Private Data Harvesting on Alexa using Third-Party Skills

Jack Corbett and Erisa Karafili[0000−0002−8250−4389]

School of Electronics and Computer Science,
University of Southampton, UK
{jc11g17, e.karafili}@soton.ac.uk

Abstract. We are currently seeing an increase in the use of voice assistants which are used for various purposes. These assistants have a wide range of inbuilt functionalities with the possibility of installing third-party applications. In this work, we will focus on analyzing and identifying vulnerabilities that are introduced by these third-party applications. In particular, we will build third-party applications (called Skills) for Alexa, the voice assistant developed by Amazon. We will analyze existing exploits, identify accessible data and propose an adversarial framework that deceives users into disclosing private information. For this purpose, we developed four different malicious Skills that harvest different pieces of private information from users. We perform a usability analysis on the Skills and feasibility analysis on the publishing pipeline for one of the Skills.

1 Introduction

The Internet of Things (IoT) is a growing phenomenon that refers to embedding internet connections in everyday objects. These objects range from small items like light bulbs and cameras to vast sensor networks capable of monitoring road networks and cities. IoT technology promises to create smart-connected homes where all household items and utilities can communicate to increase convenience and our quality of living. Many such devices are already available which enable control of your home's lighting, heating, appliances, security, and entertainment. All these devices can be controlled through smartphones, tablets, and other traditional computing devices but they also integrate with a wide range of voice assistants.

Voice assistants, also known as intelligent virtual assistants (IVAs) or intelligent personal assistants (IPAs), are software agents that can perform tasks based on natural language input such as commands or questions. There are many voice assistants available, but one of the current market leaders is Alexa [20], a voice assistant created by Amazon, which is integrated into their range of Echo smart speakers. Smart speakers are speakers with built-in microphones to record user's voice commands when a wake word is spoken. This enables users to control their smart home devices, play music, and set reminders along with a plethora of other functionality. However, having an always-listening, internet-connected,

device does come with concerns around security and privacy [4,10], that emphasise the existing challenges on authentication [11], threat discovery [19] and data access control [12,9].

Alexa integrates with a wide variety of products and services in order to provide more relevant and accurate answers. To achieve this, third-party developers build '*Skill*s' which supplement the range of inbuilt commands. The Skills vary in complexity from basic trivia games to controlling IoT devices. Skills can be installed by voice command or through the IVA mobile app and can request access to user's personal data to enrich the experience. Skills pose a threat, as they are an entry point for malicious actors to attempt to compromise user data [24]. The impact of such attacks could be wide-ranging, as the devices could help an attacker phish for a user's personal information such as their address, password or payment details.

In this work, we identify vulnerabilities in Alexa by establishing whether the tools provided by Amazon are sufficiently controlled. For the purpose of this investigation, we will look at the data which can be compromised through the legitimate development of a third-party Alexa Skill. We will not exploit vulnerabilities in the hardware or software of the device itself, only the tools provided to the developers and the publishing process.

To identify the vulnerabilities that are introduced by possible malicious Skills, we will develop an adversarial framework. On top of this framework, we will build a range of Skills that employ different strategies. Our goal is to collect different types of private information from the user without triggering suspicion from the Alexa team or the users themselves. In particular, we will develop four different Skills that will collect private information like their home address, password, credit card number, security code, and expiration date.

To ensure that the developed Skills are appropriate we evaluate their usability through a study. In this usability study, different users are asked to interact with different Skills (malicious and genuine). The result of the study is promising, as the users scored very high on the usability of our malicious Skills. On the other hand, the Skills that asked for credit card information raised some trust issues. We also performed a feasibility study where we tested the Skill publishing pipeline. We found that it was possible to publish a Skill, without the malicious component and update it with the malicious parts without raising a re-certification or issue.

The paper is organized as follows. In Section 2, we provide some background information about Alexa Skills, how they operate, some past attacks and related work. We introduce the adversarial framework our Skills will be built upon and how the data will be collected in Section 3. The four Skills and the description of their attacks are given in Section 4. The evaluation of the usability and feasibility of the Skills is introduced in Section 5. We finish in Section 6 with the conclusions, future works, and some discussions on how to strengthen Alexa's publishing pipeline and prevent these vulnerabilities.

2 Introduction to Alexa's Skills and past Attacks

In this section, we first introduce the Alexa Skills, how they are activated and where the information is stored. We also introduce past known attacks on Alexa and in particular on Alexa Skills.

2.1 Alexa Skills

A Skill is an application that can be developed by third parties to add new functionality to Alexa. They are structured differently from traditional applications as they are built entirely around natural language input, meaning responses/actions are defined and activated based on the command administered by the user.

The *voice interaction model* defines the complete flow of how users interact with a Skill. Voice commands are broken down into multiple components to ensure they are processed by the corresponding program functionality.

- Wake Word: The term used to trigger the device to start recording, for example: 'Alexa'. This is detected by a local *automatic speech recognition* (ASR) system on the device. The rest of the requests are processed in the cloud.
- Invocation Name: The keyword which is used to trigger a specific Skill.
- Intents and Slots: Intents are the tasks the Skill can carry out which take arguments called slots. Slots allow the user to provide data directly to the Skill.
- Sample Utterances: Phrases the user will say to trigger an intent. Multiple utterances should be defined for each intent to cater to the variability of natural language. An example of this would be the command: 'give me a fact' and 'tell me a fact'.

