples also exist in speech recognition systems. Vaidya et al. [52] and Carlini et al. [16] observed that attackers could issue hidden voice commands which were unrecognizable to human listeners but can be interpreted as desired commands by CMU Sphinx speech system, also in their black-box attack, the voice commands can be understood by Google Speech API. Similarly, Hadi et al. [7] use four methods to generate the noisy audios to practically attack several speech recognition models. Yuan et al. [56] successfully embedded voice commands into regular songs stealthily, which can compromise Kaldi, one popular open-sourced speech recognition system. They also showed that such samples could be played over-the-air and even transfered to another commercial black-box speech model. In addition, [44] use psychoacoustic hiding method to inject command into audios and attack Kaldi without human realization. Our work differ with them as our attack could compromise Amazon Echo and can be launched in a long range.

Voice authentication. Signal processing and machine learning can be used to defense the replay attack [13, 14, 16, 57]. In addition, many previous works demonstrate that the training data for victim's voice sample can

be collected and a voice biometric can be built for speech recognition [10, 15, 20]. However, no theoretical guarantee is provided to ensure the security of these models and replay attacks could compromise some cases. Huan et al. [25] proposed the body-surface vibrations of the user gathered by wearable devices, which can be further analyzed to determine if it matches the speech signal received by a voice assistant. This implementation would enhance the security concerns if the victim is in the noise-around situation or is conducting some confidential work. In our attack scenario, the victim is more likely to be far away from their house and Echo devices. So the wearable equipment would be unsuitable due to the transmit distance limitation.

Smart devices interacting with IFTTT. A considerable number of researches have been conducted to show that connecting a wide range of functionalities of IoT devices in smart home to each other and to different online services using trigger-action programming is feasible for ordinary users [21, 48, 51]. Surbatovich et al. [46] have proposed that IFTTT Applets can lead to privacy risks and potential harm in case that the attacker is able to exploit some trigger channels. In our attacks, this can also be achieved, considering the attacker can control Echo and use it to further trigger smart devices, which would then activate some Alexa-related IFTTT Applets.

8. Conclusion

Echo is one of the first always ready, voice controlled intelligent home appliances that connect to the social and IoT services. Based on Amazon's cloud-based voice service, Amazon provides a collection of APIs and tools such as ASK (Alexa Skills Kit), which allows third-party developers to build new functions into the Amazon Echo. Designers, developers, and brands can build engaging skills and reach millions of customers with ASK. So that Alexa can hear, comprehend, and resolve questions or commands. Besides, IFTTT Applets enrich the skills of Alexa tremendously. However, as people trust and enjoy the convenient voice control of Alexa skills via Echo, Echo dot and etc., unpredictable potential risks may be taken advantaged by injecting voice control commands to take over Echo, so that the attacker can process social network and control IoT devices stealing the owner's sensitive information, threatening his property even lives safety.

We reveal and implement the MUTAE attacks based on HackRF One, which can to inject voice commands to control Echo remotely. Moreover, We have further analyzed the impact of MUTAE attacks for IoT and social network services according to kinds of important skills. We propose to add voice pattern and answering questions as a two-factor authentication, to prevent the MUTAE attack and other potential replay MUTAE attack. Besides, we also suggest that Alexa and IFTTT platform provide a security vetting automatically based on the usage context.

Acknowledgments

IIE authors are supported in part by National Key R&D Program of China (No.2016QY04W0805), NSFC U1836211, National Top-notch Youth Talents Program of China, Youth Innovation Promotion Association CAS, Beijing Nova Program, Beijing Natural Science Foundation (No.JQ18011), National Frontier Science and Technology Innovation Project (No. YJKYYQ20170070).

References

[1] *Hacking the Samsung NX300 'Smart' Camera*. https://op-co.de/blog/posts/hacking_the_nx300/.

[2] *Listen Up: Your AI Assistant Goes Crazy For NPR Too*. http://www.npr.org/2016/03/06/469383361/listen-up-yourai-assistant-goes-crazy-for-npr-too/.

