Analysis of security and data control in smart personal assistants from the user's perspective

Cayetano Valero[a,*], Jaime Pérez[a], Sonia Solera-Cotanilla[b], Mario Vega-Barbas[b], Guillermo Suarez-Tangil[c], Manuel Alvarez-Campana[b] and Gregorio López[a,*]

[a]Institute for Research in Technology, ICAI Engineering School, Universidad Pontificia Comillas, Madrid, 28015, Spain
[b]ETSI Telecomunicación, Universidad Politécnica de Madrid, Madrid, 28040, Spain
[c]IMDEA Networks Institute, Leganés, 28918, Spain

ARTICLE INFO

Keywords:
Cybersecurity
Data control
Internet of Things
Minors
Smart Personal Assistants
Testing methodology

ABSTRACT

Advances in the fields of the Internet of Things, Speech Recognition and Artificial Intelligence have facilitated the development of Smart Personal Assistants. As a result, Smart Personal Assistants currently allow requesting a wide range of tasks naturally and intuitively through voice interaction. Their wide popularity, together with the high technological complexity of their environments, have made them an attractive target from a security point of view. Recent works have shown some of the security and privacy issues they stand upon. In this work, we propose a methodology to carry out a systematic security analysis of Smart Personal Assistants using a comprehensive set of tests designed to measure issues around the installation, the interaction, key functionality, and overall Security and Privacy controls. We apply this methodology to analyze security and data control in predominant commercial Smart Personal Assistants (SPA), including Apple HomePod, Google Home and Nest, Amazon Echo (Show and Dot), and Facebook Portal. The main findings of our research are: (i) SPA are not resilient to voice replay attacks; (ii) their skills activation mechanisms can be significantly improved to be more reliable in multi-user households; (iii) the users' control to restrict the collection and access of Personally Identifiable Information can be also improved; (iv) they lack configurations adapted to minors, which should be included to make them more appropriate for a segment of users who interact more and more with them and have especially high regulatory requirements regarding security and data protection. Among the many hot research topics within this area, we find voice authentication and authorization especially interesting since they may push the usability of Smart Personal Assistants further, as long as they are robust enough from the security perspective.

1. Introduction

One of the prominent implicit promises of new technologies, especially the Internet of Things (IoT), has always been convenience and natural integration into our environments. For instance, ask the device to perform a task and spend minimal effort on this request. Advances in Artificial Intelligence (AI) in recent years have allowed us to communicate with our devices in a way that is quite easy and natural to us: using our voice. This model of human-computer interaction is prevalent today, and users find voice command communication much more intuitive than using keyboards and remote control devices Ruan, Wobbrock, Liou, Ng and Landay (2018).

One of the technologies that has significantly contributed to the improvement of voice interaction systems through its penetration in households is the Smart Personal Assistant (SPA) Statista (2020a). The SPA connects home-integrated devices and remote services on the Internet, allowing the user to perform numerous tasks and a simple centralized orchestration of available devices. The user can, for example, query information such as the weather or the traffic status, listen to music, make voice and video calls, do online shopping, or control other devices such as smart lights and intelligent thermostats Amazon (2021b), Statista (2020b). SPA differentiates from traditional voice interaction systems in the way they process the commands collected from the user. While in a traditional system the user needs to invoke an exact command with appropriate pronunciation and a rigid structure to activate a function, SPA takes advantage of improvements in natural language processing to "understand" the user's voice commands, extracting the intent of their messages to provide the best answer to their question.

However, the high penetration in households and offices, the vast orchestration of functionalities, and the access to numerous services offered by SPA present a number of challenges to security and privacy. According to Canalys, there are more than 200 million SPA installed worldwide, and if this trend continues, the number will exceed 500 million units by 2030 Canalys (2019). It is estimated that in countries such as the US and the UK, one-third of households already have a SPA installed Lau, Zimmerman and Schaub (2018). The growth of Smart Home technologies leads to an increasing number of traditional devices in the home having a "smart" version, enabling remote control, automation, and scheduling of the devices Liu (2021). These facts introduce a key understudied factor in cybersecurity: many of these households have minors that operate a SPA. Due to the

*Corresponding author
cayetanova@alu.comillas.edu (C. Valero); gllopez@comillas.edu (G. López)
ORCID(s): 0000-0002-7981-5311 (C. Valero); 0000-0001-6044-0022 (J. Pérez); 0000-0003-3516-4489 (S. Solera-Cotanilla); 0000-0003-4506-6284 (M. Vega-Barbas); 0000-0002-0455-2553 (G. Suarez-Tangil); 0000-0003-2747-9798 (M. Alvarez-Campana); 0000-0002-7981-5311 (G. López)

pervasiveness of connected devices, such as the SPA, many fraudsters and abusers (e.g., in the context of intimate partner violence) have strategic accessibility to potential underage victims.

There are growing concerns about the privacy guarantees offered by these devices. For instance, corporations in the data analytics business may pose a threat to the general public as they attempt to collect personal data over *always-on, always listening* devices to offer targeted marketing Abdi, Ramokapane and Such (2019). According to the General Data Protection Regulation (GDPR) in force in the European Union, voice data is considered sensitive and requires the user's explicit consent for its further exploitation and the incorporation of privacy by design approach Commission (2018), which is not the case for commercial devices Abdi, Zhan, Ramokapane and Such (2021).

1.1. Motivation

In this dynamic and complex context, where new technologies are emerging at an unprecedented pace and SPAs are becoming a fundamental part in people's interaction with technology, as they are being deeply integrated with other electronic devices, it is essential to have an awareness of privacy and security issues associated with SPAs, identifying which tools and data are at stake, as this is the key to determining and correctly assessing the possible security risks to which users are exposed.

