L. Hernández Acosta and D. Reinhardt. Multi-User Privacy with Voice-Controlled Digital Assistants. Proceedings of the 20th International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Work-in-Progress), pp. 30-33, 2022.

Multi-User Privacy with Voice-Controlled Digital Assistants *

Luca Hernández Acosta
Computer Security and Privacy (CSP)
University of Göttingen
Göttingen, Germany
hernandez@cs.uni-goettingen.de

Delphine Reinhardt
Computer Security and Privacy (CSP)
University of Göttingen
Göttingen, Germany
reinhardt@cs.uni-goettingen.de

***Abstract*—With the help of Voice-controlled Digital Assistants (VCDAs), end users can perform various tasks, such as creating shopping lists, setting reminders, or controlling smart home devices via voice commands. However, in multi-user environments, the different end users of VCDAs may not have access to the same controls to protect their privacy. The primary end users who set up VCDAs usually have full control over the data collected by VCDAs, including text transcripts and audio recordings of the other end users. In order for these secondary end users to gain access to privacy settings, they must also create an account with the appropriate manufacturer and accept an invitation from the primary end user to join the respective VCDA. As a result, they depend on the primary end user and the creation of a user account to be able to protect their privacy. Through a user account, however, personal information, such as name, address, or age can be linked to audio recordings, that poses additional privacy risks to secondary end users. For both primary and secondary end users, audio recordings are still maintained on cloud servers operated by manufacturers, resulting in a lack of transparency for all end users. In this paper, we thus propose an approach to improve the protection of both primary and secondary end users that reaches from the device set-up to its utilization. Our approach is based on the concept of a local registration and offline storage of voice commands.**

***Index Terms*—voice assistants, multi-user environment, privacy, authentication**

I. Introduction

VCDAs are able to support end users in a variety of everyday tasks, leading to an increased usage in recent years [1]. The comfortable way of interaction as well as the various application areas of these assistants have contributed to their increasing adoption in private as well as professional environments [2]–[4]. Often, VCDAs are located in private households and have the capability to control other connected devices such as lights or thermostats. However, despite their convenience, interacting with a VCDA raises privacy issues. As illustrated in Fig. 1, end user's voice commands are recorded by the VCDA and then forwarded to cloud servers of manufacturers or third-party developers where they are processed and stored. The wake word detection as well as the verification of the respective end user is not only performed locally on the device, but is also supported by the cloud server [5].

In order to protect the individual end user privacy, manufacturers provide certain privacy settings. Currently, the end user can view past interactions, delete them, or even set intervals to automatically delete requests. However, these settings are limited and usually do not respect the privacy by default principle, as they are initially set to indefinitely store the collected data. Moreover, depending on the VCDA, users face different configuration possibilities, that have been shown to change over time. As a result, it becomes increasingly complex for the end user to apply these privacy settings [6], [7].

As VCDAs are increasingly used in households where multiple people interact with it, further privacy issues arise. The primary end user of the VCDA, the one who put the device into operation, has full control over it. This allows the primary end user to access the interactions of everyone in the household who has spoken to the VCDA. By default, secondary end users have no way to protect themselves from monitoring by the primary end user and rely on the primary end user to handle their data properly. The only way how secondary end users are able to protect their privacy is by creating an account with the respective manufacturer and accept an invitation sent by the primary end user to join the corresponding VCDA. Then, interactions triggered by a secondary user are only linked to that respective account and the primary end user is not capable to view those interactions anymore. Without an account the system cannot be used. Since this account is often linked to other information about the end user, such as name, place of residence, or means of payment, there is a further risk to privacy. This, along with information stored on their cloud servers from voice commands, allows manufacturers to map even more detailed user characteristics that could provide insight into their interests and other sensitive information [8].

In this paper, we present an approach to help both primary and secondary end users protecting their privacy when using VCDAs in multi-user environments. With the help of our approach end users can:

1) Create a local user account.
2) Store audio recordings locally on the VCDA.
3) Access local audio recordings with smartphone or wearable.
4) Protect recordings from other participants in a multi-user environment.