[3] *Morgan Stanley says Amazon has sold more than 11 million Echo devices*. http://www.seattletimes.com/business/amazon/amazon-has-sold-more-than-11-million-echo-devices-morgan-stanley-says/.

[4] *Belkin Wemo Home Automation devices contain multiple vulnerabilities*, 2017. http://www.kb.cert.org/vuls/id/656302.

[5] *CVE-2017-9765*, 2017. https://cve.mitre.org/cgi-bin/cvename.cgi?name=2017-9765.

[6] *Vulnerability Details : CVE-2017-9212*, 2017. https://www.cvedetails.com/cve/CVE-2017-9212/.

[7] Hadi Abdullah, Washington Garcia, Christian Peeters, Patrick Traynor, Kevin RB Butler, and Joseph Wilson. Practical hidden voice attacks against speech and speaker recognition systems. *Network and Distributed Systems Security (NDSS) Symposium*, 2019.

[8] VS Abhayawardhana, IJ Wassell, D Crosby, MP Sellars, and MG Brown. Comparison of empirical propagation path loss models for fixed wireless access systems. In *2005 IEEE 61st Vehicular Technology Conference*, volume 1, pages 73–77. IEEE, 2005.

[9] Tripwire Guest Authors. *My SecTor Story: Root Shell on the Belkin Wemo Switch*, 2015. https://www.tripwire.com/state-of-security/featured/my-sector-story-root-shell-on-the-belkin-wemo-switch/.

[10] Mossab Baloul, Estelle Cherrier, and Christophe Rosenberger. Challenge-based speaker recognition for mobile authentication. In *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG-Proceedings of the International Conference of the*, pages 1–7. IEEE, 2012.

[11] Ian Barker. *HomeHack vulnerability could allow your LG robot vacuum to spy on you*, 2017. https://betanews.com/2017/10/26/lg-hom-bot-homehack-vulnerability/.

[12] Mark Barnes. *A new hack can turn an Echo into a live microphone*, 2017. https://www.theverge.com/2017/8/1/16079044/amazon-echo-hack-microphone-listen-in-mark-barnes.

[13] Logan Blue, Hadi Abdullah, Luis Vargas, and Patrick Traynor. 2ma: Verifying voice commands via two microphone authentication. In *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*, pages 89–100. ACM, 2018.

[14] Logan Blue, Luis Vargas, and Patrick Traynor. Hello, is it me you're looking for?: Differentiating between human and electronic speakers for voice interface security. In *Proceedings of the 11th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, pages 123–133. ACM, 2018.

[15] Rudolf Maarten Bolle, Sharon Louise Nunes, Sharathchandra Pankanti, Nalini Kanta Ratha, Barton Allen Smith, and Thomas Guthrie Zimmerman. Method for biometric-based authentication in wireless communication for access control, November 16 2004. US Patent 6,819,219.

[16] Nicholas Carlini, Pratyush Mishra, Tavish Vaidya, Yuankai Zhang, Micah Sherr, Clay Shields, David Wagner, and Wenchao Zhou. Hidden voice commands. In *USENIX Security Symposium*, pages 513–530, 2016.

[17] Nicholas Carlini and David Wagner. Audio adversarial examples: Targeted attacks on speech-to-text. *arXiv preprint arXiv:1801.01944*, 2018.

[18] Catalin Cimpanu. *WikiLeaks Claims CIA Could Turn Samsung Smart TVs Into Listening Devices*, 2017. https://www.bleepingcomputer.com/news/hardware/wikileaks-claims-cia-could-turn-samsung-smart-tvs-into-listening-devices/.

[19] Lucian Constantin. *Researchers Find Vulnerability in Smart Home Control Apps*, 2017. https://motherboard.vice.com/en_us/article/pak3zg/wink-hub-insteon-hub-hacks.

[20] Amitava Das, Ohil K Manyam, Makarand Tapaswi, and Veeresh Taranalli. Multilingual spoken-password based user authentication in emerging economies using cellular phone networks. In *Spoken Language Technology Workshop, 2008. SLT 2008. IEEE*, pages 5–8. IEEE, 2008.

[21] Luigi De Russis and Fulvio Corno. Homerules: A tangible end-user programming interface for smart homes. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2109–2114. ACM, 2015.