1.2. Contributions

In this paper, we propose a methodology to carry out a security analysis of the threats that SPAs technology may pose to their users. We perform our analysis in a uniform and reproducible manner and we apply such a methodology to analyse the security and data control of predominant commercial SPAs in multi-user environments. We carry our analysis in late 2021 and we obtain the following findings:

- Main commercial SPAs are not resilient to voice replay attacks. Voice authentication mechanisms should be improved, including continuous authentication.
- Main commercial SPAs do not offer reliable enough control systems to activate skills in multi-user environments.
- Main commercial SPAs fail to offer enough control to users to restrict the collection and access of Personally Identifiable Information (PII).
- Main commercial SPAs lack basic mechanisms to control the exposure of potentially inappropriate third-party voice apps to minors. Current SPAs delegate this responsibility to their guardians without offering an effective mechanism to aid them.

With this research, we pursue two main objectives: (i) to increase transparency and user awareness around how SPA works from the security and data control perspectives; and (ii) to encourage manufacturers towards improving the security and data control of these devices, particularly enhancing the protection of the most vulnerable groups, like children.

Figure 1: Diagram of a user interacting with the SPA ecosystem. Icons created by *Smash icons* openly available in *Flaticon*.

1.3. Organization

The remainder of the paper is organized as follows. Section 1 provides an overview of the SPA ecosystem, with particular emphasis on security. Section 3 describes the proposed methodology. Section 4 presents the results of applying such a methodology to the most predominant commercial SPAs. Finally, section 5 discusses recommendations and future research and draws conclusions.

2. Related work

2.1. SPA ecosystem

The SPA contains hardware and software components to record, process, and analyse sound, although it requires a connected ecosystem in order to perform its functions properly. Figure 1 shows a simplified diagram of a user interacting with the SPA ecosystem, which consists of the following core interaction components:

- Personal Assistant: Sometimes they are also termed "Virtual Assistant". Various families of electronic devices have support for personal assistants, such as speakers, smartphones, PCs, and wearables. Their functions are to collect the user's utterances to send them to the SPA service provider's cloud, and to play the responses that arrive from the cloud to the user.
- SPA Cloud Service Provider: This component contains the "intelligence" of the assistant. It is the point where the requests are processed to extract the user's intention and find the answer that best suits his question.
- Third-party and Accessory Clouds: Users can interact with external elements outside the SPA provider, e.g., with third-party *skills* or smart devices from other manufacturers. To reach these external elements, the SPA is connected to the cloud of the providers of these elements.

Due to their usability and orchestration capabilities, the SPAs have positioned themselves as the central elements of the Smart Home environment, acting as the hub between the user and the rest of the connected devices. A growing number of smart device manufacturers are integrating their products into the SPA ecosystems to offer a better user experience and a smoother configuration process to their customers Statista

(2021), Statista (2020c), Statista (2019a). As the unifying point of communication with smart home devices, the SPA becomes a high-value target for a potential malicious actor, since gaining access would allow interacting with the other smart devices and accessing the information associated with the SPA.

The information collected and processed in the SPA is one of the most important assets from the users' point of view, due to the nature of the data itself and the apparent ease of access to the device. Personal information that a SPA processes includes user account data, purchase information, service interaction data, and the user's own voice recordings Edu, Such and Suarez-Tangil (2019) Alexa (2020) Mozilla Foundation (2020).

2.2. Security in SPA

As part of the IoT ecosystem, SPAs present security challenges intrinsic to this environment. Taking into account the architecture presented by the authors in **?**, the threats inherent to IoT devices can be classified into three layers: Perception Layer, associated with the physical layer of the elements, Network Layer, that provides communication between devices and Application Layer, that encapsulates the functionalities provided to the end users. Due to their unique architecture and operation, SPAs specific security issues are more related to the Perception and Application Layer. These security problems have been studied in different studies, to check how they can be used by a malicious actor and what impact they might cause on users. Mainly, as described in the subsequent paragraphs, the security issues in the SPAs are due to the open nature of the communication channel between the user and assistant, the use of third-party applications, and the use of AI to transcribe voice commands and provide authorization to the users.

Cheng and Roedig (2022) proposes a standardised taxonomy of security and privacy issues in voice personal assistants, consisting of four categories: Access Control, Acoustic DoS, Voice Privacy, and Acoustic Sensing. In this paper, we focus especially on the category of Access Control, where the most worrying security issues of the SPAs are concentrated. Within this category, we identified the four subcategories with the highest relevance: Weak Activation, Weak Authorization, Adversarial AI Attacks, and Use of Third-party Skills.

Weak Activation The functionality of SPA is based on a permanent listening mode. The assistant is listening at all times, waiting for the user to activate it and invoke a specific function. The activation is done by a keyword that the user transmits to the SPA, which will cause it to enter the active operation mode, during which the assistant records the user's voice to process it and generate a response. Therefore, the only authentication method available in an SPA is the word the user uses to activate it, which is chosen from a predefined and limited set Amazon (2021a), making it easy for an attacker to guess it and be able to interact with the assistant as a legitimate user. With the assistant always on, the time window in which a malicious actor could attack the SPA is increased,

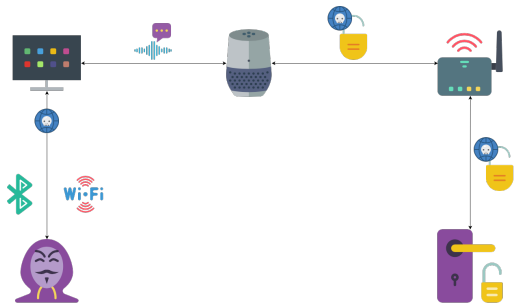

Figure 2: Example of a remote attack to a SPA

even being able to use electronic devices near the device on which the SPA is installed to inject commands without being physically present and perform actions such as unlocking the doors of the house (Figure 2) or making purchases on behalf of the user Yuan, Chen, Wang, Chen, Zhang, Huang and Molloy (2018).

Weak Authorization Some commercial SPAs have user distinction features to adapt their response based on the user invoking the action. User authentication in the SPA is based on the voice utterance collected when the user interacts with the assistant by using the corresponding wake-up word. Some of the most widely used, such as *Alexa* or *Google Assistant* Statista (2019b) include this functionality, although it is not officially offered as a security feature due to the number of false positives. Manufacturers warn of the fact that a user with a similar voice to another can access user's personal information or impersonate him/her Google (2021), so it is offered as an answer customization functionality. Depending on the settings a user has made in the assistant regarding access to personal data and other functionalities such as payment management, misidentifying a user would mean that a person could access the user's personal information through the SPA, being able to send emails or view events in their calendar or even make payments on his/her behalf.

