ISSN No:-2456-2165

Unveiling the Veil: A Comprehensive Assessment of Privacy and Security in Amazon Alexa

Deepali Handgar[1]
Department of Computer Science
University of Mumbai
Mumbai, India

Sumon Singh[2]
Department of Computer Science
University of Mumbai
Mumbai, India

Mehdi Rezaei[3]
Prof., Department of Computer Science
University of Mumbai
Mumbai, India

Dr. Jyotshna Dongardive[4]
Head, Department of Computer Science
University of Mumbai
Mumbai, India

Abstract:- In 2020, 53.6 million Amazon Echo Speakers (enabled with Alexa) were sold. This number rose to 65 million in 2021 [1]. Customers can engage with cutting-edge technology in a more natural way by directly dictating instructions to Alexa thanks to its user-friendly, personalised vocal experience. The users' vocal commands are translated into commands by the vendor's cloud services after being transmitted over the internet to the vendor's initial voice assistant devices or companion applications for smartphones and tablets. As this vast amount of data, primarily the user's personal information, moves throughout the voice assistant ecosystem, there are many locations where it is either temporarily or permanently kept. As a result, it is simple for a cyber-attacker to alter this data, raising serious privacy concerns. This, study examines various assaults on the Amazon Alexa ecosystem, including assaults on speech recognition and processing on the cloud backend and frontend audio capturing. We also discuss potential attack surface reduction techniques to make Alexa and other voice assistants more private and secure.

Keywords:- *Privacy, Security, Alexa, Attacks.*

I. INTRODUCTION

ALEXA- The design of Amazon Alexa, often known as Alexa, was inspired by the Polish speech synthesizer Ivona, which Amazon acquired in 2013. In Nov, Alexa was announced alongside Echo.

Alexa is built using nlp (Natural language processing), a collection of methods to convert speech into text and audio. In basic terms, when you talk to Alexa, Amazon keeps a record of what you say. These recordings are sent to Amazon's servers for analysis. Amazon breaks down your commands into separate sounds and tries to find the closest matching words by using a library of word pronunciations. This helps Alexa understand what you want and carry out the tasks you request. As an example, if you mention the words "sport" or "basketball," Alexa would activate the sports app. Your device receives this information from Amazon's servers, and then Alexa might start speaking. If Alexa needs to respond, she would follow the same steps as mentioned earlier, but in reverse order. [2]

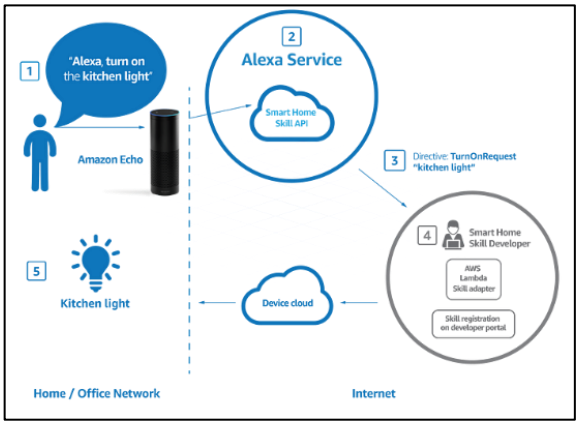

Fig 1: Alexa Ecosystem

Since launch, Alexa has gained more than 100,000 skills [1].

➢ *ALEXA ecosystem:*

Alexa is incorporated in Echo devices which contain hardware, software, and cloud-based components. With this, the customers can interact with Alexa. These devices are equipped with various technologies, including 'Automatic Speech Recognition' (ASR), 'Natural Language Understanding' (NLU), 'Text-to-Speech' (TTS), and 'Response' [8]. ASR allows us to interact with computer interfaces in a way that resembles natural human conversation. Similarly, TTS enables the accurate reproduction of human speech, while NLU empowers computers to comprehend the structure and significance of human language [8]. The Alexa system utilizes all these capabilities to formulate appropriate "responses" to user requests. Additionally, third-party services can contribute their own "skills" to provide responses. By integrating skills and voice-activated software, Alexa can continuously expand its range of functionalities, enhancing the overall user experience and customization options. The Amazon Alexa ecosystem revolves around two main components: the customer home environment and the Alexa cloud backend. The customer home environment includes devices like the

