12-2022

A survey on security analysis of Amazon echo devices

Surendra Pathak
The University of Texas Rio Grande Valley

Sheikh Ariful Islam
The University of Texas Rio Grande Valley

Honglu Jiang
The University of Texas Rio Grande Valley

Lei Xu
The University of Texas Rio Grande Valley, lei.xu@utrgv.edu

Emmett Tomai
The University of Texas Rio Grande Valley, emmett.tomai@utrgv.edu

Recommended Citation

Pathak, Surendra, et al. "A survey on Security Analysis of Amazon Echo Devices." High-Confidence Computing (2022): 100087. https://doi.org/10.1016/j.hcc.2022.100087

High-Confidence Computing 2 (2022) 100087

Contents lists available at ScienceDirect

High-Confidence Computing

homepage: www.elsevier.com/locate/hcc

A survey on security analysis of Amazon echo devices

Surendra Pathak [a,*], Sheikh Ariful Islam [a], Honglu Jiang [b], Lei Xu [c], Emmett Tomai [a]

[a] Department of Computer Science, The University of Texas Rio Grande Valley, USA
[b] Department of Computer Science and Software Engineering, Miami University, USA
[c] Department of Computer Science, Kent State University, USA

ARTICLE INFO

Keywords:
Amazon echo
Privacy
Security
Intelligent virtual assistants

ABSTRACT

Since its launch in 2014, Amazon Echo family of devices has seen a considerable increase in adaptation in consumer homes and offices. With a market worth millions of dollars, Echo is used for diverse tasks such as accessing online information, making phone calls, purchasing items, and controlling the smart home. Echo offers user-friendly voice interaction to automate everyday tasks making it a massive success. Though many people view Amazon Echo as a helpful assistant at home or office, few know its underlying security and privacy implications. In this paper, we present the findings of our research on Amazon Echo's security and privacy concerns. The findings are divided into different categories by vulnerability or attacks. The proposed mitigation(s) to the vulnerabilities are also presented in the paper. We conclude that though numerous privacy concerns and security vulnerabilities associated with the device are mitigated, many vulnerabilities still need to be addressed.

1. Introduction

The widespread adoption of Amazon Echo family of devices has made Intelligent Virtual Assistant (IVA) ubiquitous in modern homes. More than 100 million devices have been sold by January 2019 that have Alexa on board [1]. Similarly, the global smart speaker market size is growing tremendously and can reach a worth of USD 15.6 billion by 2025 [2].

The device's popularity is partially attributed to its ability to carry out tasks using voice commands, which promotes human-computer interaction to a higher stage and abandons touch-based or other physical interactions-based interface. Though the new avenue of interaction has transcended device usability, it also introduces unforeseen security concerns. In 2017, a broadcast event triggered Amazon Echo in multiple households while covering an incident related to Amazon Echo [3]. Malicious skills that have similar names to genuine skills can be created to collect user information [4]. Additionally, inaudible voice commands can be used to exploit Alexa and carry out attacks [5]. In addition, since Intelligent Virtual Assistants (IVAs) are very intrusive to users' personal space, proper security and privacy concerns must be assessed. Thus, these systems become more prone to attacks without proper research and analysis on underlying security vulnerabilities and privacy concerns. Multiple articles in the literature summarize the security and privacy concerns of smart homes and smart speakers [6–10]. Despite the extensive publication on smart home and speakers, very few works deal with vulnerabilities specific to Amazon Echo [11,12]. Edu *et al.* [9] have comprehensively studied on security, privacy, and attacks on smart speakers. Our findings corroborate the results of the study and expand with more evidence. We have added additional vulnerabilities specific to Amazon Echo and presented/analyzed mitigation techniques. Studies specific to Amazon Echo [11,12] have left out details of vulnerabilities/attacks and mitigation techniques. We have added new vulnerabilities to this review literature. We have also systematically presented detailed software, hardware, system vulnerabilities, and adversary attacks to provide an inclusive review.

The major contributions of this work are:

- We outline the major components of Amazon Echo ecosystem.
- We studied and listed vulnerabilities of Amazon Echo classified into software, hardware, and system vulnerability.
- We discuss mitigation to those vulnerabilities.
- We present the results of a few adversary attacks on Amazon Echo.

We have presented the paper in the following structure: We outline different components of Amazon Echo ecosystem in Section 2. After that, we discuss the vulnerabilities of Amazon Echo categorized into software, hardware, and system vulnerability in Section 3. Simultaneously, we present mitigation to the vulnerability found in literature review. We also present a few vulnerabilities exploited by attackers in this section. Finally, we provide the conclusion of the paper along with future works in Section 4.

* Corresponding author.
E-mail addresses: pathak.surendra01@gmail.com (S. Pathak), sheikhariful.islam@utrgv.edu (S.A. Islam).

https://doi.org/10.1016/j.hcc.2022.100087
Received 22 August 2022; Received in revised form 12 October 2022; Accepted 12 October 2022

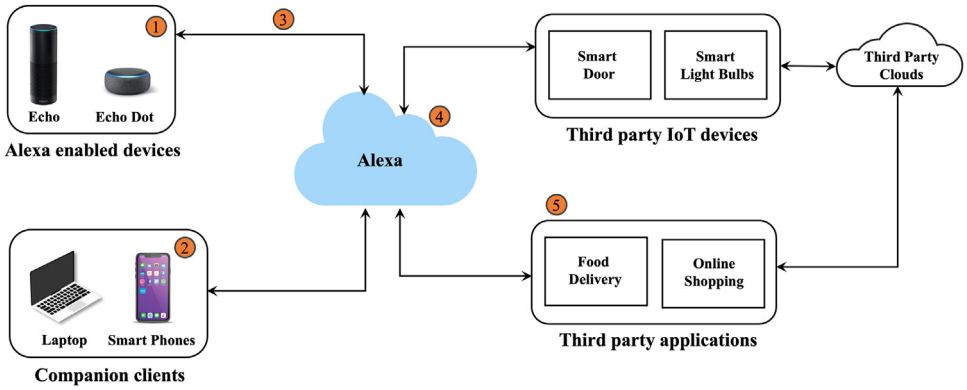

Fig. 1. Amazon Alexa Ecosystem and Corresponding Vulnerabilities [13]: ① Booting into Device Firmware, Dolphin Attack, Always listening Mechanism. ② Cross-Site Scripting Vulnerability. ③ Network Traffic Analysis Vulnerability, Device Pairing Protocol Vulnerability. ④ Cloud vulnerability, Automatic Speech Recognition Errors. ⑤ Skill Squatting Attack, Voice Masquerading Attack.

2. Amazon alexa ecosystem

Amazon Echo has become a very popular virtual assistant in the past few years. The services offered by the device have benefited many households and even businesses. Using the Amazon Echo, users can easily carry out multiple actions by just speaking to the device, a futuristic living experience that was thought of as fiction a few years ago. Though many people may not be aware of it, there is a solid architecture to carry out these functionalities. There are several entities playing roles in this ecosystem. Each of the entities shown in Fig. 1 is discussed in this section.