Once the Alexa-enabled device has recorded the request, the audio is sent to the *Alexa Voice Service* (AVS). AVS is a cloud-based system that performs ASR to identify the user's intent and calls the corresponding backend function. To decide on the action to take, it uses natural language understanding to calculate the probability of a user's intent based on keywords and language rules [8]. Afterward, text to speech (TTS) is used to build the audible response.

Once the AVS has identified that the user is making a request to a Skill, it calls the Alexa Skills Kit which acts as the frontend for Alexa Skills. It has a reference to the backend which is a web service endpoint that runs the Skill logic. Once the user's intent has been identified, it sends a request to the backend in JSON [1]. The backend handles the processing and provides a response, also in JSON format. This can be supplemented with Speech Synthesis Mark-up Language (SSML) which controls Alexa's vocal delivery. It is also possible to host your own web service to handle Alexa requests. Before a Skill is published to the store for users to install, it must pass a *certification process*. Amazon provides a submission checklist for developers and it must also follow Alexa policy guidelines and meet security requirements. To update a Skill, officially, it must pass the certification process again.

2.2 Previous Exploits on Alexa and Related Work

Let us now give an overview of the most interesting vulnerabilities discovered in Alexa. The original Amazon Echo was vulnerable to a physical attack due to debug pads on the device's base. This enabled an attacker to root the device, turning it into a wiretap [5]. It posed a major threat but required physical access to the device and considerable time to exploit, which decreased the risk. A vulnerability like this would be a greater issue today, given the number of Alexa devices and their increased use in shared spaces like hotels [22]. The vulnerability was removed in 2017.

Other exploits are based around nuances with spoken language input. The first to be discovered were Voice Squatting [15] and Voice Masquerading attacks [18]. Voice Squatting involves setting the invocation name of a malicious Skill similar to that of a genuine Skill, either by removing words or using homonyms. The user then believes they are communicating with a service they trust when in fact Alexa has enabled the malicious Skill. Voice Masquerading is similar but instead comes into effect when a user tries to switch between Skills. If they do not preface their switch command with the wake word, Alexa thinks they are making a request to the current Skill. This can be exploited by including an intent that matches the name of another popular Skill. The user then believes they are communicating with a different Skill and the malicious Skill can emulate the responses in order to attempt to compromise user data [24]. Another similar attack is the Surfing attack [23]. This exploit allows attackers to interact with an IVA over a long distance.

In 2019, Security Research (SR) Labs demonstrated two vulnerabilities that can exploit both Alexa and Google Assistant enabling vishing/phishing and eavesdropping attacks [14]. The first uses silence to convince the user that the Skill has stopped executing, before asking them to confirm their password to install a software update. SR Labs also demonstrated that the backend code for an Alexa Skill can be updated without requiring it to be resubmitted for the certification process [2]. This enables malicious actors to completely change how the user's data is processed and the responses Alexa gives. The only aspect that cannot be changed is the frontend (meaning new intents and slots) cannot be added without having to re-certify.

Some solutions have been proposed for preventing such attacks. However, they do not provide a solution for the vulnerabilities in the publication pipeline that are identified in this paper. The work [21] proposes a solution that is an extension introduced between the user and the voice assistant. This extension ensures that certain security and privacy properties are respected. Another similar solution is introduced in [7], which proposes a system that jams the microphone to prevent it from recording the user's speech. A study about older adult interactions with voice assistance and ways to reduce their privacy risks have been proposed in [6].

Studies on the required privacy policy for third-party Skills and the certification process have been conducted recently. The following work [17] provides a study on the effectiveness of the privacy policies provided by the app developers

but they do not analyze the possibility of malicious apps. Furthermore, another study [3] was conducted that analyzed the certification process of the Skills for Alexa and Google Assistant. The work in [16] performs a systematic analysis of the Alexa Skill ecosystem and identifies similar vulnerabilities as in our paper. While the approaches in [3] and [16] are wider as large numbers of Skills are analyzed, in our paper, we go for a more specific approach, as we develop four new malicious Skills and explain in detail how the information is harvested and how the adversarial framework eludes the publication pipeline.