Adversarial AI Attacks The use of AI in the SPA entails a number of risks inherent to this technology, such as the use of the AI itself to attack the speech recognition models Carlini and Wagner (2018). The environments in which the models will be used are assumed to be secure, but due to the open nature of the communication channel between the user and the SPA, a possibility exists that a malicious actor could inject samples specially crafted to cause potentially unwanted behavior in the SPA, being able to force malicious actions, such as downloading *malware* on the device where the SPA is installed by accessing a URL, or causing a denial of service Li, Qu, Li, Szurley, Kolter and Metze (2019).

Use of third-party skills Third-party *skills* extend the functionality of a SPA, allowing users to include third-party applications in their assistants so that they can be invoked just as they would with a native assistant function. However, this increases the risk to which a user is exposed, as the attack surface against a malicious actor increases, with an attacker

being able to create malicious *skills* with a pronunciation similar to a legitimate one to hijack user requests directed to the legitimate *skill*, or being able to create *skills* that take control of services that are normally reserved for the SPA, such as starting and terminating interactions with the *skills*, to make the user believe that a *skill* has been successfully executed and the SPA has stopped listening, while in reality, the malicious *skill* is still active, recording and sending information to the attacker without the user being aware Zhang, Mi, Feng, Wang, Tian and Qian (2019).

3. Proposed methodology

3.1. Overview

Figure 3 shows an overview of the methodology proposed in this paper to analyse, in a reproducible manner, security and data control in commercial SPA from the user perspective in multi-user environments.

First, we need to determine how the user interacts with the SPA. The objective of this first stage is to develop different categories that fully characterize the relation with a SPA from the user perspective, from the moment of installation to the configuration of the different features. In this work, four categories are proposed to cover this relation, namely: installation, interaction, functionality, and security and data control.

Once the categories are established, we proceed with the definition of the different tests in each of them, in order to identify and evaluate the processes that a user conducts in each of the phases. Finally, it is necessary to define a common testing procedure to ensure that the tests are performed under equivalent conditions on all the devices to preserve the reproducibility of the tests.

The following sections develop in detail each of the defined categories and the common testing process.

3.2. Categories and tests

The feature tests were divided into four categories, covering the different stages of the interaction of a user with the SPA and the possible security and data control adjustments that can be made. With this, we aim to obtain a global overview of security and data control in SPAs, focusing the study on common configuration aspects, and actions that a user can perform on the SPA.

3.2.1. Installation

This category covers both the installation of the device with the SPA itself and the registration of new accessories and third-party skills.

- *Installation process of the SPA*. This test studies how the installation of the SPA is carried out, which additional devices and applications are necessary, and what configurations can be performed.
- Installation of new connected devices. The installation process of accessories connected to the SPA is examined together with the permissions or configurations that can be set to restrict their use.

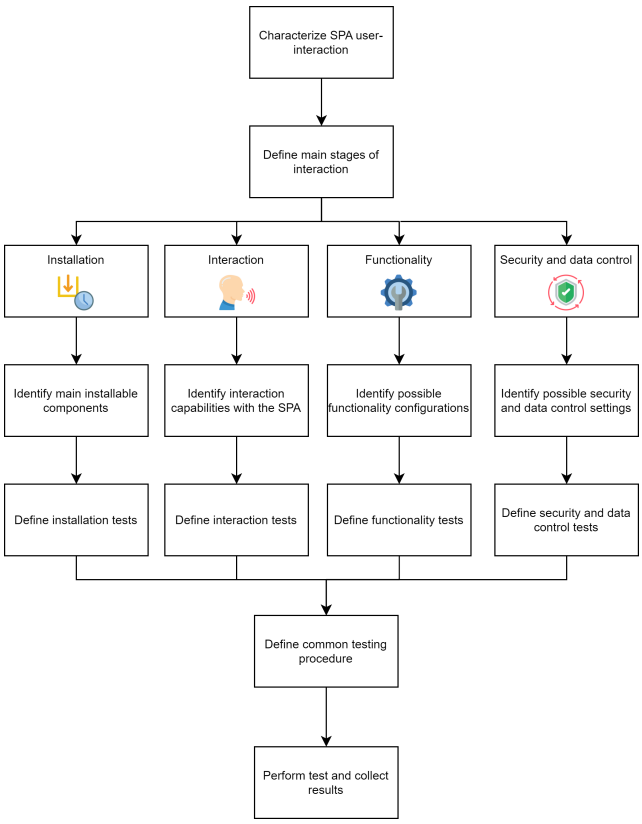

Figure 3: Overview of the proposed methodology

- *Installation of third-party skills*. The installation process of third-party skills or developments to be invoked from the SPA are analysed if the SPA has this functionality available.

3.2.2. Interaction

This category covers the possible options that a user has for managing the interaction with the SPA, with connected devices and third-party skills.

- *Interaction with the SPA. and connected devices* This test checks how a user can control the interaction with the SPA and the connected accessories.
- *Interaction with third-party skills*. It is analysed if there are controls that allow defining user profiles for third-party skills incorporated into the SPA.

3.2.3. Functionality

This category covers options to control assistant functionalities, such as payments or multimedia content playback.

- *Payments and transactions*. The possibility of making payments from the SPA is studied, verifying the configurations and restrictions that a user can define.
- *Default profiles for minors*. It is checked if the SPA has options to create restricted profiles for minors in an easy manner, therefore allowing for control of multimedia

content that is played by the SPA, the accessories with which it can interact, payments that can be made, etc.

3.2.4. Privacy and data control

This category includes tests that evaluate the security features of the SPA, as well as the options that a person has to control the use of his/her personal information.

- *Control of answers containing personal information*. SPA can include users' personal information as responses to some of the requests. This test evaluates if a user can control the sharing options of his personal information.
- *Authentication methods*. The authentication methods available in the SPA are analysed, as well as their real effectiveness.
- *Non-human voice filtering*. It is checked whether the SPA is susceptible to be activated by voices of artificial origin, such as a recorded message or a Text To Speech (TTS) system.
- *Interaction with the conversation history*. These tests cover the options provided to the user to control and display information about the conversation history with the assistant.