Alexa-enabled devices. Alexa-enabled devices are the Amazon Echo family of devices that user interacts with, usually by speaking out a command. The device consists primarily of a microphone and speaker and is connected to the Internet. A wake word is used to activate the device. After activation, it starts recording voice which is passed to the Alexa voice service, where computation is done. When the computation is complete, it receives a response that is played as sound.

Alexa cloud services. Most of the computation of Intelligent Voice Assistant is carried out in Alexa cloud services. The voice commands are sent to Amazon Echo, and the response is stored in Alexa cloud services. Alexa cloud service composes entities that carry out Automatic Speech Recognition, Speech-Language Understanding, Natural Language Understanding, Text-to-Speech conversion, etc.

Companion clients. Devices running one of the Alexa companion applications, such as Amazon Alexa, are companion clients. Apart from interacting with Alexa using voice commands, users can interact with them through a companion app. Though there is no specific companion application native to personal computers, users can still access Alexa using the web browser from a personal computer.

Third-party Internet of Things (IoT) devices. Compatible IoT devices increase the usability of Amazon Echo by adding additional voice-controlled functionalities. With a growing adaptation of Amazon Echo, the number of compatible IoT devices is also increasing. Some of the popular compatible IoT devices include Philips Hue, Lifx Mini, August WiFi Smart Lock, etc.

Third-party applications. The functionality of Amazon Echo is enhanced by many third-party applications that extend Alexa's capabilities. In addition, the "skills" extend Alexa's functionality, enriching user experience and enabling user-tailored services. Some examples are *Lyft* (ride-sharing), *Domino's* (food ordering), *The Wall Street Journal* (news updates), etc.

3. Vulnerabilities and mitigation technologies

This section reviews and summarizes the major vulnerabilities of Amazon Echo devices and corresponding mitigation techniques.

3.1. Software vulnerabilities

3.1.1. Skill squatting attack

Skills are the voice-driven capabilities developed for Alexa to power Amazon devices, such as Echo [14]. These skills that enrich Echo capabilities are developed using *Alexa Skills Kit*. More than 130,000 Alexa skills have been developed by 2021 [15]. A common skill usage scenario is illustrated in Fig. 2. Though skills extend Alexa functionality, they introduce a new attack vector.

Skill squatting attack exploits predictable errors, including homophones, compound words, and phonetic confusion, to wrongly direct users to malicious skills. Attackers create malicious skills with a similar invocation and intent name to legitimate skills [12]. A user intending to access a benign skill may be routed to malicious skill due to phonetic confusion. When a malicious skill gets access to the user device, further attacks can be carried out from there. Skill squatting attack is comparable to domain name typo-squatting in web applications where domain name's common typos are exploited.

Kumar et al. [4] carried out an experiment by developing multiple pairs of skills having similar invocation names. A total of 27 pairs of skills (target skill and squatted skill) are developed to examine whether Alexa triggers squatted skill instead of the requested target skill. Twenty-five of the total 27 pairs of skills are squatted at least once, giving a success rate of 92.6%. Skill squatting attacks can be extended to target a specific demographic which is termed spear skill squatting. "Squattable" words in the language of a demographic are exploited to target and attack the group in spear skill squatting.

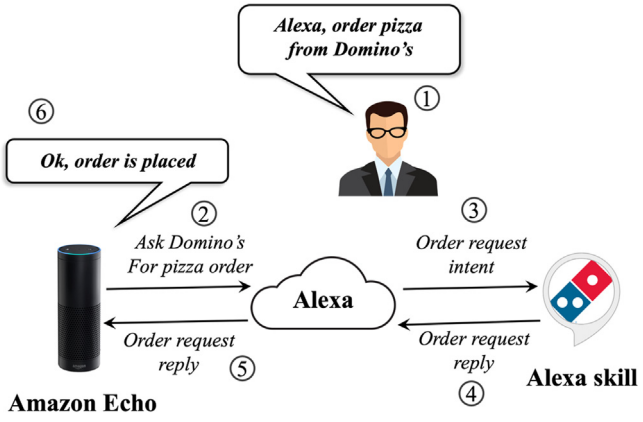

Fig. 2. A User-Alexa interaction to order a pizza [4].

Table 1
Summary of Amazon echo vulnerabilities and Attacks.

Exploitation mechanism	Vulnerability	Threat	Mitigation
Software	Skill Squatting Attack [4,12]	Malicious skill gets control of device	Screening of new skill's name using Word-based and phenom-based techniques
	Voice Masquerading Attack [17]	Malicious skill eavesdrops user's communication	Skill response checker and User intention classifier
	Network Traffic Analysis Vulnerability [18]	Adversary can detect user-device interaction time	-
	Device Pairing Protocol Vulnerability [19]	Associate a previously deregistered device to another amazon account	Initially associating new Amazon Echo with Amazon account used to purchase the devic
	BlueBorne attack vector [20]	Linux kernel and SDP server threats	Amazon published security patches
	Broadcast Media Vulnerability [3]	Echo triggered by broadcasting events	On-the-cloud system to detect media audios
	Automatic Speech Recognition Errors [21]	Alexa misunderstands words and triggers	Command discarded after looking on Amazon server
	Lack of Authorization Mechanism [22]	Any person can command Alexa	User-voice authentication mechanism
	Cross-Site Scripting Vulnerability [23]	Access, install and remove user's skills list	Findings shared with Amazon and issue fixed
	Cloud vulnerability [24,25]	Large amount of data stored in single place may be vulnerable	-
Hardware	Dolphin Attack [5]	Inject inaudible commands using ultrasonic channel	Utilizing non-linearity traces that can not be erased during signal modulation
	Booting into Device Firmware [26] [27]	Echo can be exploited by gaining root shell access	Issue fixed in later iterations of Amazon Echo
System	Always listening Mechanism [28]	Alexa records and streams conversation without utilizing wake word	Turing Echo mic off while not using the device
	Lack of Physical Presence Detection Mechanism [22]	Echo picks up commands from outside window/door	VSButton to check physical presence of user
Miscellaneous	REEVE Attack Vulnerability [29]	Use household audio device to remotely control Alexa	Two-factor authentication mechanism
	DoS Attack Vulnerability [30]	Makes Echo unavailable for use during attack period	-

Mitigation. Skill squatting vulnerability can be mitigated by adding a screening process during skill certification to scrutinize whether a skill can be confused with another registered skill. Currently, there are 30 skills with the name "Cat Facts", however the mechanism of how amazon routes a request is unknown. Such vulnerable skills can be mitigated by thorough scrutiny at the screening process such that each skill has a unique name. Though this mechanism may mitigate the vulnerability, skill publishers may have a conflict over skill names. They may want a simple skill name which may lead to name scarcity.