3 Overview of the Adversarial Framework

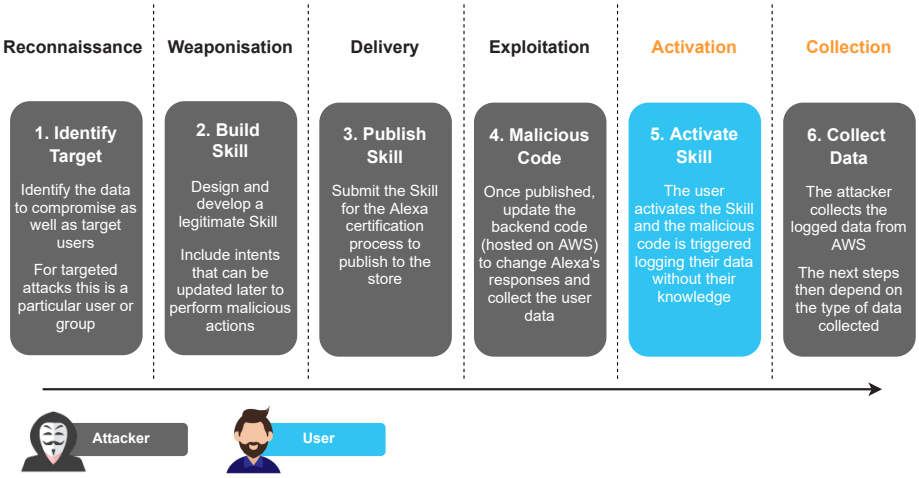

Fig. 1. Adversarial Framework Construction

Let us now introduce our adversarial framework, where our malicious Skills will be based upon. First, the attacker needs to identify the target and the data they want to compromise. The *weaponisation* stage involves the attacker building the IVA Skill. The initial version of the Skill does not include any malicious code but is designed to support malicious actions once the Skill is published. The code is delivered to the IVA store for certification. Once the Skill is approved and published, the attacker exploits the ability to update the backend code without triggering re-certification. The updated code implements malicious actions and can include changing or removing the IVA responses, introducing silence (in order to convince a user that the Skill interaction has ended), or changing the Skill's functionality to simulate an existing Skill. A graphical representation of the various phases of our adversarial framework is provided in Figure 1.

Let us provide an overview (see Figure 2) of how the adversarial framework works and how it is able to compromise the user's data. The user activates

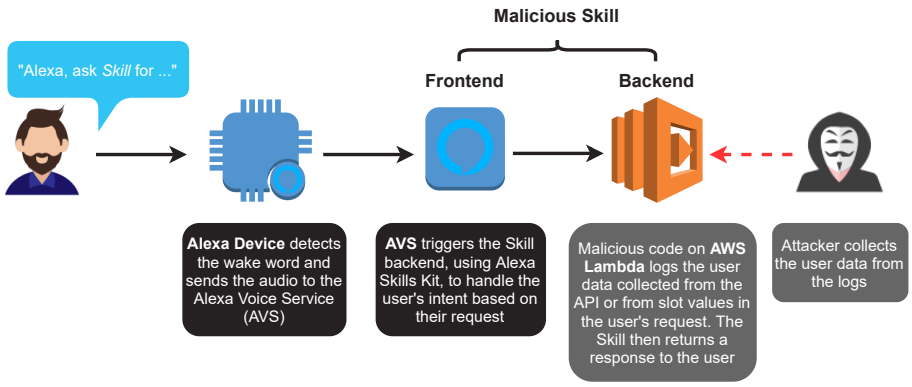

Fig. 2. Overview of the Adversarial Framework Attack Flow

Alexa and includes the Skill's name in the request. The audio is recorded by the Alexa enabled device and is interpreted by the voice service (AVS), which uses the Alexa Skills Kit to start an AWS Lambda instance. The latter contains the Skill's backend logic that includes the malicious code which will be used to collect the user's data. If this attack is successful, the user's data is compromised and collected by the attacker from the backend logs. The attacker's next action depends on the collected data, e.g., they could use compromised payment details to make fraudulent purchases or sell the users account credentials online.

3.1 Data Compromising through the Adversarial Framework

There are two main approaches to compromise users' data. If the information is available through the Alexa data API, provided the user accepts permission for the Skill to access the data, it can be collected directly. The user only needs to accept permission once, when they first access the Skill. However, the data available is limited and it must be believable to both Amazon's approval process and the user that the Skill requires the data to function.

Alternatively, the data can be collected directly from the user using slot values. Once the Skill has been approved the attacker can update the backend code to ask for different information in the existing slots. The diagram in Figure 3 demonstrates the flow of information between the user, AVS and malicious Skill. The diagram also shows how the attacker can harvest the collected information. The Skill can be hosted in AWS (as shown in Figure 3) or self-hosted. We decided that the attacker would collect the harvested data from the AWS CloudWatch Logs or if the Skill is hosted on their infrastructure, directly from their own server logs. This decision was made to avoid unnecessary external API calls with hosting on AWS, as it was sufficient to collect the data from the logs.