3.3. Testing procedure

Each SPA model follows a different operating procedure but can be generalized. In this way, it is possible to systematize the evidence acquisition process during the execution of each test defined for each device. Thus, it is possible to obtain a uniform set of results and avoid excluding relevant data and evidence. In this work, the following common testing procedure is applied:

- Switching on the SPA and the mobile device.
- Connecting the SPA to the Internet via Wi-Fi.
- Registration/login to the application that manages the SPA. Personal account required (*Google*, *Amazon*, *iCloud*, *Facebook*, etc.).
- SPA recognition, synchronization and configuration process from the mobile application.
- Tests of interaction with the SPA:
  - Basic tests: voice authentication, recognition of non-human voices, exchange of personal data, storage of conversation histories, control of responses with personal content, etc.
  - Tests with devices connected to the SPA.
  - Testing with the SPA and third party *skills*, if any.
- Testing of voice payments and transactions.

3.4. Device selection

On a global scale, and with the exception of personal assistants only available in the Chinese market, the most widely used personal assistants at the time the research was carried out were *Amazon*'s *Alexa*, with an estimated market share of 31.7 % (2019), *Google*'s *Google Assistant* with an approximate market share of 31.4 % (2019), and *Apple*'s *Siri*, with a market share of 6 % (2019) Statista (2019b). Although Siri's market share is lower than direct competitors, it is the assistant that is installed on the largest number of devices in the world, due to the popularity of Apple's smartphones and tablets Statista (2020a).

The most popular SPA, from the four leading brands, were included in this research. The main features of the selected SPAs are summarized and compared in Table 1. In addition, for the sake of reproducibility, Table 2 shows the models, devices, and applications used in this research. This information is especially relevant since we can only claim that the obtained results hold for this combination of hardware and software.

4. Analysis of test results

The results obtained in the analysis are summarized in the Tables 3 and 4, which show the findings for the tests on each of the aforementioned considered devices: *Apple HomePod Mini*, *Google Home Mini*, *Google Nest Audio*, *Amazon Echo Show 5*, *Amazon Echo Dot 4* and *Facebook Portal*. Table 3 shows the tests conducted regarding the installation and interaction with the SPA; whereas Table 4 shows the results of the functionality and security tests on the SPA. As mentioned, these are the results of the tests carried out during 2021 with the combination of hardware and software presented in Table 2.

The next subsections analyzed the obtained results in detail.

4.1. SPA installation

The installation process of the SPA is similar in all devices, as shown in Table 3. In order to use the SPA, it is mandatory to have a provider's account (*Apple*, *Google*, *Amazon* or *Facebook* account), which can be created free of charge in all cases. It should be noted that all tested devices, except the *Apple HomePod Mini*, can be installed with mobile devices with different operating systems. The *Apple HomePod Mini* requires an *Apple* mobile device to be used. Unlike the previous ones, *Facebook Portal* does not require another device to perform the installation, the device itself (thanks to the touch screen) allows the user to perform this process. In fact, it facilitates this process of associating the user account with the portal by using a parity code.

4.2. Installation of new connected accessories

Control when installing new devices differs considerably between manufacturers. The *Apple HomePod Mini* approach stands out for its good trade-off between usability and security, allowing to differentiate which users can install or edit devices from the application. In the *Amazon Echo Show 5* and *Amazon*

Table 1
Comparison of the main features of the analysed SPAs

Features	*Apple HomePod Mini*	*Google Home Mini*	*Google Nest Audio*	*Amazon Echo Show 5*	*Amazon Echo Dot 4*	*Facebook Portal*
Model and release date	1st Gen. (Nov 2020)	1st Gen. (Oct 2017)	-	1st Gen. (Jun 2019)	4th Gen. (Oct 2020)	2nd Gen. (Oct 2019)
Personal Assistant	*Siri*	*Google Assistant*		*Amazon Alexa*		
Companion application	*Home* Application	*Google Home* Application		*Amazon Alexa* Application		
Supported OS	iOS	iOS and Android				
Wake-up word activation	It can be turned off, but the wake-up word cannot be changed	It cannot be deactivated, and the wake-up word cannot be modified		It cannot be deactivated, but the wake-up word can be selected from a set of three		
Microphone	It cannot be deactivated	It can be turned off				
Camera	No			Yes. It can be deactivated	No	Yes. It can be deactivated
Voice recognition	Yes					

Echo Dot 4 only the administrator user can add or edit accessories and users. However, in the *Google Home Mini* and *Google Nest Audio* when a new user is added to the home they become an administrator, being able to install, edit and delete connected devices, with no possibility of defining more restricted permissions to certain users. In the same way, anyone can enable the use of skills and plugins from the *Facebook Portal* software portal without confirmation, and even re-activate add-ons previously disabled by the main user.

4.3. Third-party *skills* installation

Evaluation of the process of installing third-party *skills* in SPA has provided results that have corroborated the importance of studies conducted on the security of this functionality Zhang et al. (2019).

With the exception of the *Apple HomePod Mini*, which does not provide third-party *skills* support, the rest of the devices analysed do allow its use. While the installation process for third-party *skills* on these devices differs in specification, they lead to the same result, an insecure *skills* installation process that does not allow control over who can add third-party developments to the SPA. In the *Google Home Mini* and *Google Nest Audio* there is no *skills* installation process, knowing the activation phrase of the *skill* it can be activated without additional controls. In the *Amazon Echo Show 5* and *Amazon Echo Dot 4* there is an installation process of *skills* from the mobile application, where it is also possible to see which *skills* are activated. But this control process loses its sense if interacting with the SPA in a direct way, since in this way it is not checked if a *skill* has been activated or deactivated by the administrator. The same holds for *Google Home Mini* and *Google Nest Audio*: it is sufficient to know the activation phrase to use the *skill*. Even if the administrator deactivated it from the mobile app, it can be reactivated from the SPA by simply invoking the *skill* again. In the case of *Facebook Portal*, it is allowed by restriction to the device software portal.