ISSN No:-2456-2165

Alexa smartphone app, Amazon Echo, and other IoT gadgets, creating a smart living space for users who have embraced Alexa as part of their connected home setup. The Alexa cloud backend, on the other hand, relies on Alexa voice services and additional Amazon Web Services (AWS) cloud offerings. These include the Lambda serverless computing service, the DynamdoDB database service, and the Amazon S3 storage service [9]. Together, these cloud services support Alexa's functionalities and enable seamless communication between the user's home devices and Amazon's cloud infrastructure..

II. PRIVACY CONCERNS

Users can interact with Alexa through Amazon Alexa skill, which is an application built by a third party. Examples include Spotify for audio, Easy Meal Ideas for recipes, and Alexa Guard for home protection. Third-party skills are not required by the Amazon privacy policy to reveal the methods used for data collection and usage. In the US, just 28.5% of third-party services have proper privacy rules that clearly outline how they collect and use user data. Even more concerning is that only 13.6% of skills targeted at minors have legitimate privacy policies in place [3].

A formal description of an attack surface for a computer system is the total number of potential entry sites or "attack vectors" (also known as unauthorized users or "attackers") from which data can be taken or extracted. Any software must take precautions to keep its attack area as small as possible [4]. Alexa has extensive access to a wide range of customer data, going beyond the typical information used by smartphones and laptops, such as messages and schedules. It can also control and gather data from various sources, including house locks, personal shopping lists, speech recordings, conversations, and customer voice profiles. Moreover, with the integration of "smart home" devices like "smart locks" with gate codes and key codes, as well as security cameras providing live footage of customer homes, Alexa has gained unparalleled access to customer data, surpassing many other modern software systems in terms of the data it can handle.

There have been concerns raised regarding Amazon's potential access to private conversations within homes and other non-verbal cues that could reveal who is present. These concerns arise from the fact that Alexa-enabled devices continuously listen to audio in their surroundings. In response to these worries, Amazon asserts that the devices only transmit recordings from the user's home when the user activates them using the "wake word." [5]

In 2018, after learning that one of her husband's co-workers had obtained audio recordings of the couple conversing inside their house, a Portland woman going by the name of Danielle contacted Amazon to voice her displeasure. Amazon claimed its product captured conversational cuts on several unfortunate occurrences and mistakenly picked up the background noise [6].

In early 2018, security researchers at Checkmarx demonstrated the capability to transform an Amazon Echo device into a covert spying tool. They achieved this by creating a malicious Alexa Skill that could secretly record unsuspecting users and then send the transcriptions of their conversations to an attacker [7]. To use Alexa, you first need to use the wake word for the device. The user then would ask Alexa something and Alexa would respond accordingly. Usually, the conversation ends here and the mic stops transmitting any further sound. Using the developer function “shouldEndSession”, the researchers programmed the skill such that Echo could keep listening. During their research, the investigators discovered that Alexa typically issues a verbal prompt, known as "readback," to inform the user that it is still listening and active. However, they found a way to exploit this feature by entering null values instead of actual words. This caused the Echo to remain silent, making it not notify the user that the session was still ongoing [6]. This eavesdropping strategy has some obvious drawbacks. Attackers would only have received transcriptions of a target's talks, not audio recordings. Furthermore, despite being theoretically limitless, even a prolonged session would in reality only last for a few minutes before the Echo would usually terminate it. As a result, attackers would only be able to monitor users for brief periods of time, and even then, only after they purposefully engage with a compromised Echo [7].

III. FORMS OF ATTACKS:

Due to their ability to access user voice commands, Skills—speech programmes that handle user requests—have emerged as the new target for attackers. Malicious skills have the capability to either steal user data or manipulate voice commands to perform actions that deviate from the user's intended command.