3.1.2. Voice masquerading attack

In Voice Masquerading Attack (VMA), users are unaware of skill eavesdropping on their conversations. As a result, an adversary can exploit the vulnerability to extract a user's private information. There are two major types of VMAs [17]:

- In-communication skill switch; and
- Faking termination.

In-communication skill switch is an opportunistic attack where a skill pretends to be another skill. The attack may occur when a user tries to switch skills during interaction with Alexa. A malicious skill pretends to hand over execution to the target skill by impersonating the target skill. As a result, the user may share the information intended for the target skill with malicious skill, which causes a serious privacy concern. Additionally, an adversary can exploit the acquired personal information to attack the user in the future.

Faking termination is a VMA where malicious skill fake skill termination to eavesdrop a user. Users may rely on skill's response to determine skill termination. For instance, users infer skill termination if the skill prompts "goodbye" or remains silent after execution. An instance of faking termination is shown in Fig. 3. A list of users' perceived indicator of end of conversation is summarized in Table 2. Malicious skills may create fake termination while keeping eavesdropping on sensitive information of Amazon Echo users.

Mitigation. The mitigation to VMA employs user's command and skill's response to establish an attack. There are two major components of the mitigation mechanism [17], *skill response checker* and *intention classifier*.

Skill Response Checker (SRC) examines skill's suspicious response to establish an attack. SRC examines whether a skill mimics Alexa response

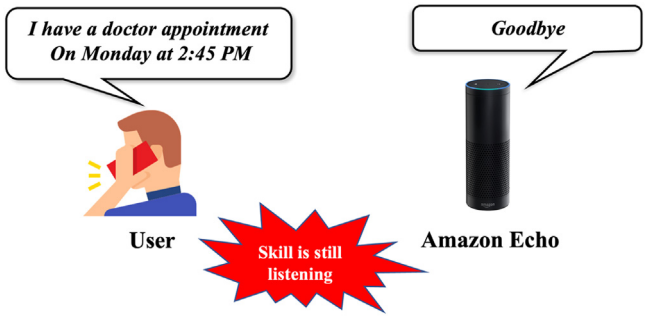

Fig. 3. A skill faking termination: User shares personal information thinking skill is terminated [16].

Table 2
Survey responses of Amazon Echo users [17].

Indicator of end of conversation	Users
Echo says "Goodbye" or something similar	23%
Echo does not talk	52%
The LED light on Echo is off	25%

by utilizing a set of common response patterns in Alexa. An alarm is triggered to notify Alexa of the event whenever a similar response is detected.

User Intention Classifier (UIC) examines a user's voice commands to ascertain if a user is wrongly attempting to switch to a new skill. UIC utilizes the semantics and context of a user command to check the user's intent. An approach based on contextual information is used here. For example, users use words semantically related to Alexa (such as "open sleep sounds") to switch contexts. Additionally, UIC compares user commands to system commands and current skills context to determine user intent. However, recognizing a user's intent is a challenging task. The pattern in which users interact with their devices can vary. Thus, finding a typical user pattern can be exacting. Moreover, determining the context of a command utilizes natural language processing. Even though the technology is helpful, it is ever-evolving, which introduces new challenges.

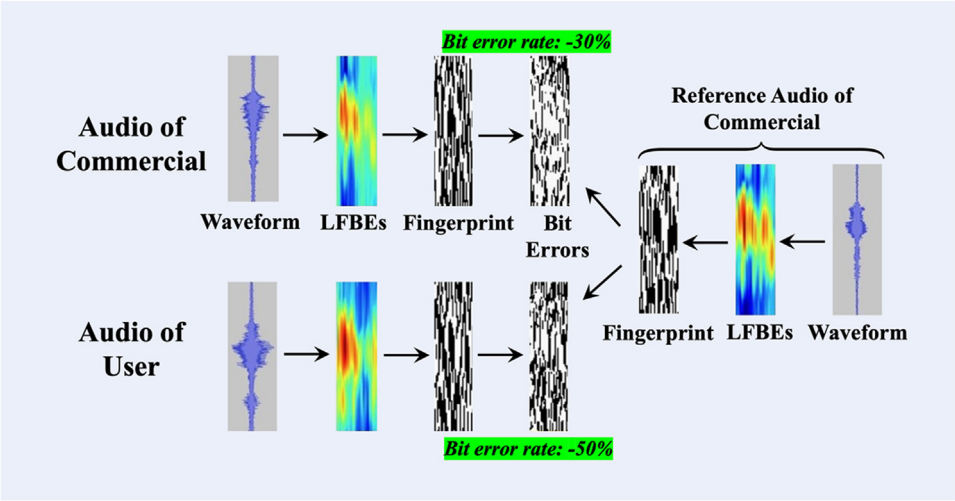

Fig. 4. Fingerprints being employed to detect audio match. Two audio with "bit error rate" less than 50% are likely to be from the same instance [34].

3.1.3. Network traffic analysis vulnerability

There are multiple works on network traffic analysis of Amazon Echo in the literature. Apthorpe *et al.* [18] carried out a study in that direction by setting up a laboratory smart home environment. The setup consists of Raspberry Pi 3 Model B to record IoT devices incoming and outgoing signal packets. The setup is utilized in a three-step strategy to identify IoT devices residing in a home and infer user behaviors. Firstly, network traffic is recorded and separated into streams. Secondly, an IoT device most likely to be associated with the stream is identified. DNS queries of each stream match the device. Thirdly, traffic rates of the stream are monitored to reveal user behaviors.

In the experiment, Amazon Echo was asked a series of questions to observe the device's network traffic. The authors were able to identify the instances of user-device interactions using network traffic data. The knowledge of the user-device interaction time to an adversary may have unwanted implications and privacy concerns.

IVA and IVA-enabled devices mostly communicate over a secure channel using encrypted HTTPS [31]. However, such encryption cannot protect specific communication patterns like payload sizes, data rates, and source/destination. Many state-of-art machine learning techniques can leverage such information to infer user behaviors such as duration of user-device interaction, listening to music, and ordering products or services. In addition to that, machine learning algorithms may be used to predict user commands [32,33].

3.1.4. Broadcast media vulnerability

In January 2017, a six-year-old girl from Dallas accidentally ordered a dollhouse while playing with Amazon Echo [3]. The device ordered a dollhouse when the girl asked Echo, "Can you play dollhouse with me and get me a dollhouse?". Later, Echo devices in multiple households were triggered when a morning show covered the event. The Amazon Echos listening to the news, tried to order a dollhouse.

Mitigation. In response to such events, Amazon developed an on-the-cloud system to distinguish media audios. The system uses broadcast audio to teach Alexa about recorded instances of Alexa's trigger words and use this knowledge to detect recorded sounds in the future. In addition, the system utilizes a technique called acoustic fingerprinting, an efficient mechanism that is robust to audio distortion and interference produced by television and other digital devices [34]. Usage of fingerprinting in such a process is shown in Fig. 4. However, some false positive results were observed when testing several videos for the fingerprint match [35]. In addition, some videos that do not contain a wake word resulted in a fingerprint match, which raises a question about the technique's robustness.