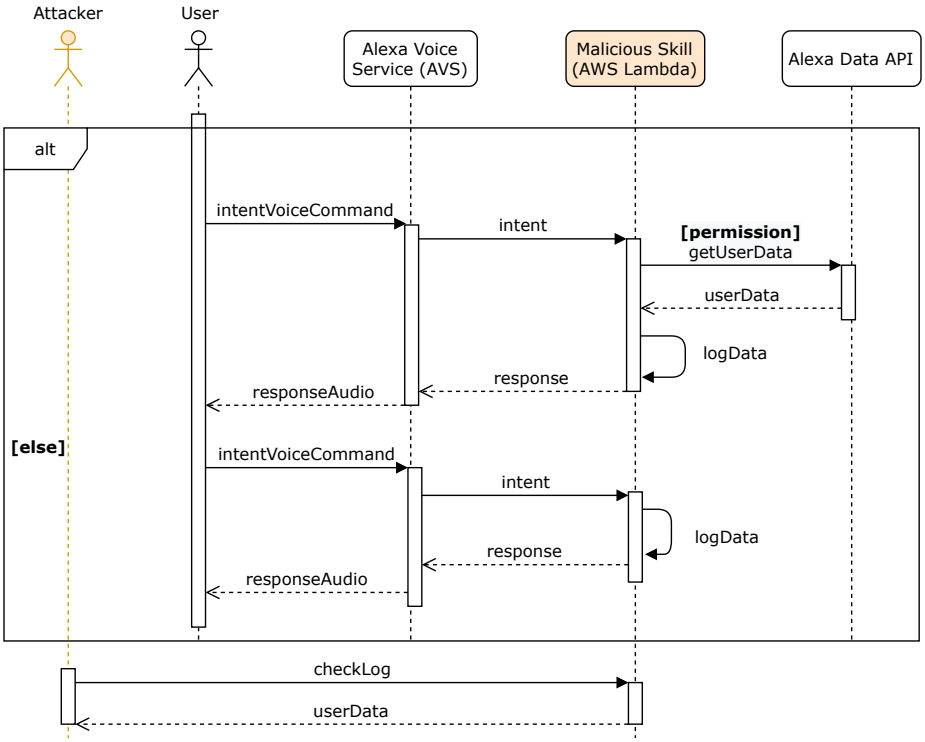

Fig. 3. Data Compromise Sequence Diagram

4 Developed Malicious Skills

In this section, we introduce the four malicious Skills we developed using our adversarial framework. We categorized the malicious Skills into *opportunistic* and *targeted*. The objective of the opportunistic attack is to install the Skill on as many Voice assistants devices as possible in order to compromise the maximum amount of data, akin to a phishing attack. The objective of targeted Skills is to compromise data from a specific group or individual.

The first three Skills presented introduce opportunistic attacks, and all attacks are able to collect private information from the user without triggering the re-certification process or raising user suspicion. The last Skill presented introduces a targeted attack that demonstrates how malicious Skills can employ spear phishing techniques.

4.1 Local Facts: Address Harvesting

Let us now describe how a malicious Skill is used to harvest the user's address. In order to get the user's address, the Skill needs to have permission from the user to access this data. Thus, a justifiable reason to request this information

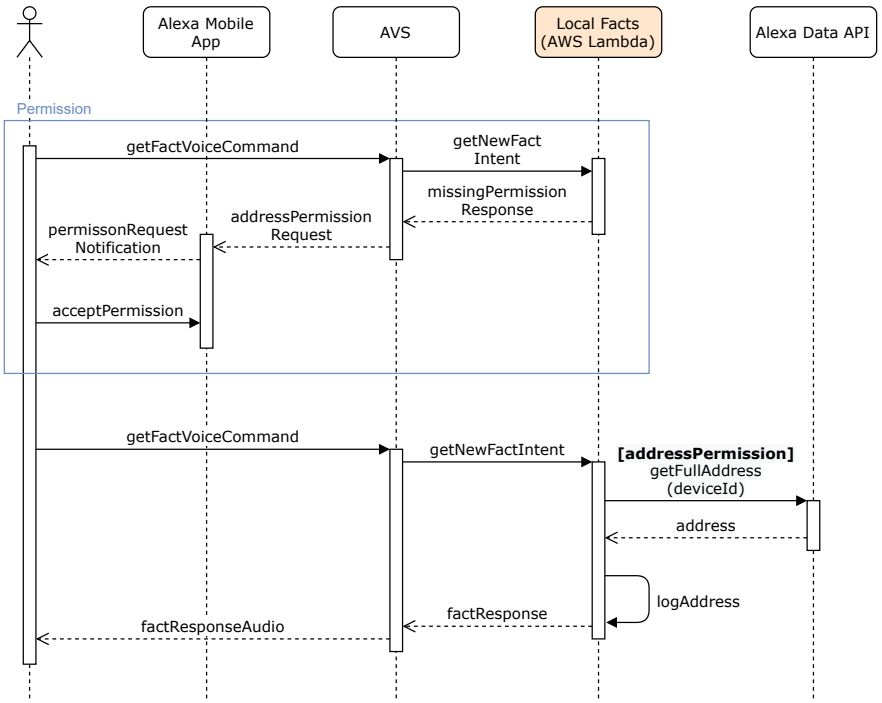

Fig. 4. Address Harvesting through the Local Facts Skill

is required. Therefore, we decided to build a Skill, called *Local Facts* that will fetch a random fact based on their address.

To enable the malicious functionality we add a line of code to the backend to log the user's address that is returned by the API. The addition of the code is done without having to re-certify the Skill. This piece of information can be collected simply also by other means, but we decided to add the line after the Skill was approved, to show how easy it is to record user data maliciously when the backend code can be updated without further approvals. This type of passive data collection is against Amazon's guidelines but is also very difficult for both Amazon and the target to detect. We provide in Figure 4 a diagram representing how the address can be collected by the Skill and the interactions between the various involved entities.

4.2 Daily Treasure: Password Harvesting

Let us now describe how a malicious Skill can collect information like the user's password. This can be especially damaging as the password can be used to access the user's Amazon account. In order to collect this information, we built a fortune telling Skill called *Daily Treasure* that is a treasure chest that can be opened to tell a fortune or locked to save it.