4.4. Interaction with the SPA and connected devices

Through these tests, the level of control of the interaction with the SPA that can be established has been assessed. As shown in Table 3, the devices offer different approaches. In the *Apple HomePod Mini* it is possible to enable and disable voice interaction or touch control, as well as disable the playback of media content for certain users and the use of certain devices. In the *Google Home Mini* and *Google Nest Audio* there is a physical button to disable the device's microphones. You can disable media playback on the device, but you need to create a family in an external application to incorporate the users you want to control. Such control is configured in the form of filters, which are applied by user groups (*all* or *members of the family*), and it is not possible to set individual controls. In the *Amazon Echo Show 5*, *Amazon Echo Dot 4* and *Facebook Portal* physical controls are provided to activate and deactivate camera if any, and microphones, but since there is no user registry, it is not possible to define interaction permissions for specific users.

4.5. Interaction with third-party *skills*

As discussed in Table 3, it is not possible to use third-party *skills* on the *Apple HomePod Mini*. On the *Google Home Mini*, as there are no individual user controls, it is not possible to distinguish the use of *skills* on a per-user basis. Applying filters on the device allows the use of third-party *skills* to be disabled for all users or registered household members. Controls are available in the *Amazon Echo Show 5* and *Amazon Echo Dot 4* to enable and disable *skills*, but controls have no effect if interacting directly with the SPA, as no verification is made to check which user is attempting to activate the *skill*. In the case of *Facebook Portal*, the administrator can permit the use of *skills* from the *Facebook Portal* software portal before they are used in the device, but any user can trigger an add-on without confirmation.

Table 2
Summary of the models, devices, and applications used in the study

Features	*Apple HomePod Mini*	*Google Home Mini*	*Google Nest Audio*	*Amazon Echo Show 5*	*Amazon Echo Dot 4*	*Facebook Portal*
Model information						
Model	1st Gen.	1st Gen.	-	1st Gen.	4th Gen.	2nd Gen.
Release date	November 2020	October 2017	-	June 2019	October 2020	October 2019
Personal Assistant	*Siri*	*Google Assistant*	*Google Assistant*	*Amazon Alexa*	*Amazon Alexa*	*Amazon Alexa*
General features						
Wake-up word activation	It can be turned off, but the wake-up word cannot be changed	It cannot be deactivated and the wake-up word cannot be modified	It cannot be deactivated and the wake-up word cannot be modified	It cannot be deactivated, but the wake-up word can be selected from a set of three	It cannot be deactivated, but the wake-up word can be selected from a set of three	It cannot be deactivated, but the wake-up word can be selected from a set of three
Microphone	It cannot be deactivated	It can be turned off	It can be turned off	It can be turned off	It can be turned off	It can be turned off
Camera	No	No	No	Yes. It can be deactivated	No	Yes. It can be deactivated
Linked device information						
Device	*iPhone X* (2017)	*iPhone X* (2017)	*Samsung A71*	*iPhone X* (2017)	*iPhone 12*	*iPhone 12*
Version	*iOS 14.6*	*iOS 14.6*	*Android 10*	*iOS 14.6*	*iOS 14.5*	*iOS 14.5*
Companion application information						
Application	*Home*	*Google Home*	*Google Home*	*Amazon Alexa*	*Amazon Alexa*	*Amazon Alexa*
Version	*iOS 14.6* version	2.39104	-	2.2.422194.0	-	-
Voice recognition	Yes	Yes	Yes	Yes	Yes	Yes

4.6. Payments and transactions

The possibility to make payments from the SPA is not available on the *Apple HomePod Mini*. On the *Google Home Mini* and *Google Nest Audio*, this option is disabled by default and can be configured from the mobile app. Only the users themselves can enable payments with the SPA, as they need to enter a payment method linked to their Google account, but once configured, they can be used by any user. To protect against misuse, it includes the ability to authenticate with a second factor on the mobile device to which the account has been linked, using the capabilities of the mobile device itself, such as fingerprint or facial recognition. The *Amazon Echo Show 5* and *Amazon Echo Dot 4* also offer the ability to make purchases, which requires activation of the *Amazon 1 Click* payment method from the account on the *Amazon* website. Authentication mechanisms are available, based on a four-digit code that will be required by the SPA or a voice profile of the user making the purchase. If the voice profile matches one of the users who has configured his voice model in the application, the purchase will be accepted.

4.7. Default profiles for minors

At the time of carrying out the research and in Spain, there is no possibility of creating profiles for minors with restricted use in a simple and quick way in any of the SPA. To restrict the use of features in the *Apple HomePod Mini*, it is necessary to adjust the options one by one, with cases in which it is necessary to disable a feature for all users of the device, for example, the playback of explicit multimedia content has to be disabled for all requests made to the *Apple HomePod Mini*, not being able to differentiate by users. From the mobile application, it is possible to restrict the use of connected accessories for family members, to prevent minors from interacting with elements such as doors or windows from their mobile device, but nothing prevents them from acting on devices with voice commands, for which there is

Table 3
Comparison of installation and interaction features of SPAs

Feature	*Apple HomePod Mini*	*Google Home Mini* and *Nest Audio*	*Amazon Echo Show 5* and *Amazon Echo Dot 4*	*Facebook Portal*
Installation				
SPA Installation	An *iCloud* account and an *Apple* device are required.	A *Google* account is required.	An *Amazon* account is required.	A *Facebook* account is required.
Connected accessories installation	Configurable permissions per user.	Anyone can add or edit accessories.	Only the house administrator can add or edit devices.	Only the house administrator can add or edit devices.
Third-party *skills* installation	Third-party *skills* cannot be configured.	Any action can be used without installation.	A basic user can activate any *skill* and re-activate a previously disabled *skill*.	It is allowed by restricted to the device software portal.
Interaction				
Interaction with the SPA and connected devices	Media playback and devices can be controlled per user.	It is not possible to define permissions.	There is no possibility to differentiate between users or to define permissions.	There is no possibility to differentiate between users or to define permissions.
Interaction with third-party *skills*	There is no possibility to interact with third-party *skills*.	There are no third-party action controls per user.	Any user can trigger a *skill* without confirmation.	Any user can trigger an add-on without confirmation.