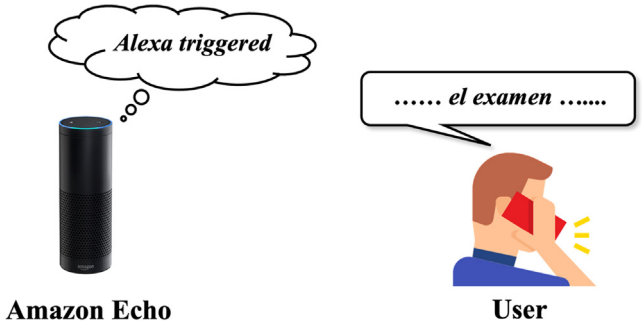

Fig. 5. Speech recognition error: El examen interpreted as Alexa triggers the device.

3.1.5. Automatic Speech Recognition (ASR) errors

Though Amazon tried hard to address Alexa's broadcast media vulnerability, Alexa is still vulnerable to various automatic speech recognition errors. Castell-Uroz et al. [21] experimented with an audio database with a few interesting findings.

A Spanish language audio database (approx. 1700 files) was employed to surveil Echo's reaction to distinct sounds. Database audio was reproduced nearby Echo, where some words (e.g., "el examen", "economia") from the database triggered Alexa due to speech recognition error. A situation where Alexa got triggered due to an ASR error is shown in Fig. 5. After getting triggered, Alexa looked up commands on Amazon server but was eventually discarded. The results designate that Amazon's security mechanism discards this kind of false positive.

Mitigation. Due to the limitations of speech recognition technology, ASR errors are unavoidable. However, attempts are made to minimize errors and improve performance. For example, Swarup et al. [36] diminished ASR errors by enhancing existing baseline model architecture with learned features. Similarly, Wang et al. [37] injected noise into error-free ASR-generated text data to train the dialog model with augmented data. The authors claimed to make VPA robust to ASR errors.

3.1.6. Lack of authorization mechanism

A user commands Echo by speaking out a trigger word. The trigger word is "Alexa" by default, however, it can be configured to be one of the "Amazon", "Computer", or "Echo" [38]. There is an absence of an additional authentication layer to control the access to the device, which is a serious vulnerability. Amazon Echo does not check if a command is issued by an authorized user or someone else, making it vulnerable to

attackers who manage to get access to the device. In addition to that, Amazon Echo can be triggered by machine-generated voices due to the lack of an authentication mechanism [22]. MP3 audio files generated via an online resource have successfully accessed the device and executed commands. MP3 audio from various devices such as Bluetooth speaker, laptop, desktop, and mobile phone, is capable of issuing commands to Alexa.

Mitigation. A layer of authentication can be implemented by adapting a biometrics-based authentication scheme in Amazon Echo. A camera module can be integrated to identify users and help enforce authentication schemes. Authorized users are verified by a face-recognition system when they gaze into the device [39]. A face-recognition algorithm wakes up the camera and authenticates users enabling a secure authorization mechanism. Once a user is authenticated, echo can listen and execute user commands securely. The biometric-based authentication can be implemented in future models of Amazon Echo. However, it is challenging to implement the authentication procedure in the current and previous models in the user households due to the hardware nature of mitigation.

3.1.7. Bluetooth associated vulnerability

IoT devices, including Amazon Echo, can be vulnerable to Bluetooth-associated vulnerability, which may compromise the device and user data. Additionally, Bluetooth-enabled devices are vulnerable to "BlueBorne" attack vector that endangers the integrity of digital devices [20]. BlueBorne attack vector has eight zero-day vulnerabilities critical to IoT device security. Specifically, there are two vulnerabilities of Amazon Echo:

- Linux kernel: Remote code execution vulnerability
- SDP server: Information leak vulnerability

BlueBorne permits attackers to compromise a device even when Bluetooth is not in discoverable mode. For Amazon Echo, there is an absence of a mechanism to turn Bluetooth off given the device's limited user interface, making it vulnerable to BlueBorne attack. Additionally, Echo devices constantly scan for Bluetooth communications increasing the risk of attack.

Mitigation. Armis Labs apprised Amazon regarding BlueBorne attack vector-associated risks. Amazon issued an update in response to security fixes. In addition, Amazon Echo users (version>v591448720) have been automatically updated with the security patch.

3.1.8. Cross-Site scripting vulnerability

Cross-Site Scripting is an injection attack where malicious scripts are injected into harmless websites. Alexa can be vulnerable to Cross-Site Scripting (XSS), according to a study in August 2020 [23]. A Cross-Origin Resource Sharing (CORS) token can be extracted using XSS that is exploited to perform actions using the victim's identity. The attack shown in Fig. 6 is carried out as follows:

- The user receives a malicious link with code-injection capability that redirects the user to amazon. User clicks on the malicious link.
- An AJAX request using the user's cookies is sent to access the list of user's installed skills on his/her Alexa account. The CSRF (Cross-Site Request Forgery) token is retrieved as a part of the response.
- CSRF token is misused to remove a skill from the user's list of installed skills.
- Attacker now installs a skill whose invocation phrase is identical to the deleted skill.
- Malicious skill is triggered when the user uses invocation phrase.

An adversary can exploit certain vulnerabilities in Alexa subdomains to carry out attacks targeting Alexa users. Adversary takes advantage of these vulnerabilities to carry out multiple actions in multiple stages to attack targeted users. The attack initiates when the user clicks on a malicious link. An attacker can carry out the following attacks [40]:

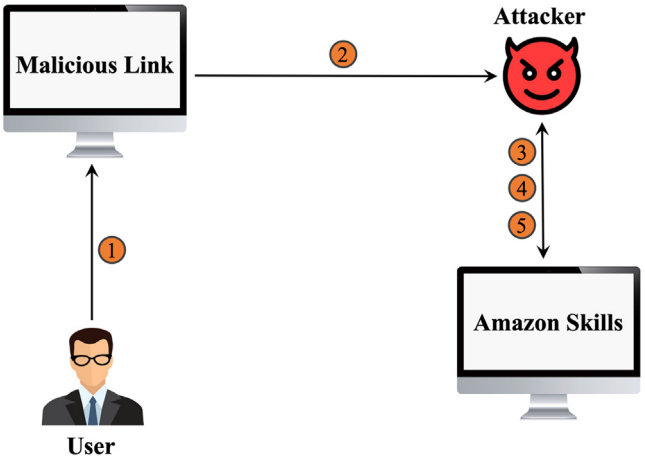

Fig. 6. Attack flow using XSS and CSRF token: ① User clicks on malicious link; ② Adversary gets user cookies; ③ Adversary sends AJAX request with user cookies; ④ Adversary gets skill list and CSRF token in response; and ⑤ Adversary adds/deletes skills using CSRF token and skill id.