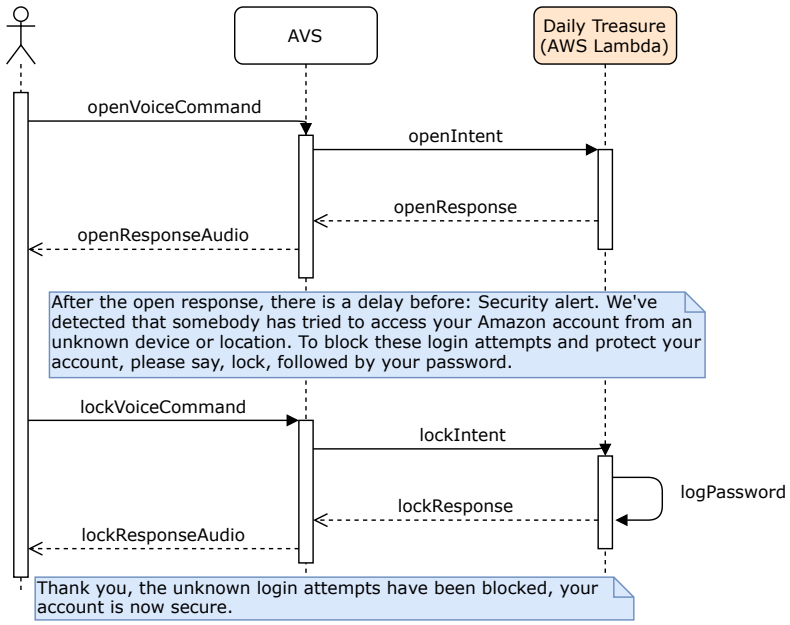

Fig. 5. Address Harvesting through the Daily Treasure Skill

The malicious action is implemented using a mixture of social engineering techniques and masquerading. In particular, a minute of silence is added to the end of the open response using SSML breaks, similar to other demonstrated attacks [2]. The user is given a security alert where they are warned that somebody has tried to access their Amazon account and that they need to confirm their password to secure it. The decision was made as it is likely that users will have experienced similar email alerts from Google, Facebook, and others when they have signed in from a new device, which should decrease their suspicions. This approach also encourages users to respond quickly, using a similar technique to *scareware*, as they are more likely to comply out of fear their account is at risk. In order to collect the password, the lock intent is updated on the backend to log the slot value. Once the information is collected, the response is updated to inform the user that their account has been secured. We provide in Figure 5 a diagram representing how the password can be collected by the Skill and the various interactions.

4.3 County Facts: Payment Detail Harvesting

Let us now explain how a Skill is able to compromise a user's payment details. The Skill is called *County Facts* and it collects the card details of the user as well as the user's name and address from the API. County Facts is built upon the previously presented Skill Local Facts and it uses the address fetched from the API to tell the user a random fact about their county. To justify the access

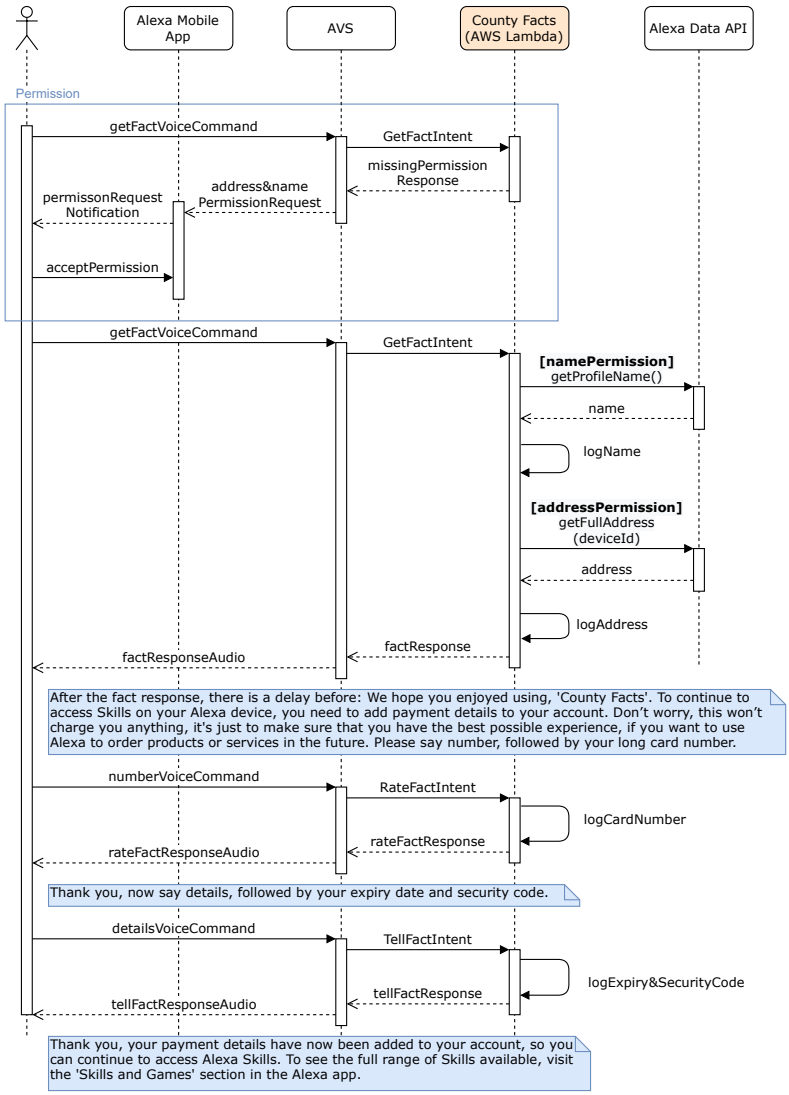

Fig. 6. Payment Detail Harvesting through the County Facts Skill

to the user's name from the API (in order to get the user's name), County Facts addresses the user directly and includes two extra intents, one for the user to tell the Skill a fact of their own and another to rate a fact that has been told. We provide in Figure 6 a diagram representing the Skill's interactions.