Table 4
Comparison of functionality, privacy and security features in SPA

Feature	*Apple HomePod Mini*	*Google Home Mini* and *Google Nest Audio*	*Amazon Echo Show 5*, *Amazon Echo Dot 4* and *Facebook Portal*
Functionality			
Payments and transactions	Payments or purchases are not supported	Payments can be set up from the SPA. They support an additional authentication method through the linked mobile device	Payments can be made with the SPA via *Amazon 1 Click*. Additional confirmation methods can be configured via a voice profile or a four-digit code.
Default profiles for minors	There is no possibility to create a user profile for minors	This option is only available on *Android* devices. For other devices, it is required to manually create content filters that affect all the users.	There is no dedicated option to set a safe user profile for minors.
Security and data control			
Control of answers containing personal information	Voice recognition is necessary to provide personal information and sensitive information require additional authentication through smartphone.	Personal responses are linked to the voice profile of the user. They can be disabled.	There is no possibility to disable responses containing personal information. Any user can invoke them.
Authentication methods	Authentication through voice recognition and additional authentication step with the smartphone.	The device supports voice authentication, however, using a recording of a user invoking the SPA, it is possible to impersonate him (Figures 4 and 5)	Voice recognition is only used for personalization functions. A malicious actor can impersonate a legitimate user using a recording of the activation message (Figures 4 and 5)
Non-human voice filtering	Messages from recordings and synthetic voices are not filtered.	Messages from recordings and synthetic voices are not filtered.	Messages from recordings and synthetic voices are not filtered.
Interaction with conversation history	Conversation history can be deleted but is not visible	History can be viewed, paused, and automatically deleted	History can be viewed, paused, and automatically deleted

no control beyond completely disabling the activation of the SPA. On the *Google Home Mini* and *Google Nest Audio*, the option to include minors in the assistant's user group is only available on *Android*. On the rest of the device, it is necessary to set filters, which, as seen previously, affect all users of the assistant equally. Filters can be used to restrict the playback of multimedia content, calls, and the use of third-party *skills*. However, as discussed in Table 4, payments are still active in the SPA unless they are disabled for all users. The *Amazon Echo Show 5* and *Amazon Echo Dot 4* also do not have the ability to set restricted profiles for minors, so the various settings have to be adjusted manually. Restrictions are configured from the device itself, with the password of the linked Amazon account having to be entered each time changes are made. It is possible to restrict the use of the web browser, multimedia content playback or payments. As for the connected accessories, like the rest of the devices, it is not possible to restrict their use.

4.8. Control of answers containing personal information

The processing of responses containing personal information can be divided into two groups, formed by the *Apple HomePod Mini*, *Google Home Mini* and *Google Nest Audio* in one case and by the *Amazon Echo Show 5*, *Amazon Echo Dot 4* and *Facebook Portal* in the other case. On the *Apple HomePod Mini*, *Google Home Mini*, and *Google Nest Audio* the personal responses are related to the selected voice recognition setting. If the user has the voice profile saved and voice recognition is active, when the SPA is asked for information such as calendar events, messages, or reminders, among other examples, it will check which user initiated the query, to provide an answer according to that person's data. If another user asks a question about personal information, they will receive an answer with their own information, if their voice profile is stored, or they will be denied access to the information in the answer if it is a user unknown to the SPA. In contrast to these two devices, there is no method of controlling responses containing personal information in the *Amazon Echo Show 5*, *Amazon Echo Dot 4*, and *Facebook Portal*. The only option to prevent external users from accessing such information is not to use the service that can generate it.

4.9. Authentication methods

Authentication in the SPA analysed is performed through voice recognition. In the *Apple HomePod Mini*, voice recognition authentication is not available in Spanish, while in the rest of the devices analysed this possibility is offered. Under normal use, the devices are able to distinguish between different voices and recognise users, denying access to personal information between users in a satisfactory manner. Authentication is generally not used as a security measure, but as an ancillary functionality to personalise the experience with the SPA. This is due to the insecurity of the speech recognition feature, as explicitly found in the analysis and as stated in the documentation of *Google* Google (2021).This ineffective authentication behavior is caused by the fact that

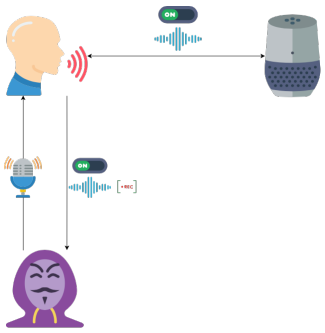

Figure 4: Recording of the activation message by a malicious actor.

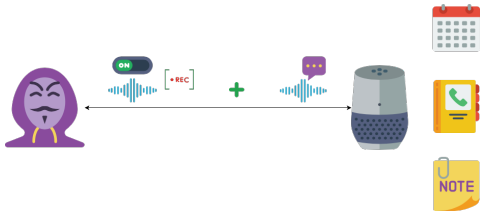

Figure 5: Example of an impersonation attack using a recorded activation message.

in the messages sent by the user to the SPA, only the wake-up word is authenticated. The voice authentication process is only active during the invocation of the assistant. Once the user's identity has been validated in that portion of the message, the SPA assumes that the rest of the message will come from the same user. This leads to the possibility of a malicious actor obtaining a recording of the user invoking the assistant, since, having this information, he/she would be capable of executing all possible requests to the assistant, without the need to have recorded complete messages uttered by the legitimate user.

The process of impersonating a user has been carried out in two phases. In the first phase, the legitimate user makes a request to the SPA, asking for certain personal information. With these tests, it can be verified that the voice recognition system of the SPA is able to distinguish users successfully. During the second phase, an attack by a possible malicious actor in possession of a recording of the user pronouncing the activation message of the SPA was simulated. Within this phase, two stages can be distinguished. In the first stage, the malicious actor has to obtain the voice recording of the legitimate user activating the SPA, as shown in Figure 4. In the second stage, the attacker uses the recording of the legitimate user to activate the SPA and send requests for access to the user's personal information, bypassing the voice authentication, as shown in Figure 5.