A. Skill misuse

Although skills increase Alexa's functionality, they also open up a new path for the attack. By taking advantage of anticipated mistakes like homophones, complex words, and phonetic confusion, skill squatting attacks misdirect users to harmful skills. Attackers create harmful skills with names that sound similar to legitimate ones, making it easy to confuse users seeking a genuine skill. This phonetic confusion could lead users to unintentionally access malicious skills while trying to find a legitimate one. Once a malicious skill gains access to a user's device, it can launch further attacks. This type of attack, known as skill squatting, is akin to domain name typo-squatting seen in web applications, where attackers exploit common typos in domain names to deceive users.

In a study conducted by Kumar et al. [11], they created multiple pairs of skills with identical invocation names to investigate whether Alexa activates the squatted skill instead of the intended target skill. They built a total of 27 pairs of skills, each consisting of a target skill and a squatted skill. Interestingly, 92.6% of the time, Alexa activated the squatted skill instead of the intended one, as 25 out of the 27 pairs of skills were successfully squatted at least once. Spear

skill squatting is an advanced form of skill squatting where attackers specifically target a particular demographic or group. In this technique, the attackers use "squattable" terms in the language specific to the targeted population to execute skill squatting attacks [10].

Mitigation: To minimize the vulnerability to skill squatting, one effective approach is to introduce a screening process during skill certification. This procedure would assess whether a skill could potentially be confused with another registered skill, helping to identify and prevent instances of deceptive naming. For example, if there are multiple skills with the same name, such as "Cat Facts," implementing a screening mechanism could aid in ensuring that user requests are appropriately routed to the intended skill. Despite the existence of 30 skills with the name "Cat Facts," the specific routing method used by Amazon remains unclear [10]. By giving every skill a distinct name during the screening process, such vulnerable skills can be reduced. Although this defence may lessen the vulnerability, skill publishers might disagree on the names of certain skills. They might prefer a straightforward skill name, which could result in name shortages.

B. Voice Masquerading Attack

Voice Masquerading Attack (VMA) operates stealthily, leaving users unaware that their conversations are being eavesdropped on. This covert approach creates an opportunity for malicious actors to exploit the vulnerability and illicitly obtain a user's personal data. There are primarily two categories of VMAs [12]:

- Switching skill during communication
- Pretending to terminate

An opportunistic attack known as a, "in-communication skill switch" occurs when one skill impersonates another. When a user attempts to switch between skills while engaging with Alexa, there is a potential risk of an attack taking place. By imitating the target skill, a malevolent skill claims to provide the target skill execution. This poses a major privacy risk since the user might provide malevolent skill access to data intended for the target skill.

In order to spy on a user, malevolent skills will fake termination as a VMA. Users can determine skill termination by looking at the response from the skill. Users can assume that a skill has been terminated if it prompts "goodbye" or is silent after execution. Malicious skills may fabricate termination while continuing to listen in on users of Amazon Echo and gather sensitive data.

Mitigation: The VMA mitigation strategy makes use of the user's command and skill in order to launch an assault. The mitigation mechanism consists of two main parts: the intention classifier and the skill response checker. In order to determine an attack, the Skill Response Checker (SRC) checks a skill's questionable response. SRC uses a collection of typical Alexa response patterns to determine whether a skill, duplicates Alexa's responses. Every time a comparable response is found, an alarm is set off to alert Alexa of the situation. The User Intention Classifier (UIC) plays a vital role in identifying whether a user unintentionally attempts to switch to a different skill by analyzing their voice commands. To verify the user's intent, UIC relies on the semantics and context of the command. Users can influence the context by using terms closely related to Alexa, like "open sleep sounds." To understand the user's intention more accurately, UIC also compares their commands with system commands and considers the context of their previous interactions with skills. However, discerning user intent can be complex since users' device usage behavior can vary, and finding a consistent usage pattern can prove challenging. Utilizing natural language processing, the system strives to comprehend the command's context. While technology is beneficial, it continuously evolves, presenting new challenges in ensuring accurate intent recognition and enhancing user experience.