- Access user's Alexa voice history.
- Install skills to user's Alexa without user's knowledge.
- View the list of User's Alexa skills.
- Remove a user's skill without user's knowledge.
- Access user's personal information that includes bank details, personal details, addresses, phone numbers, and etc.

Mitigation. The findings of the study illustrating the vulnerabilities were shared with Amazon. Amazon responded to it by fixing issues and pushing updates. No manual update is required from Echo users to mitigate the vulnerability.

3.1.9. Device pairing protocol vulnerability

Studies have found network vulnerability associated with pairing clients with Amazon Echo. Janak et al. [19] found an issue with the out-of-the-box experience (OOBE) protocol. Amazon Echo uses the OOBE protocol to provide a device's network credentials and establish the device with an Amazon account. One of the certificates, i.e., X.509, used in the pairing process is self-signed, which makes the method vulnerable to Man-in-the-Middle (MITM) attacks. Thus, an adversary detecting the pairing exchange can acquire a link code and compromise the system. The authors found that it is possible to link a previously de-registered device to another amazon account. The vulnerability can be mitigated by initially associating the new Amazon Echo with an Amazon account used during device purchase.

The authors found another device pairing protocol vulnerability associated with Amazon web application. There exists a vulnerability in web application pairing clients of Amazon Echo. The web application is initially rendered via HTTPs, which is secured during the device pairing process. However, after a user logs in, the application rendered uses HTTP to download the JavaScript pairing client. Since the mechanism uses HTTP, it leaves the application vulnerable to attackers. The attackers can exploit the vulnerability to carry out code injection attacks which can cause severe consequences.

3.1.10. Cloud vulnerabilities

Cloud is an integral part of VPAs like Alexa, and provides both scalable computing service and virtually unlimited storage. Amazon stores gigabytes of Alexa voice recording data every hour in its cloud storage. Though VPA's cloud reliance has multiple benefits, it also introduces a unique attack avenue for an adversary. The storage of a large amount of information at a single point poses security concerns. Alexa may store personal or sensitive information in the cloud. An attacker can exploit such information if the data storage is compromised.

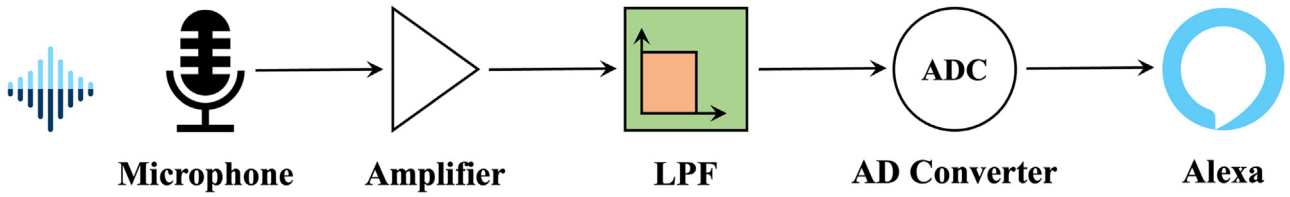

Fig. 7. Demonstration of modulated tone traversing the signal pathway of a audio device in terms of FFT [5].

Amazon utilizes user data to train Alexa's speech recognition and natural language understanding algorithm [24]. The data is fed into the algorithm to better understand user commands and take appropriate action. The algorithm learns from past data to make better responses in the future. With the availability of a large amount of data, the algorithm can train extensively, and Alexa can better interact with users. Despite the advantages obtained from the procedure, Amazon can infer information on user lifestyles such as daily routine, user preferences, etc. Moreover, the information can be exploited to harm user if someone with an evil intention gets access to the data.

Besides speech recognition and natural language understanding algorithm, Amazon also employs people to provide human feedback to train the algorithms [25]. The team is responsible for listening to users interacting with their devices and assisting in training software algorithms. The team listens to voice recordings captured at users' homes, transcribing and annotating them before sending feedback to the system. The process assists Amazon in training speech recognition and natural language understanding so that Alexa can better comprehend and respond to human requests. In its marketing policy, Amazon does not explicitly mention human listening being involved even though they say that users' requests are used to train their speech recognition and natural language understanding systems.

3.2. Hardware vulnerabilities

3.2.1. Dolphin attack

Dolphin attack is an inaudible attack that exploits the ultrasound channel and underlying hardware vulnerability to inject inaudible voice commands at VPAs. The attack uses modulated audio commands on ultrasound carriers (frequency >20 kHz), making the command inaudible to the human ear [5]. The modulated command is demodulated and interpreted at voice capture hardware and speech recognition system, respectively, at VPA. The modulated audio signal can be successfully demodulated by leveraging the non-linearity of microphone circuits. Modulated signal traversing an audio capture device is illustrated in Fig. 7. The attack exploits Micro Electro Mechanical Systems (MEMS) microphones that accept inaudible ultrasound signals as legitimate commands. Since the attack employs synthesized ultrasound signals, an attacker requires proximity to the target device. For example, Amazon Echo can pick up and execute inaudible audio commands from a distance of 165 cm. The attack range was further increased to 25ft by exploiting the non-linearity of the Echo's microphone [41].

Mitigation. Dolphin attacks can be abused to carry out unsolicited actions on Amazon Echo. Therefore, defense strategies should be employed to address the unwanted attacks. Hardware-based defense strategies such as microphone enhancement can be an approach in that direction. Since the current MEMS microphones can sense high frequency (>20 kHz) signals, they can be enhanced to suppress such signals. Similarly, there have been defense attempts utilizing the non-linearity traces, which cannot be erased during the signal modulation [41].

3.2.2. Booting into device firmware

Amazon Echo can be exploited physically, allowing an adversary to gain root shell access to the underlying Linux OS. Amazon Echo has two underlying vulnerabilities. [26]:

- Exposed debug pads at its base.
- Hardware configuration setting that permits booting device via an external Secure Digital (SD) card.

These vulnerabilities can be exploited and allow the attacker to boot into the underlying Linux environment from an SD card [27]. Furthermore, an attacker can boot into the device's firmware and install a persistent backdoor that allows remote root shell access to the device. After the root access is obtained, the attacker can install malware, steal authentication tokens, and wiretap the device remotely. Rooting Amazon Echo requires physical access to the device, which may not be a concern for a device in a secure location such as a personal household. However, adapting Amazon Echo to places such as hotel rooms provides an avenue for attacking [11].

3.3. System vulnerabilities

3.3.1. Always listening mechanism

Studies have shown that Amazon Echo starts recording and transmitting audio only after it gets triggered with a wake word [42]. Till then, it stays in a dormant state of buffering and re-recording until a wake word is detected. Ford and Palmer [28] carried out an experiment in that direction where they analyzed Echo Dots' network traffic over 21 days in a private household. Nobody in the household interacted with the devices on purpose utilizing a wake word during this period. Analyzing the logged audio reveals that 70% of logged response cards were Television sounds and 30% were human voices. This demonstrates that Amazon Echo records private conversations without utilizing a wake word. This can be a significant privacy concern where personal or sensitive audio is leaked accidentally or by an attacker.