The personal information that can be accessed from the SPA will depend on the user's account settings and the information he/she chooses to share with the provider.

The described impersonation procedure has been tested in all SPAs, with satisfactory results on all devices, but with differences. While in the *Google Home Mini*, *Google Nest Audio*, *Amazon Echo Show 5*, *Amazon Echo Dot 4*

and *Facebook Portal* the process of impersonating a user was entirely successful, being able to impersonate a user and access their personal information, the *Apple HomePod Mini* has an additional security measure that prevents a user interacting directly with *Siri* through the *Apple HomePod Mini* from accessing personal information, having to go through an additional authentication process based on the unlocking of the mobile device associated with the *Apple HomePod Mini*, as shown in Figure 6.

Figure 6: Authentication method in HomePod Mini for accessing personal information.

4.10. Non-human voice filtering

The SPAs analysed have proven to be susceptible to activation by artificially generated voices, such as recordings and TTS systems [Figure 7]. They do not implement any kind of protection measures against synthetic voices beyond turning off the microphones of the devices when they are not being physically controlled. This finding, which by itself may not seem to present a serious problem, combined with a poor authentication process, as detailed in the previous section, leads to the possibility that a malicious actor with a recording of the user's activation message can make any query to the SPA, bypassing voice recognition controls and being able to extract personal information from the device, interact with connected accessories in the home, etc.

Figure 7: Activation of SPA through artificial voice systems.

4.11. Interaction with conversation history

The options for interacting with the conversation history are different for each of the devices. The *Apple HomePod Mini* offers the least possibilities, only being able to send a request to delete the conversation recordings stored on *Apple's* servers. The *Google Home Mini* offers a full privacy settings section for viewing and deleting voice recordings, as well as setting up automatic deletion of recordings and pausing the storage of recordings. The options that *Amazon* incorporates into the *Amazon Echo Show 5* and *Amazon Echo Dot 4* are similar to those of *Google*, offering a full set of options for reviewing conversations with the SPA, deleting them, and setting up automatic deletion of data. In the case of *Facebook Portal*, since it uses *Amazon Alexa* as a voice agent, something similar happens. In fact, *Facebook Portal* records voice clips when the users activate the Smart Assistant by saying "Hey portal" and it sends back to *Facebook* these clips. Also, these voice clips are recorded and sent back to *Amazon*. The data extracted from conversations identified by *Facebook Portal* is used to display advertising across *Facebook*. The company may also share specific demographic and audience engagement data with advertisers and analytics partners.

5. Discussion and conclusions

SPAs are becoming prevalent devices in our lives and homes, and are therefore a potential target for malicious actors. In addition, the unique interaction characteristics between the user and the device widen the attack surface. Of particular concern are attacks in the Access Control category, which allow malicious actors to interact with the SPA, access sensitive information, and even make purchases on behalf of the user.

Particularly relevant in this paper are the findings regarding the user authentication systems that the assistants incorporate, with deficient protection measures against impersonation attempts involving a voice recording of an activation message uttered by a legitimate user. This implies that an unauthenticated user can perform orders even without having a similar voice to the one registered. In fact, tests show that it does not even have to be a naturally reproduced human voice, the SPA device does not correctly differentiate between female human voices, male human voices, recorded and reproduced voices, and even simulated voices through an artificial generator.

Our paper shows a gap in the literature offering robust systems for authenticating users in SPA, connected with a lack of adoption from manufacturers of prospective security and privacy mechanisms developed for other platforms. For example, a promising future line of work is continuous speech recognition, which attempts to prevent the attacker from using a recording of the user solely uttering the activation command. With a continuous authentication system, the entire message received by the attendee would be authenticated, as well as subsequent interactions in the conversation, to hider attacks leveraging replay attacks of authentication challenges. Moreover, an additional message processing layer would be needed to identify whether the voice in the message is from an artificial source (e.g., deep fake audio) or a human voice. This implies, of course, challenges from the AI algorithms point of view, which should also be robust against adversarial AI attacks **?**, **?** in order to be fully operational.

Users should be aware of the security and data control flaws that these devices currently present, and should be careful with SPA configurations to prevent access to personal data without extra identification (e.g., two-factor identification on the smartphone). Manufacturers should also

design the SPAs in such a way that users configure the safety options while being aware of the dangers to which they are exposed and giving their explicit consent to the privacy options, as required by data protection regulations such as GDPR Commission (2018).

Since SPAs are more and more used by minors **?**, it is of major importance the possibility to set up profiles for minors that limit some functionalities and protect them from inappropriate content and from potential stalkers and fraudsters. The protection of minor users through simple and quick settings is practically non-existent in the conditions analysed (see Table 2), whereas this has already been included in other related applications, such as streaming platforms like *Netflix*, which already includes a specific default profile for them. In the case of SPAs, however, it is necessary to go through the entire set of device settings, even having to switch between different configuration menus, to disable all features that may be unsafe for an unsupervised minor user. As a result, the access of minors to inappropriate content or actions in the case of SPAs rely heavily on the technical knowledge and skills of their guardians, which is unfair for minors and may entail not fulfilling current security and data regulation that is typically more strict in their case.

It has been also found that, in many cases, restricted device settings affect all users equally, leaving features unusable for all members of the household, which is impractical and does not encourage users to establish usage controls.

Due to the complexity and variety of options that modern SPAs enable to protect minors, we see the need to improve usability. This can be done through the definition of different pre-configured security profiles, with access policies that allow controlling in a simple but precise fashion how users interact with the devices, as it is done in other multimedia content platforms, such as *Netflix* or *HBO*. In the ecosystem of the SPA, the creation of security profiles for minors is more complex, since the functionalities of the devices, and therefore the control requirements, are more varied and complex than in a multimedia playback application, so it is necessary to thoroughly analyze the characteristics of the SPA to identify and assess the risks that the use by a minor can produce.