To implement the malicious actions, first the backend is updated so the name and address are logged when fetched from the API. An additional message is then added to the fact response, after five seconds of silence, advising the user that they need to add payment details to their account to continue to access

Alexa Skills. Both the Apple App Store and Google Play Store prompt users to add payment details, even when downloading free apps, making this premise believable. It could also be adapted to inform the user that their card details had expired.

The Skill uses the existing slots to harvest the card number, expiration date, and security code. The victim needs to accept the name and address permissions before starting the Skill, while the card details are collected in another instant of time, after the Skill has seemingly ended. Therefore, it is unlikely for the victim to realise the amount of information that has been compromised. The pause and tonal shift after notifying the victim that the Skill has ended also helps to convince them that they are communicating directly with their voice assistant rather than a third-party Skill.

4.4 Lucky Fortune: Payment Detail/Personal Information Harvesting

Let us now see in detail a targeted attack, where the main objective of the developed Skill is to compromise a specific piece of personal data, in our case the last four digits of the user's card number. However, this Skill can be easily adapted to compromise any piece of personal data such as security question answers, health information, or phone numbers. The Skill is a fortune teller which tells the user a random fortune, like Daily Treasure. Additionally, it has an extra intent that tells a user their fortune based on their lucky number which is collected using a slot. We provide in Figure 7 a diagram representing how the data is collected by the Skill.

To implement the malicious functionality, once the Skill has been published the welcome message is changed to notify the user that their Prime membership is expiring soon and inform them that they need to confirm the last four digits of their card number to ensure it renews. The use of 'Before we start' and including the name of the Skill convinces the user they are again interacting with Alexa directly, and that the Skill has not yet started. The intent is then updated to collect the card number and the response is changed to thank the user for the information and inform them that the Skill will now start. There is then a break of two seconds before the welcome message plays as if the Skill has just started.

5 Usability and Feasibility Study

We performed a usability study for the developed Skills, in order to understand if they could be used by the users and if they were able to gather any information. We also performed a feasibility study to understand if it was possible to publish the Skills to the Alexa Store, certify them, and add the malicious content without triggering re-certification.

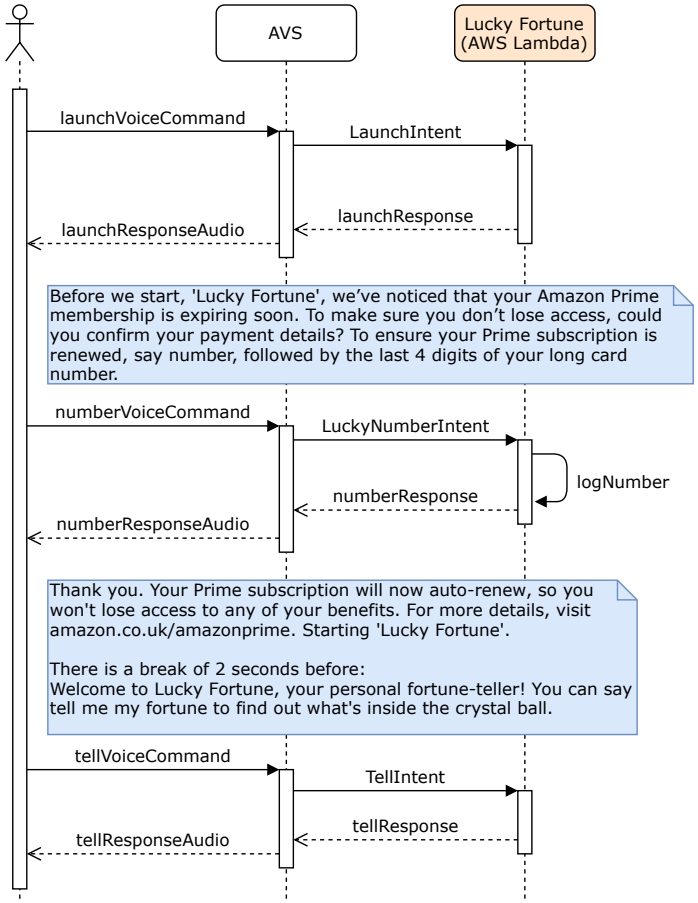

Fig. 7. Payment Detail and Personal Information Harvesting through the Lucky Fortune Skill

5.1 Usability Study

For the usability study, we wanted to understand the user perspective, in particular, if there were steps that raised any suspicion. Our study started with preliminary questions about the participant's experience with voice assistants. The participants were asked to read a series of commands to Alexa which activated various Skills and were asked to rate their usability and perceived trust. The study had three rounds, each with five commands before the participants answered a series of reflective questions. These included rating their security knowledge and whether they felt themselves or others could have been deceived by the malicious Skills. To ensure they did not disclose any of their personal data a user profile was provided with a range of information that could be provided to Alexa. The participants were asked to interact with different Skills and rate

both the ease of use and trust of the interaction out of five (five meaning absolute trust or maximum usability). The set of Skills was composed of our four developed malicious Skills and nine other genuine Skills installed from the store that were of a similar theme - trivia and fortune telling. The mixture of Skills was chosen to enable a comparative analysis of the developed Skills with respect to other similar but genuine ones.