C. Attack on ASR

The objective of Automatic Speech Recognition (ASR) is to convert spoken words into written text. However, an attacker can disrupt this process by modifying the audio signal. They can create a specific audio signal or introduce noise to an existing signal, which is referred to as adversarial examples or obfuscated examples when used to challenge deep learning systems. In this study's context, an obfuscated example is an audio signal that the personal voice assistant perceives as a command but appears as noise to humans. On the other hand, an adversarial example aims to deceive the personal voice assistant while appearing harmless as a regular audio transmission to humans.

Adversarial examples can be classified into two types: targeted and nontargeted. In a targeted adversarial example, the attacker's focus is on causing a specific command transcription error. They are interested in manipulating the ASR to produce a particular wrong command. On the other hand, in a nontargeted adversarial situation, the attacker is not concerned about a specific command being decoded incorrectly. They aim to create any kind of transcription error in the ASR, regardless of the exact command that results.

For attackers to construct adversarial or disguised examples, having access to the ASR's internal workings provides an advantage. A white-box attack relies on exploiting the ASR's internal knowledge, including its structure and parameters. On the other hand, in a grey-box setup, the attacker lacks access to the target model's internal details, such as its structure and parameters. In order to create adversarial examples, attackers can query the output of the target model's last layer, in addition to obtaining numerical confidence or prediction scores. This querying process allows the attacker to gather relevant information to craft effective attacks without complete knowledge of the target model's internal workings [13].

The attack's optimization process heavily relies on these parameters to guide its actions. In a black-box scenario, the attacker is restricted to observing only the final decision outcomes, making it the most challenging but also

the most probable option in real-world scenarios. To gauge the effectiveness of hostile command attacks, the attack success rate, also known as accuracy, efficiency, or effectiveness, is calculated. A successful attack occurs when the ASR accurately transcribes every single word in the target command. The success rate, which evaluates sentence-level accuracy, is calculated as the proportion of successful attacks out of all the attack trials [13].

D. Voice Capture

The Amazon Echo operates by requiring a wake word to initiate voice recording. Prior to detecting the wake word, the device remains in a dormant state, buffering and re-recording audio. In an experiment conducted by M. Ford and W. Palmer [14], they analyzed the network traffic of Echo Dots in a residential home for 21 days. During this period, no intentional interactions occurred with the devices using a wake word. Despite the lack of intentional commands, the analysis of the recorded audio revealed that 30% of the recorded content consisted of human voices, while 70% was attributed to television noises. This finding highlights a concerning revelation that Amazon Echo has the potential to record private conversations without the explicit use of a wake word. This unintended audio capture poses a significant privacy risk, particularly if private or sensitive audio is accidentally exposed or accessed by unauthorized individuals.

To address this concern, a practical solution is for users to physically turn off the microphone of the device before speaking any private information aloud. When the microphone is disabled, Alexa ceases to send audio data to the Amazon AVS cloud, ensuring that no unintended recordings occur. However, despite the availability of the mic button and its functionality, many users tend to underutilize it. Some users perceive that turning off the microphone compromises the device's hands-free functionality, which might discourage them from utilizing this safeguard regularly.

IV. LITERATURE REVIEW

In their research, Pradhan et al. [15] propose a voice replay detection system as a defense against replay attacks. This technology offers comprehensive room-scale detection without the need for wearable devices, ensuring privacy protection. By leveraging speech and WiFi features, the system becomes capable of identifying various types of replay assaults. It capitalizes on inherent differences between live and recorded voices, as well as human breathing patterns during speech, which are captured through WiFi signals. The system has proven successful in detecting replay attacks; however, the authors acknowledge the potential for further improvements. To enhance its capabilities, they suggest expanding the detection range, which is currently limited to two meters. Additionally, diversifying the training datasets would enable better generalization of the voice and WiFi models, thus enhancing overall accuracy. Moreover, the inclusion of additional physiological signals like heart rate and other biometric measurements could further bolster the system's replay detection capabilities. These advancements would reinforce the system's effectiveness in thwarting replay attacks and contribute to stronger security measures in voice-based authentication systems.