[22] CVE Details. *Heap-based buffer overflow in IrfanView before 4.32*, 2012. http://www.cvedetails.com/cve/CVE-2011-5233/.

[23] Wenrui Diao, Xiangyu Liu, Zhe Zhou, and Kehuan Zhang. 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*, pages 63–74. ACM, 2014.

[24] Adi Shamir Eyal Ronen. Extended functionality attacks on iot devices: The case of smart lights. In *2016 IEEE European Symposium on Security and Privacy*, pages 1–12. 2016 EuroSP, 2016.

[25] Huan Feng, Kassem Fawaz, and Kang G Shin. Continuous authentication for voice assistants. *arXiv preprint arXiv:1701.04507*, 2017.

[26] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6):82–97, 2012.

[27] Dr. Shankar Banik Ike Clinton, Lance Cook. *A Survey of Various Methods for Analyzing the Amazon Echo*, 2016. https://vanderpot.com/Clinton_Cook_Paper.pdf.

[28] ISACA. *Alexa, Can You Hear Me? Demystifying the Amazon Echo Through Theoretical Bug Hunting*, 2016. http://www.isaca.org/knowledge-center/research/researchdeliverables/pages/alexa-can-you-hear-me.aspx.

[29] Yeongjin Jang, Chengyu Song, Simon P Chung, Tielei Wang, and Wenke Lee. A11y attacks: Exploiting accessibility in operating systems. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pages 103–115. ACM, 2014.

[30] Yunhan Jack Jia, Qi Alfred Chen, Shiqi Wang, Amir Rahmati, Earlence Fernandes, Z Morley Mao, Atul Prakash, and Shanghai JiaoTong Unviersity. Contexiot: Towards providing contextual integrity to appified iot platforms. In *Proceedings of The Network and Distributed System Security Symposium*, volume 2017, 2017.

[31] Chaouki Kasmi and Jose Lopes Esteves. "iemi threats for information security: Remote command injection on modern smartphones". *IEEE Transactions on Electromagnetic Compatibility*, 57(6):1752–1755, 2015.

[32] Swati Khandelwal. *Bluetooth Hack Affects 20 Million Amazon Echo and Google Home Devices*, 2017. https://thehackernews.com/2017/11/amazon-alexa-hacking-bluetooth.html.

[33] David Lodge. *Steal your Wi-Fi key from your doorbell?*, 2016. https://www.pentestpartners.com/security-blog/steal-your-wi-fi-key-from-your-doorbell-iot-wtf/.

[34] Guaranty Media. *95 percents OF US HOUSEHOLDS HAVE AT LEAST ONE RADIO RECEIVER*, 2017. http://guarantymedia.com/95-of-u-s-households-have-at-least-one-broadcast-radio-receiver/.

[35] Mitre. *CVE-2017-9765 Detail*, 2017. https://nvd.nist.gov/vuln/detail/CVE-2017-9765.

[36] Nicole Casal Moore. *Hacking into homes: 'Smart home' security flaws found in popular system*, 2016. http://ns.umich.edu/new/multimedia/videos/23748-hacking-into-homes-smart-home-security-flaws-found-in-popular-system.

[37] Mike Newton. *gSOAP remote code execution*, 2017. https://netvu.org.uk/is-the-gsoap-vulnerability-really-a-surprise/.

[38] Pierluigi Paganini. *Flaws in BMW ConnectedDrive Infotainment System allow remote hack*, 2016. http://securityaffairs.co/wordpress/49149/hacking/bmw-connecteddrive-hacking.html.

[39] Sheng-Lung Peng, Souvik Pal, and Lianfen Huang. *Principles of Internet of Things (IoT) Ecosystem: Insight Paradigm*. Springer, 2020.

[40] Zinaida Benenson Philipp Morgner, Stephan Mattejat. All your bulbs are belong to us: Investigating the current state of security in connected lighting systems. pages 1–13, 2016.

[41] Ian Poole. *Radio Signal Path Loss*. http://www.radio-electronics.com/info/propagation/path-loss/rf-signal-loss-tutorial.php.