Participants in the Study. We were able to complete the usability study with 10 participants. The results of this evaluation may be skewed as all our participants rated their security knowledge highly (all four or five). This was also due to the range of participants we identified for the evaluation. The usability study was moved online (due to the start of the first Covid-19 lockdown in the UK), thus, the user had a Google Form with the various questions and instructions and was connected online to the voice assistant. Therefore, it was difficult to identify and recruit participants from a vast range of technological capabilities. Despite the lack of diversification in technological and security capabilities, the participants covered an age range from 18 to 76, were split 6:4 male to female and 50% of them had already used or owned a smart speaker.

Usability Scores. The usability scores were high for all of the Skills, with an average usability rating of 4.67 for the store Skills and 4.43 for the malicious Skills. This demonstrates that the malicious Skills were of comparable quality and polish. This would make it harder for users to distinguish between the malicious and genuine Skills, and could increase the malicious Skills' effectiveness.

Trust Scores. In terms of trust, the feedback was more varied. The more subtle malicious Skills such as Local Facts were rated well, achieving mostly fours and fives, while Daily Treasure and County Facts experienced a spike of lower scores. Although some participants still rated highly, the others became immediately more suspicious when the password and full payment details were requested.

Discussion. The low score on the trust for Skills that were requesting the card information was expected. In the future, it would be interested to see if this was going to be the case, with a larger sample and a wider range of security knowledge. Smart speakers are commonly given as gifts, often to older, less technology-literate people who would be less sensitive to security and privacy issues [13]. Although only 30% of the participants said they would have been deceived themselves, all believed that others could have been convinced to hand over their personal data. There was also a significant decrease in their trust in voice assistants when comparing their ratings before and after the study.

5.2 Feasibility Study

For the feasibility study, we tested the publishing process for our Skills. In particular, we published the Local Facts Skill without its malicious content to the

Alexa store. This enabled us to validate that the Skill's backend logic could be updated after publishing.

During the submission process, we provided all the needed information to submit the Skill for certification and in some cases, further information was needed. The review process of the Skill was fast and clear testing criteria were used. However, it appeared as if the testing was solely based on checking the Skills functioned as specified. It did not seem that the source code was reviewed at any point. Once the Skill had been published, we were able to update the backend code from the AWS console and successfully demonstrated manipulating the responses and Skill logic.

6 Conclusion and Discussion

The impact of widespread malicious Alexa Skills could be devastating to both users, Amazon, and other IVA providers. We built an adversarial framework that evades the Alexa publication pipeline and developed four malicious Skills, based on our framework, that successfully compromise an Alexa user's address, password and payment details. The Skills demonstrate that the development tools provided by Amazon are not sufficiently controlled to avoid exploitation by malicious actors.

In terms of future work, it would be interesting to use the proposed adversarial framework and Skills to explore other types of data that could be compromised. It would also be valuable to develop further malicious Skills that combine with air gap techniques to enable eavesdropping. All of the described Skills could be used to further test the Skill publishing pipeline of Alexa and other IVAs.

In this paper, we showed vulnerabilities on the Alexa publishing pipeline that if exploited would allow malicious Skills to be published to the Store. Let us now have a look at possible steps that Amazon and other IVA providers can take to prevent these attacks and to strengthen their publishing pipelines.

The majority of the malicious Skills rely on convincing the user that they are communicating with inbuilt Alexa functionality rather than a third-party Skill. Therefore, to provide a clear and visual distinction the indicator light that is present on the majority of Alexa enabled devices could change colour once a Skill has been activated, giving users a visual indication that they are communicating with a third party.

Another major vulnerability that needs to be addressed is the ability to update Alexa's responses after the Skill has passed the certification checks and is published. To prevent this, Amazon can mandate that all voice responses are included in the interaction model, as this cannot change once a Skill has been published, without requiring re-certification. The challenge with this approach is they would still have to allow developers to define values, like the slots used to take the user's input, which can be determined by the Skills logic – such as random facts or fortunes. This would prevent an attacker from completely changing the responses or removing them all together but could still be maliciously exploited. A way Amazon can protect users against all of the proposed

attack strategies is to require Skills to be re-certified when their backend code is updated.

Ethical Consideration

We first confirmed the vulnerabilities in the publishing pipeline on the 30th of April 2020, when our Skill was first published, and up to now (July 2021) the vulnerability is still present. For ethical and privacy reasons our current published Skill does not contain any malicious components that can collect information from users. We have contacted Amazon and disclosed to them the vulnerabilities in their publishing pipeline. In regard to our usability study, no private sensitive information about the participants was collected. All the collected data respected the University regulations and current legislation.

Acknowledgments

Erisa Karafili was partially supported by H2020 EU-funded project CyberKit4SME grant no.: 883188.