The research conducted by M. Ford and W. Palmer [14] brought valuable insights to the understanding of Echo and Alexa network behavior by addressing three key research questions and identifying discrepancies in the quantity of logged response cards to AVS data points. This study's findings have contributed significantly to the knowledge base on how Echo and AVS interact. The results of this study have raised new inquiries and sparked curiosity about Echo's voice training process and its potential impact on reducing unintentional audio recordings. Furthermore, the study piques interest in exploring whether other Alexa services, such as making purchases from Amazon's shop, generate distinct and observable network traffic patterns.

To gain a comprehensive understanding and answers to these additional queries, further research is required in this specific area. By delving deeper into the network behavior of Echo and AVS, future studies can shed more light on optimizing these voice-assisted technologies and enhancing their privacy and security features.

In the study conducted by Mitev et al., they employed a combination of third-party extensions and new malicious skills to demonstrate a man-in-the-middle attack on Alexa, which proved to be more potent than previously believed possible. To circumvent Alexa's skill interaction architecture, the authors discovered that skill functionality could be exploited when coupled with popular inaudible (ultrasound) attack techniques. This allowed malicious attackers to gain control and manipulate interactions between the user and other benign skills.

The attack was skillfully crafted, making it difficult for users to detect, and it had the capability of intercepting and influencing discussions between the user and the voice assistant. What made this assault particularly powerful was its ability to operate during active user engagement, precisely when the user was speaking to the device and continuing the conversation from their perspective. This active manipulation of the user's interactions went beyond merely issuing straightforward, pre-prepared commands, enabling the attacker to have a more significant impact on the user's interactions with Alexa. The study revealed crucial vulnerabilities that necessitate robust security measures to safeguard voice-activated systems like Alexa from such sophisticated attacks.

In Zhang et al.'s research, they introduced the concept of the "dolphin attack," wherein voice instructions beyond the human hearing range (frequency > 20 KHz) can interact with voice assistants (VAs). The paper not only demonstrated the success of this attack on various VAs, such as Siri, Cortana, Google Now, Samsung's S Voice, and Alexa, but also provided a quantitative analysis of different attack parameters. The researchers conducted experiments and organized the results in a table format, outlining the VA

device used, whether the command was understood, and if the command activated the device. Additionally, they recorded the greatest distance from the microphone on the VA device, highlighting that proximity to the target is crucial for the attack's effectiveness. Notably, the maximum distance recorded was 1650 mm, indicating that the attacker still needs to be relatively close to the target VA for the attack to work.

Zhang et al.'s study sheds light on the potential security risks posed by inaudible voice commands and the importance of implementing measures to safeguard voice assistants from such sophisticated attacks.

Yıldırım et al.'s paper presented an exploration of Amazon and Google Voice Assistants (VAs) as a potential source of digital forensic evidence [17]. The study conducted a quick qualitative analysis to examine activity history records associated with voice commands given on a VA device, specifically a Samsung smartphone. The research methodology involved searching for data related to voice commands, and the collected information included timestamps, the written form of the voice command, and the corresponding assistant's response. By analyzing this data, the researchers aimed to understand the potential implications of using VAs as a valuable resource for digital forensic investigations.

In the study conducted by W. Li et al. [18], the authors proposed a novel approach to enhance the security of user voice instructions using granule computing technology. Their research focused on 'encrypting' user voice commands to protect them from potential attacks. The success of their approach was quantitatively evaluated in the article.

Unlike existing voice assistant (VA) client endpoints that heavily rely on cloud processing, the authors' method shifted most of the computing tasks to the VA device itself. This decentralized approach aimed to minimize the risk of exposing sensitive voice data during cloud processing, where potential security vulnerabilities may exist. To ensure robust encryption, the authors employed the advanced encryption standard (AES) to encrypt each sound with a unique key for every voice command. This added layer of encryption made it more difficult for hostile attackers to crack the content's encryption and gain unauthorized access to user voice instructions.