Mitigation. The vulnerability can be mitigated by turning the device mic off with a physical mechanism while speaking out private information. Alexa does not stream audio to Amazon AVS cloud while the device mic is turned off [28]. Echo light turns red when microphone is turned off as shown in Fig. 8. However, many users

Fig. 8. User turn device mic off while revealing confidential information. Echo's LED light turns red while mic is off.

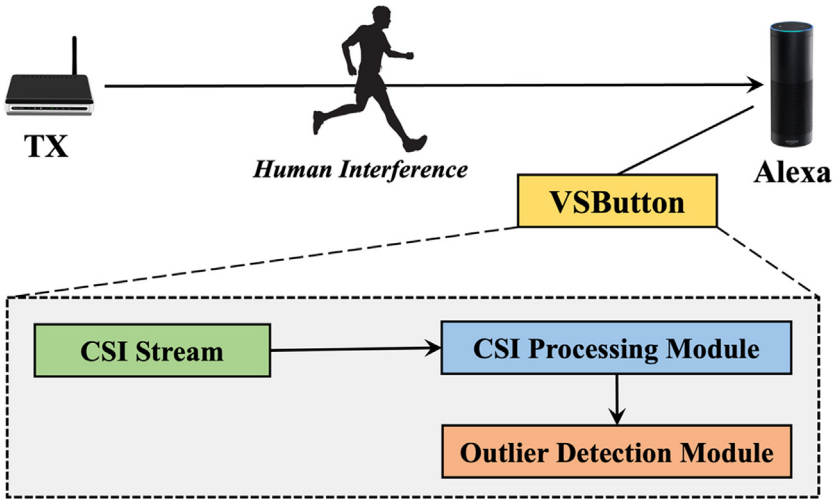

Fig. 9. Design of VSButton [22]. VSButton utilizes Channel State Information (CSI) to detect human motion. Noises in CSI values are eliminated in CSI Processing Module. CSI patterns of movement is detected in Outlier Detection Module.

do not use the mic button despite being aware of the functionality. Multiple users perceive the technique negates the device's hands-free accessibility [43].

3.3.2. Lack of physical presence detection mechanism

Amazon Echo does not require a user to be physically present near the device to request a service. Due to the absence of a mechanism to detect the physical presence, an Alexa-enabled device executes any command that it can hear, provided that the command is loud enough. Any service request that reaches Amazon Echo at 60dB (or higher) sound pressure level gets served by the device. It is a severe vulnerability that can be exploited in multiple ways. For instance, an adversary can issue a command from the facade to access Amazon Echo inside a household. The adversary can then aggravate the attack by utilizing other devices connected to the Echo. Alternatively, an adversary can control Echo if he gets access to one of the speakers in proximity to the Echo in the household. The attacker can abuse the speaker to play audio containing wake words and commands to compromise Alexa-enabled devices.

Mitigation. A user's physical presence can be detected by the Virtual Security Button (VSButton), which is an access control technique that utilizes the physical presence of a user. The secure access mechanism allows access to Alexa only when VSButton is in a push state. The virtual button is pushed whenever a human presence is detected nearby. The access control mechanism utilizes a home WiFi network to detect user movement. VSButton monitors the Channel State Information (CSI) of home WiFi to detect human motion. A user can push VSButton simply by waving his hand. The variation in CSI values within a room can be leveraged to detect human motion. Movements inside a room cause considerable variation in CSI values, while movement outside the room/house causes only a tiny variation. The phenomenon is employed to determine if movement is occurring inside the room.

The human movement detection by VSButton consists of two major steps:

- CSI processing phase;
- Outlier detection phase.

In CSI Processing Phase, noises in CSI values are eliminated. The output is then utilized in Outlier Detection Phase to detect CSI patterns of movements inside the room. A real-time hyper-ellipsoidal outlier detection mechanism is employed in the later phase to detect human movement. The components of VSButton are shown in Fig. 9.

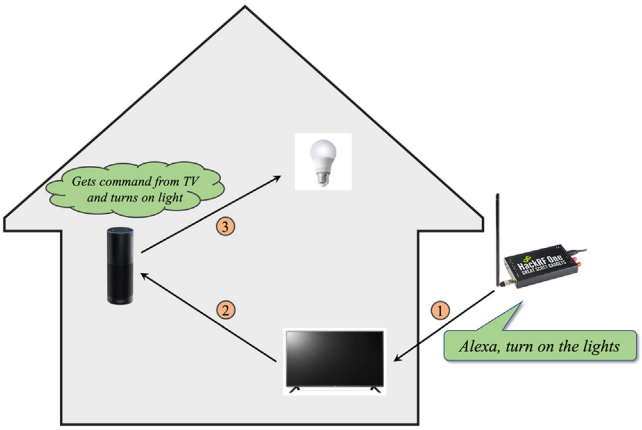

Fig. 10. REEVE attack flow: Television receives recorded commands from HackRF One and speaks out the commands to Alexa. Alexa triggers and executes the commands [29].

3.4. Vulnerabilities exploitation by attackers

3.4.1. Remote voice control (REEVE) attack

REEVE attack is a remote attack that exploits devices like radio, TV, speaker, etc., to attack Amazon Echo and carry out unauthorized actions. TV speaker, radio speaker, or Bluetooth speaker controlled by an attacker is exploited to attack Amazon Echo. An experiment investigated a scenario where a TV speaker is used to control the Amazon Echo by replacing the video content of the current TV channel with a video containing several Echo commands [29]. The brute force approach can determine the current TV channel for the attack. Channel 48, i.e., HBO, was used for the attack in the experiment. The Amazon Echo commands recorded originally in mp4 format are later converted into the "Transport Stream" format that HackRF uses for ATSC (Advanced Television Systems Committee) broadcast. Amazon Echo was placed at a distance of six meters from the Television in the same room, while HackRF One was kept at a distance of four meters from Amazon Echo outside the room. The setup is illustrated in Fig. 10. After that, GNU radio software was configured with a central frequency of 677 MHZ and a bandwidth of 6 MHz to broadcast fake programs to replace HBO content. HackRF One is used to transmit recorded programs to the TV. When the Television is turned on, it plays the fake recorded program containing several Alexa commands executed when Echo hears them.