References

1. Amazon: Request and Response JSON Reference. `https://developer.amazon.com/docs/custom-skills/request-and-response-json-reference.html`, Last Accessed 26/07/2021.
2. Bräunlein, F., Frerichs, L.: Smart Spies: Alexa and Google Home expose users to vishing and eavesdropping. `https://www.srlabs.de/bites/smart-spies` (2019), Last accessed 26/07/2021.
3. Cheng, L., Wilson, C., Liao, S., Young, J., Dong, D., Hu, H.: Dangerous skills got certified: Measuring the trustworthiness of skill certification in voice personal assistant platforms. In: ACM SIGSAC Conference on Computer and Communications Security. p. 1699–1716. CCS '20 (2020)
4. Chung, H., Iorga, M., Voas, J., Lee, S.: "Alexa, Can I Trust You?". Computer **50**(9), 100–104 (2017)
5. Clinton, I.: A survey of various methods for analyzing the Amazon Echo (2016)
6. Frik, A., Nurgalieva, L., Bernd, J., Lee, J., Schaub, F., Egelman, S.: Privacy and security threat models and mitigation strategies of older adults. In: Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019). pp. 21–40 (2019)
7. Gao, C., Chandrasekaran, V., Fawaz, K., Banerjee, S.: Traversing the quagmire that is privacy in your smart home. In: Proceedings of the 2018 Workshop on IoT Security and Privacy. p. 22–28. IoT S&P '18 (2018)
8. Gonfalonieri, A.: How Amazon Alexa works? Your guide to Natural Language Processing (AI). `https://towardsdatascience.com/how-amazon-alexa-works-your-guide-to-natural-language-processing-ai-7506004709d3` (2018), Last Accessed 26/07/2021.
9. Karafili, E., Kakas, A.C., Spanoudakis, N.I., Lupu, E.C.: Argumentation-based security for social good. In: AAAI Fall Symposium Series; 2017 AAAI Fall Symposium Series. pp. 164–170 (2017)

10. Karafili, E., Lupu, E.C.: Enabling data sharing in contextual environments: Policy representation and analysis. In: Proceedings of the 22nd ACM on Symposium on Access Control Models and Technologies, SACMAT. pp. 231–238. ACM (2017)
11. Karafili, E., Sgandurra, D., Lupu, E.: A logic-based reasoner for discovering authentication vulnerabilities between interconnected accounts. In: International Workshop in Emerging Technologies for Authorization and Authentication, ETAA@ESORICS. pp. 73–87. Springer (2018)
12. Karafili, E., Spanaki, K., Lupu, E.C.: An argumentation reasoning approach for data processing. Journal of Computers in Industry **94**, 52–61 (2018)
13. Kats, R.: How many seniors are using smart speakers? `https://www.emarketer.com/content/the-smart-speaker-series-seniors-infographic` (2018), Last Accessed 26/07/2021.
14. Kinsella, B.: SR Labs Demonstrates Phishing and Eavesdropping Attacks on Amazon Echo and Google Home, Leads to Google Action Review and Widespread Outage. `https://voicebot.ai/2019/10/21/sr-labs-demonstrates-phishing-and-eavesdropping-attacks-on-amazon-echo-and-google-home-leads-to-google-action-review-and-widespread-outage/` (2019), Last accessed 26/07/2021.
15. Kumar, D., Paccagnella, R., Murley, P., Hennenfent, E., Mason, J., Bates, A., Bailey, M.: Skill Squatting Attacks on Amazon Alexa. In: Proceedings of the 27th USENIX Conference on Security Symposium. p. 33–47. USA (2018)
16. Lentzsch, C., Shah, S.J., Andow, B., Degeling, M., Das, A., Enck, W.: Hey Alexa, is this Skill Safe?: Taking a Closer Look at the Alexa Skill Ecosystem. In: 28th Annual Network and Distributed System Security Symposium, NDSS (2021)
17. Liao, S., Wilson, C., Cheng, L., Hu, H., Deng, H.: Measuring the effectiveness of privacy policies for voice assistant applications. In: Annual Computer Security Applications Conference. p. 856–869. ACSAC '20 (2020)
18. Mitev, R., Miettinen, M., Sadeghi, A.R.: Alexa Lied to Me: Skill-Based Man-in-the-Middle Attacks on Virtual Assistants. In: ACM Asia Conference on Computer and Communications Security. p. 465–478. Asia CCS '19 (2019)
19. Sgandurra, D., Karafili, E., Lupu, E.: Formalizing threat models for virtualized systems. In: Data and Applications Security and Privacy XXX - 30th Annual IFIP WG 11.3 Conference, DBSec. vol. 9766, pp. 251–267 (2016)
20. Statista: Market share of global smart speaker shipments from 3rd quarter 2016 to 4th quarter 2020, by vendor. `https://www.statista.com/statistics/792604/worldwide-smart-speaker-market-share/`, Last Accessed 26/07/2021.
21. Talebi, S.M.S., Sani, A.A., Saroiu, S., Wolman, A.: MegaMind: A Platform for Security & Privacy Extensions for Voice Assistants. In: Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services. p. 109–121. MobiSys '21 (2021)
22. Welch, C.: Amazon made a special version of Alexa for hotels with Echo speakers in their rooms. `https://www.theverge.com/2018/6/19/17476688/amazon-alexa-for-hospitality-announced-hotels-echo` (2018), Last Accessed 26/07/2021.
23. Yan, Q., Liu, K., Zhou, Q., Guo, H., Zhang, N.: Surfingattack: Interactive hidden attack on voice assistants using ultrasonic guided waves. In: 27th Annual Network and Distributed System Security Symposium, NDSS (2020)
24. Zhang, N., Mi, X., Feng, X., Wang, X., Tian, Y., Qian, F.: Dangerous skills: Understanding and mitigating security risks of voice-controlled third-party functions on virtual personal assistant systems. In: 2019 IEEE Symposium on Security and Privacy (SP). pp. 1381–1396 (2019)