In a study conducted by Lau et al. [19], the researchers investigated end-users' behavior regarding voice assistants (VAs) and their perceptions and concerns related to privacy. The study employed qualitative analysis through structured interviews and a diary study. The diary study involved 17 users and an equal number of non-users of VAs who participated in semi-structured interviews. Over the course of a week, the users recorded instances of using the VA device and unintentional wake-word triggers at least once a day. Afterward, interviews were conducted with the users, focusing on details like the device's placement and reasons for their choices. For non-users, the interview questions centered on the reasons behind their decision not to use a VA and any privacy concerns that influenced their choice. To analyze the interviews, the researchers utilized a derived codebook to identify recurrent themes and emerging categories.

The study's findings shed light on speaker usage trends, who was responsible for placing the speaker (the user or someone else), and the ideal speaker placement to enhance perceived privacy, particularly in living rooms. This research provides valuable insights into user behaviour and privacy concerns associated with voice assistants, which can inform the design and development of future voice-enabled technologies.

In the research conducted by Turner et al. [20], they demonstrated a security attack known as "phoneme morphing" targeting voice assistants (VAs). This attack involved deceiving the VA by impersonating the registered user's voice through a modified recording of the attacker's voice. The researchers quantitatively examined the variation in several characteristics of this attack. The attack was based on a technique that translated phonemes, the smallest units of sound in the English language (44 in total), from a known speaker to mimic the victim's speech. The process involved three phases: first, recording the victim's speech to map the phonemes between the source and target; next, modifying the audio to resemble the victim's voice; and finally, broadcasting the modified audio to the VA system. The phoneme clustering of the source voice was performed offline.

To evaluate the attack's effectiveness, four keywords were used, revealing a significant range in success rates. The lowest success rate was 42.1%, while the highest was 89.5%. These results highlight the potential security risks associated with voice impersonation attacks on VAs and underscore the need for robust defenses to safeguard user privacy and device integrity.

V. CONCLUSION

In this paper, we have looked at several types of attacks that have been made against Amazon Alexa as well as potential future threats. We have determined the significance of proactively addressing these security threats by examining past instances and vulnerabilities. We've also outlined a number of ways to reduce these dangers and improve Amazon Alexa's overall privacy and security.

We have emphasised the need for safe authentication methods through our review in order to guard against unauthorised access to Alexa devices. In order to stop future exploitation, we have also emphasised the importance of ongoing monitoring and timely vulnerability patching. In addition, we have discussed the significance of unambiguous user permission and strong data encryption procedures in data collecting and storage practises.

Furthermore, we want to emphasise the importance of improved user controls and privacy settings on Amazon Alexa as a result of our investigation into privacy issues.

Privacy hazards may be greatly reduced by giving consumers the ability to manage their data and choose how much information is shared with the voice assistant.

The overall goal of this article is to increase understanding of potential security and privacy issues related to Amazon Alexa while providing workable solutions.

VI. FUTURE WORKS

There are a number of areas where we might concentrate on future improvements to better strengthen the privacy and security of Amazon Alexa.

First off, enhancing Alexa's understanding of and ability to effectively respond to user requests while retaining privacy may be done by investing in powerful machine learning algorithms and natural language processing skills. Inadvertent triggers or unauthorised access to sensitive material would be less likely as a result.

Second, Amazon can keep improving the way it gathers and stores data. While maintaining personalised experiences, privacy may be improved by putting in place more stringent restrictions and providing users with more granular alternatives to manage their data.

Additionally, regular safety inspections and vulnerability evaluations must be carried out to spot and swiftly remedy any possible security vulnerabilities. By taking preventive measures, we can guard against assaults in the future and maintain Alexa's resistance to new dangers.

Furthermore, encouraging engagement with the security research community can promote ethical vulnerability disclosure, facilitating quicker detection and resolution of possible security concerns.