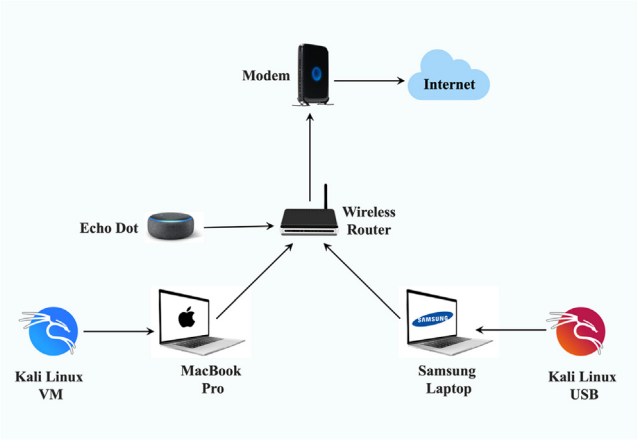

Fig. 11. Network diagram for penetration testing [30].

Mitigation. REEVE attack can be defended by addressing two root causes of the attack [29]:

- Lack of user authentication;
- Lack of additional security layer for sensitive services.

Due to the absence of user authentication for accessing Echo, Alexa can be triggered by unauthorized users and electronic devices. Two-factor authentication can be implemented to mitigate the issue by verifying the authorized user's voice pattern. Moreover, Alexa can ask users questions on the fly, where the user must answer questions to continue accessing the device. These approaches verify the physical presence and user identity while performing security-sensitive operations. In addition to that, Echo can enforce an authorization policy where only authorized users can perform security-critical activities such as opening the front door of the house.

3.4.2. Denial-of-Service attack

Overstreet et al. [30] carried out penetration testing to test the vulnerability of Amazon Echo against Denial of Service (DoS) attack. Kali Linux operating system was used for carrying out the attack and network monitoring. One instance of Kali Linux was set up in Samsung 900X and the other in Virtual Box hypervisor in a macOS system to carry out network monitoring and DoS attack, respectively. After that, an open and unsecured network was created using ASUS WL-500W wireless router. The firmware of the router was flashed and OpenWRT was installed. OpenWRT is a Linux-based OS for the network device. The network setup is illustrated in Fig. 11.

A network scan was carried out after the setup. Firstly, nmap scan was run to collect the MAC and IP addresses of devices connected to the network. Another nmap scan was run to find information about TCP ports on the devices via the SPARTA tool in Kali Linux. Then Samsung's network interface card was changed from managed to monitor mode to trace the Amazon Echo's activity on the test network. Airodump-ng command was run to dump network traffic to the specified directory in pcap format. The capture file was imported into Wireshark to analyze packets and infer the activities occurring over the network. After the network scan was complete, a DoS attack was carried out on Amazon Echo. Metasploit was used to launching a syn-flood attack to crash Amazon Echo. Data was captured before and during the attack to observe network traffic which was analyzed later using Wireshark. Before the attack, network traffic was regular, but packets were dropped on the network during the attack. Furthermore, the syn-flood attack crashed Amazon Echo, making the Echo's LED indicator (located at the top of the device) turn red instead of the standard blue color. The device's functionality returned to normal after the attack concluded.

There are significant consequences of the DoS attack on Amazon Echo. First, a user is not able to access information when required, which can be concerning during critical situations. An attacker can prolong the DoS attack to make Echo nonfunctional for a longer duration which can be more concerning. Finally, if an attacker has proper knowledge and tools, he can utilize freely available resources within Kali Linux to carry out the attack.

4. Conclusion

Alexa has changed the way users interact with personal assistance devices in a revolutionized way. With the popularity of the Amazon Echo family of devices, users can carry out tasks like making phone calls, online shopping, and controlling the smart home with voice commands. Due to the widespread use of Echo devices, it is imperative to ensure proper security and privacy measures to avoid an unwanted scenario in the future. Few vulnerabilities of Echo devices discovered during the literature review are discussed in this paper. The vulnerabilities are classified as software, hardware, or system vulnerability and specific adversary attacks. Mitigation of the vulnerabilities is also discussed simultaneously. We observed that though few vulnerabilities are mitigated, many others remain unresolved.

The primary theme of this research was to investigate and summarize the existing vulnerabilities of Amazon Echo. For future work, we plan to narrow down to a specific vulnerability for a detailed study. Furthermore, we intend to get insights into the network behaviors of Amazon Echo by analyzing the device's network traffic data. Finally, detailed research on network behavior using machine learning models may give an outlook on new vulnerabilities or confirm existing findings, which would be an exciting line of future work.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] D. Bohn, Amazon says 100 million Alexa devices have been sold - what's next?, Verge. https://www.theverge.com/2019/1/4/18168565/amazon-alexa-devices-how-many-sold-number-100-million-dave-limp.
[2] Smart speaker market with COVID-19 impact analysis by IVA, markets and markets. https://www.marketsandmarkets.com/Market-Reports/smart-speaker-market-44984088.html?gclid=Cj0KCQjw6uT4BRD5ARIsADwJQ1-vRVTShO8MMfltZr-VGKAYbcvlyLuF4igwuFIw829WC2lmyPrthEMoaAm-0EALw_wcB.
[3] A. Liptak, Amazon's Alexa started ordering people dollhouses after hearing its name on TV, Verge. https://www.theverge.com/2017/1/7/14200210/amazon-alexa-tech-news-anchor-order-dollhouse.
[4] D. Kumar, R. Paccagnella, P. Murley, E. Hennenfent, J. Mason, A. Bates, M. Bailey, Skill squatting attacks on Amazon Alexa, in: 27th USENIX Security Symposium (USENIX Security 18), 2018, pp. 33–47.
[5] G. Zhang, C. Yan, X. Ji, T. Zhang, T. Zhang, W. Xu, Dolphinattack: inaudible voice commands, in: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017, pp. 103–117.
[6] N. Komninos, E. Philippou, A. Pitsillides, Survey in smart grid and smart home security: Issues, challenges and countermeasures, IEEE Commun. Surv. Tutor. 16 (4) (2014) 1933–1954.
[7] M. Plachkinova, A. Vo, A. Alluhaidan, Emerging trends in smart home security, privacy, and digital forensics (2016).
[8] H. Touqeer, S. Zaman, R. Amin, M. Hussain, F. Al-Turjman, M. Bilal, Smart home security: challenges, issues and solutions at different iot layers, J. Supercomput. 77 (12) (2021) 14053–14089.
[9] J.S. Edu, J.M. Such, G. Suarez-Tangil, Smart home personal assistants: a security and privacy review, ACM Comput. Surv. 53 (6) (2020) 1–36.
[10] V.S. Gunge, P.S. Yalagi, Smart home automation: a literature review, Int. J. Comput. Appl. 975 (8887-8891) (2016).
[11] C. Jackson, A. Orebaugh, A study of security and privacy issues associated with the Amazon Echo, Int. J. Internet Things Cyber-Assurance 1 (1) (2018) 91–100.
[12] Y. Lit, S. Kim, E. Sy, A survey on Amazon Alexa Attack surfaces, in: 2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC), IEEE, 2021, pp. 1–7.
[13] H. Chung, J. Park, S. Lee, Digital forensic approaches for Amazon Alexa ecosystem, Digit. Investig. 22 (2017) S15–S25.
[14] A. Alhadlaq, J. Tang, M. Almaymoni, A. Korolova, Privacy in the Amazon Alexa skills ecosystem, Star 217 (11) (1902).