[42] Dikla Barda Roman Zaikin and Oded Vanunu. *HomeHack: How Hackers Could Have Taken Control of LG's IoT Home Appliances*, 2017. https://blog.checkpoint.com/2017/10/26/homehack-how-hackers-could-have-taken-control-of-lgs-iot-home-appliances/.

[43] Rafael Scheel. *Smart TV Hacking (Oneconsult Talk at EBU Media Cyber Security Seminar)*, 2017. https://www.youtube.com/watch?v=b0J_8QHX60A.

[44] Lea Schönherr, Katharina Kohls, Steffen Zeiler, Thorsten Holz, and Dorothea Kolossa. Adversarial attacks against automatic speech recognition systems via psychoacoustic hiding. *Network and Distributed Systems Security (NDSS) Symposium*, 2019.

[45] Bob Sorokanich. *Researcher: BMW, Mercedes Vulnerable to Remote-Unlocking Hack*, 2015. https://blog.caranddriver.com/researcher-bmw-mercedes-vulnerable-to-remote-unlocking-hack/.

[46] Milijana Surbatovich, Jassim Aljuraidan, Lujo Bauer, Anupam Das, and Limin Jia. Some recipes can do more than spoil your appetite: Analyzing the security and privacy risks of ifttt recipes. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1501–1510. International World Wide Web Conferences Steering Committee, 2017.

[47] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*, 2013.

[48] Kazuki Tada, Shin Takahashi, and Buntarou Shizuki. Smart home cards: tangible programming with paper cards. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pages 381–384. ACM, 2016.

[49] Scott Tenaglia. *Breaking BHAD: Abusing Belkin Home Automation Devices*, 2016. https://www.blackhat.com/docs/eu-16/materials/eu-16-Tenaglia-Breaking-Bhad-Abusing-Belkin-Home-Automation-Devices.pdf.

[50] Iain Thomson. *Backdooring the Frontdoor Hacking a perfectly secure smart lock*, 2016. https://media.defcon.org/DEF%20CON%2024/DEF%20CON%2024%20presentations/DEFCON-24-Jmaxxz-Backdooring-the-Frontdoor-UPDATED.pdf.

[51] Blase Ur, Elyse McManus, Melwyn Pak Yong Ho, and Michael L Littman. Practical trigger-action programming in the smart home. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 803–812. ACM, 2014.

[52] Tavish Vaidya, Yuankai Zhang, Micah Sherr, and Clay Shields. Cocaine noodles: exploiting the gap between human and machine speech recognition. *Presented at WOOT*, 15:10–11, 2015.

[53] Mathy Vanhoef. *Key Reinstallation Attacks: Forcing Nonce Reuse in WPA2*, 2017. https://papers.mathyvanhoef.com/ccs2017.pdf.

[54] Vijay. *Hackers can spy on what you say by hacking Sony made Android TVs*, 2016. https://www.techworm.net/2016/05/hackers-can-spy-say-hacking-sony-made-android-tvs.html.

[55] Xuejing Yuan, Yuxuan Chen, Aohui Wang, Kai Chen, Shengzhi Zhang, Heqing Huang, and Ian M Molloy. All your alexa are belong to us: A remote voice control attack against echo. In *2018 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2018.

[56] Xuejing Yuan, Yuxuan Chen, Yue Zhao, Yunhui Long, Xiaokang Liu, Kai Chen, Shengzhi Zhang, Heqing Huang, Xiaofeng Wang, and Carl A Gunter. Commandersong: A systematic approach for practical adversarial voice recognition. *USENIX Security 2018*, 2018.

[57] Guoming Zhang, Chen Yan, Xiaoyu Ji, Tianchen Zhang, Taimin Zhang, and Wenyuan Xu. Dolphinattack: Inaudible voice commands. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 103–117. ACM, 2017.

[58] Steve Zurier. *Wemo IoT Vulnerability Lets Attackers Run Code On Android Phone*, 2016. https://www.darkreading.com/iot/wemo-iot-vulnerability-lets-attackers-run-code-on-android-phone-/d/d-id/1327362?