Last but not least, giving customers the tools, they need to make educated decisions about their data and privacy settings includes offering clear and simple privacy regulations and boosting user education about Alexa's privacy and security capabilities.

REFERENCES

[1]. Intriguing Amazon Alexa Statistics You Need to Know in 2023
[2]. https://safeatlast.co/blog/amazon-alexa-statistics/#gref
[3]. How Amazon Alexa works? Your guide to Natural Language Processing (AI)
[4]. https://towardsdatascience.com/how-amazon-alexa-works-your-guide-to-natural-language-processing-ai-7506004709d3#:~:text=How%20does%20Alexa%20work%3F,Amazon%20records%20your%20words.
[5]. Amazon Alexa Skills Present Security Risks
[6]. https://www.esecurityplanet.com/trends/amazon-alexa-security-risks/
[7]. P. Manadhata and J. M. Wing, “Measuring a system’s attack surface,”
[8]. Carnegie Mellon University, Tech. Rep., 2004.
[9]. How private is Amazon Echo?
[10]. https://www.slashgear.com/how-private-is-amazon-echo-07354486
[11]. Is Alexa Spying on You? Amazon Responds After Rogue Echo Device Leaks Couple's Private Chat
[12]. https://www.newsweek.com/alexa-spying-you-amazon-responds-after-rogue-device-secretly-records-private-944557
[13]. Turning an Echo into a Spy Device Only Took Some Clever Coding
[14]. https://www.wired.com/story/amazon-echo-alexa-skill-spying/
[15]. N. A., “White paper - alexa privacy and data handling overview,” July 20, 2018, retrieved October 23, 2020. [Online]. Available:
[16]. https://d1.awsstatic.com/product-marketing/A4B/White Paper - Alexa Privacy and Data Handling Overview.pdf
[17]. Li Y., et al, “A Survey on Amazon Alexa Attack Surfaces”, 2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC), 2021
[18]. Pathak S., et al, “A survey on security analysis of Amazon echo devices”, High-Confidence Computing, Elsevier B.V., 2022
[19]. 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.
[20]. 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.
[21]. P. Cheng and U. Roedig, “Personal Voice Assistant Security and Privacy—A Survey,” Creative Commons Attribution 4.0 License, Proceedings of the IEEE | Vol. 110, No. 4, April 2022.
[22]. 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.
[23]. S. Pradhan, W. Sun, G. Baig, and L. Qiu, “Combating Replay Attacks Against Voice Assistants,” Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, pp. 100:1–100:26, 2019.
[24]. Mitev, R.; Miettinen, M.; Sadeghi, A.R. “Alexa Lied to Me: Skill-based Man-in-the-Middle Attacks on Virtual Assistants”. In Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security, Asia CCS ’19, Auckland, New Zeland, 9–12 July 2019.
[25]. Yıldırım, ˙I.; Bostancı, E.; Güzel, M.S. Forensic Analysis with Anti-Forensic Case Studies on Amazon Alexa and Google Assistant Build-In Smart Home Speakers. In Proceedings of the 2019 4th International Conference on Computer Science and Engineering (UBMK), Samsun, Turkey, 10–15 September 2019.

[26]. Li,W.; Chen, Y.; Hu, H.; Tang, C. Using Granule to Search Privacy Preserving Voice in Home IoT Systems. IEEE Access 2020, 8, 31957–31969.
[27]. Lau, J.; Zimmerman, B.; Schaub, F. Alexa, Are You Listening? Privacy Perceptions, Concerns and Privacy-seeking Behav-iors with Smart Speakers. In Proceedings of the ACM on Human-Computer Interaction, 2018; Available online: https://www.key4biz.it/wp-content/uploads/2018/11/cscw102-lau-1.pdf (accessed on 25 March 2020).
[28]. Turner, H.; Lovisotto, G.; Martinovic, I. Attacking Speaker Recognition Systems with Phoneme Morphing. In Proceedings of the ESORICS 2019: Computer Security, Luxembourg, 23–27 September 2019.