[15] J. Blankenburg, Introducing more than 50 features to build ambient experiences, drive growth with Alexa, Alexa Live. https://developer.amazon.com/en-US/blogs/alexa/alexa-skills-kit/2021/07/more-than-50-features-to-build-ambient-experiences.
[16] D. Anniappa, Y. Kim, Security and privacy issues with virtual private voice assistants, in: 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), IEEE, 2021, pp. 0702–0708.
[17] N. Zhang, X. Mi, X. Feng, X. Wang, Y. Tian, F. Qian, Dangerous skills: Understanding and mitigating security risks of voice-controlled third-party functions on virtual personal assistant systems, in: 2019 IEEE Symposium on Security and Privacy (SP), IEEE, 2019, pp. 1381–1396.
[18] N. Apthorpe, D. Reisman, N. Feamster, A smart home is no castle: privacy vulnerabilities of encrypted iot traffic, arXiv preprint arXiv:1705.06805 (2017).
[19] J. Janak, T. Tseng, A. Isaacs, H. Schulzrinne, An analysis of Amazon Echo's network behavior, arXiv preprint arXiv:2105.13500 (2021).
[20] A. Author, BlueBorne cyber threat impacts amazon echo and Google home, Armis. https://www.armis.com/blog/blueborne-cyber-threat-impacts-amazon-echo-and-google-home/.
[21] I. Castell-Uroz, X. Marrugat-Plaza, J. Solé-Pareta, P. Barlet-Ros, A first look into Alexa's interaction security, in: Proceedings of the 15th International Conference on emerging Networking EXperiments and Technologies, 2019, pp. 4–6.
[22] X. Lei, G. Tu, A.X. Liu, C. Li, T. Xie, The insecurity of home digital voice assistants-vulnerabilities, attacks and countermeasures, in: 2018 IEEE Conference on Communications and Network Security (CNS), IEEE, 2018, pp. 1–9.
[23] D. Barda, R. Zaikin, Y. Shriki, Keeping the gate locked on your IoT devices: vulnerabilities found on Amazon's Alexa, Check Point Res. https://research.checkpoint.com/2020/amazons-alexa-hacked/.
[24] J. Cao, D. Bass, Why Google, Microsoft and Amazon love the sound of your voice, Bloomberg. https://www.bloomberg.com/news/articles/2016-12-13/why-google-microsoft-and-amazon-love-the-sound-of-your-voice#xj4y7vzkg.
[25] M. Day, G. Turner, N. Drozdiak, Thousands of Amazon workers listen to Alexa users' conversations, Time. https://time.com/5568815/amazon-workers-listen-to-alexa/.
[26] M. Barnes, Alexa, are you listening?, F-Secure Labs. https://labs.f-secure.com/archive/alexa-are-you-listening/.
[27] I. Clinton, L. Cook, S. Banik, A survey of various methods for analyzing the amazon echo, Citadel Milit. Coll. South Carolina (2016).
[28] M. Ford, W. Palmer, Alexa, are you listening to me? An analysis of Alexa voice service network traffic, Pers. Ubiquitous Comput. 23 (1) (2019) 67–79.
[29] X. Yuan, Y. Chen, A. Wang, K. Chen, S. Zhang, H. Huang, I.M. Molloy, All your alexa are belong to us: A remote voice control attack against echo, in: 2018 IEEE Global Communications Conference (GLOBECOM), IEEE, 2018, pp. 1–6.
[30] D. Overstreet, H. Wimmer, R.J. Haddad, Penetration testing of the Amazon echo digital voice assistant using a denial-of-service attack, in: 2019 SoutheastCon, IEEE, 2019, pp. 1–6.
[31] H. Chung, M. Iorga, J. Voas, S. Lee, Alexa, can I trust you? Computer 50 (9) (2017) 100–104.
[32] S. Kennedy, H. Li, C. Wang, H. Liu, B. Wang, W. Sun, I can hear your alexa: voice command fingerprinting on smart home speakers, in: 2019 IEEE Conference on Communications and Network Security (CNS), IEEE, 2019, pp. 232–240.
[33] C. Wang, S. Kennedy, H. Li, K. Hudson, G. Atluri, X. Wei, W. Sun, B. Wang, Fingerprinting encrypted voice traffic on smart speakers with deep learning, in: Proceedings of the 13th ACM Conference on Security and Privacy in Wireless and Mobile Networks, 2020, pp. 254–265.
[34] M. Rodehorst, Why Alexa won't wake up when she hears her name in Amazon's Super Bowl ad, Amazon Sci.. https://www.amazon.science/blog/why-alexa-wont-wake-up-when-she-hears-her-name-in-amazons-super-bowl-ad.
[35] L. Schönherr, M. Golla, T. Eisenhofer, J. Wiele, D. Kolossa, T. Holz, Exploring accidental triggers of smart speakers, Comput. Speech Lang. 73 (2022) 101328.
[36] P. Swarup, R. Maas, S. Garimella, S.H. Mallidi, B. Hoffmeister, Improving ASR confidence scores for Alexa using acoustic and hypothesis embeddings, in: Interspeech, 2019, pp. 2175–2179.
[37] L. Wang, M. Fazel-Zarandi, A. Tiwari, S. Matsoukas, L. Polymenakos, Data augmentation for training dialog models robust to speech recognition errors, arXiv preprint arXiv:2006.05635 (2020).
[38] D. Bohn, You can finally say 'Computer' to your Echo to command it, Verge. https://www.theverge.com/tldr/2017/1/23/14365338/amazon-echo-alexa-computer-wake-word-star-trek.
[39] B. Sudharsan, P. Corcoran, M.I. Ali, Smart speaker design and implementation with biometric authentication and advanced voice interaction capability, in: AICS, 2019, pp. 305–316.
[40] R. Nash, Amazon Alexa virtual assistant security bug fixed after cybersecurity firm discovered vulnerabilities, 8NewsNow. https://www.8newsnow.com/news/local-news/amazon-alexa-virtual-assistant-security-bug-fixed-after-cybersecurity-firm-discovered-vulnerabilities/.
[41] N. Roy, S. Shen, H. Hassanieh, R.R. Choudhury, Inaudible voice commands: the {Long-Range} attack and defense, in: 15th USENIX Symposium on Networked Systems Design and Implementation (NSDI 18), 2018, pp. 547–560.
[42] S. Gray, Always on: privacy implications of microphone-enabled devices, in: Future of Privacy Forum, 2016, pp. 1–10.
[43] J. Lau, B. Zimmerman, F. Schaub, Alexa, are you listening? Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers, Proc. ACM Hum.-Comput. Interact. 2 (CSCW) (2018) 1–31.