However, based on the need to rely on the voice samples that the assistant captures at the time of invocation by a user to authenticate him, it is necessary to develop a robust system that allows distinguishing between minor and adult users each time a query is sent to the assistant, in order to apply security profiles automatically. This would facilitate the use of the security options included in the SPA, since they would be applied transparently to users, just as the authentication mechanisms are applied in current smartphones.

This research contributes to shed some light and verify some security flaws that have already been identified in the literature in commercial SPAs with the aim of increasing awareness and promoting progress in security and data control in SPAs, paying special attention to vulnerable groups such as minors.

6. Acknowledgements

This work has been funded by the Horizon 2020 program of the European Union through the RAYUELA project (contract no. 882828). Guillermo Suarez-Tangil was partially funded by the "Ramon *y* Cajal" Fellowship RYC-2020-029401-I supported by TED2021-132900A-I00 and COMODIN-CM funded by European Regional Development Fund (ERDF) and the Comunidad de Madrid respectively. The European Commission is not responsible for any use that may be made of the information contained therein.

The authors have no relation with the manufacturers of the analysed Smart Personal Assistants. The studied Smart Personal Assistants were bought and have been exclusively used for research purposes. No human beings have been involved in the performed tests except from the researchers who carried out the tests themselves, so the data managed by the Smart Personal Assistants is fake data and thus does not represent sensitive data at all.

References

Abdi, N., Ramokapane, K.M., Such, J.M., 2019. More than smart speakers: Security and privacy perceptions of smart home personal assistants, in: Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019), USENIX Association, Santa Clara, CA. pp. 451–466. URL: https://www.usenix.org/conference/soups2019/presentation/abdi.

Abdi, N., Zhan, X., Ramokapane, K.M., Such, J., 2021. Privacy norms for smart home personal assistants, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/3411764.3445122, doi:10.1145/3411764.3445122.

Alexa, 2020. Alexa internet privacy notice. https://www.alexa.com/help/privacy. Accessed: March 2021.

Amazon, 2021a. Alexa accesibility - how to change your wake word. https://www.amazon.com/-/es/b?ie=UTF8Źnode=21341305011. Accessed: June 2021.

Amazon, 2021b. Alexa features. https://www.amazon.com/b?ie=UTF8Źnode=21576558011. Accessed: March 2021.

Canalys, 2019. Global smart speaker installed base to top 200 million by end of 2019. https://www.canalys.com/newsroom/canalys-global-smart-speaker-installed-base-to-top-200-million-by-end-of-2019. Accessed: March 2022.

Carlini, N., Wagner, D., 2018. Audio adversarial examples: Targeted attacks on speech-to-text, in: 2018 IEEE Security and Privacy Workshops (SPW), pp. 1–7. doi:10.1109/SPW.2018.00009.

Cheng, P., Roedig, U., 2022. Personal voice assistant security and privacy–a survey. Proceedings of the IEEE , 1–32doi:10.1109/JPROC.2022.3153167.

Commission, E., 2018. General data protection regulation (gdpr). https://gdpr-info.eu/.

Edu, J.S., Such, J.M., Suarez-Tangil, G., 2019. Smart home personal assistants: A security and privacy review. CoRR abs/1903.05593. URL: http://arxiv.org/abs/1903.05593, arXiv:1903.05593.

Google, 2021. Link your voice to your devices with voice match. https://support.google.com/assistant/answer/9071681}vm_pr. Accessed: March 2021.

Lau, J., Zimmerman, B., Schaub, F., 2018. Alexa, are you listening? privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. Proc. ACM Hum.-Comput. Interact. 2. URL: https://doi.org/10.1145/3274371, doi:10.1145/3274371.

Li, J.B., Qu, S., Li, X., Szurley, J., Kolter, J.Z., Metze, F., 2019. Adversarial music: Real world audio adversary against wake-word detection system. arXiv:1911.00126.

Liu, S., 2021. Smart home in the united states - statistics & facts. https://www.statista.com/topics/6201/smart-home-in-the-united-states/. Accessed: March 2021.

Mozilla Foundation, 2020. *privacy not included. https://foundation.mozilla.org/en/privacynotincluded/amazon-echo-dot/. Accessed: March 2021.

Ruan, S., Wobbrock, J.O., Liou, K., Ng, A., Landay, J.A., 2018. Comparing speech and keyboard text entry for short messages in two languages on touchscreen phones. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1, 1–23. URL: http://dx.doi.org/10.1145/3161187, doi:10.1145/3161187.

Statista, 2019a. Number of smart home devices supported by google assistant worldwide from january 2018 to may 2019. https://www.statista.com/statistics/933532/worldwide-google-assistant-device-support/. Accessed: March 2021.

Statista, 2019b. Smart speaker with intelligent personal assistant market share in 2018 and 2019, by platform. https://www.statista.com/statistics/1005558/worldwide-smart-speaker-market-share/. Accessed: March 2021.

Statista, 2020a. Digital voice assistant installed base worldwide in 2019, by brand. https://www.statista.com/statistics/967412/worldwide-digital-voice-assistant-installed-base-brand/. Accesed: June 2021.

Statista, 2020b. Smart speaker use case frequency in the united states as of january 2020. https://www.statista.com/statistics/994696/united-states-smart-speaker-use-case-frequency/. Accessed: March 2021.

Statista, 2020c. Total number of smart home devices that are compatible with amazon's alexa as of july 2020. https://www.statista.com/statistics/912893/amazon-alexa-smart-home-compatible/. Accessed: March 2021.

Statista, 2021. Total number of brands compatible with amazon alexa from january 2018 to july 2020. https://www.statista.com/statistics/912903/amazon-alexa-brand-use-growth/. Accessed: March 2021.

Yuan, X., Chen, Y., Wang, A., Chen, K., Zhang, S., Huang, H., Molloy, I.M., 2018. All your alexa are belong to us: A remote voice control attack against echo, in: 2018 IEEE Global Communications Conference (GLOBECOM), pp. 1–6. doi:10.1109/GLOCOM.2018.8647762.

Zhang, N., Mi, X., Feng, X., Wang, X., Tian, Y., Qian, F., 2019. 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. doi:10.1109/SP.2019.00016.