*The final publication is available at IEEE Xplore via http://dx.doi.org/10.1109/PerComWorkshops53856.2022.9767270

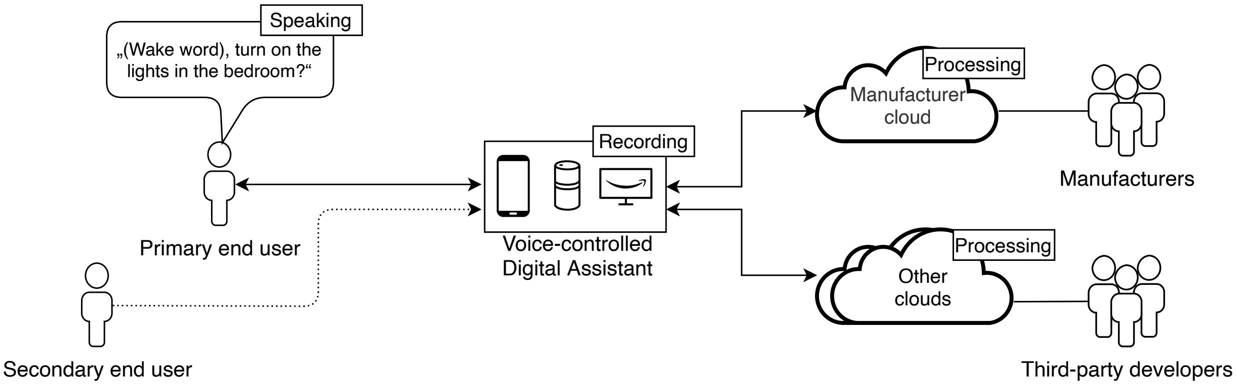

Fig. 1. The *Voice-controlled Digital Assistant (VCDA) ecosystem*

The remainder of this paper is structured as follows. In Sec. II we discuss related work. Further, in Sec. III we describe the adopted scenario which we are focusing on. In Sec. IV we describe our proposed approach. In Sections V and VI we state out potential limitations of our approach and give an outlook on future research, respectively. Finally we conclude our research in Sec. VII.

II. Related Work

Different articles have highlighted the fact that privacy is not adequately protected when using VCDAs [6], [7]. The main reasons are the processing and storage of data on the cloud servers of the manufacturers. Currently, raw voice data is processed on the servers of manufacturers and third-party developers. Thus, in addition to the content of the voice command and the voice profile, other sensitive information can be derived from the voice. As Kröger et al. have shown in [8], some features of the voice are sufficient to draw conclusions about age, gender, or health aspects.

In order to avoid inferences about personal characteristics, solutions have been proposed that make use of voice obfuscation [9], [10]. In addition to this obfuscation, the content to be transmitted can also be filtered so that user-defined sensitive words are replaced before being processed in the cloud [9], [10].

To circumvent online processing of voice data, offline processing solutions have also been proposed to enhance end user privacy [11], [12]. However, this requires the end users to perform some additional effort in training and testing the assistant with user specific voice commands, leading to a negative impact on user acceptance.

So far privacy in multi-user environments, has only been discussed theoretically, although it has been considered an important aspect for potential end users [13]–[15]. Nevertheless, Google recently proposed an initial solution to protect end users by requiring each end user to set up their own account and enable personalized responses by creating a voice profile. With those personalized responses Google Assistant can then differentiate between different end users by their voice and only show results that are relevant to the specific end user. Voice recordings are then mapped to the respective account and stored online on Google cloud servers. Likewise, Apple offers this option for personalized results, however, this option is not available for all languages.

While we can observe that some efforts have been made to protect the privacy of all end users in multi-user environments, issues such as the need to create a user account with the manufacturer and the online storage of voice recordings still exist. In contrast to related work, we focus on an approach that enables the creation of local accounts on the VCDA and offline storage of voice recordings on the VCDA rather than on cloud servers.

III. Scenario

We assume that multiple people are interacting with the same VCDA. In our approach when setting up a VCDA end users have to use their Smart Device (SD) (e.g. smartphone or smartwatch) in order to connect to the VCDA and create a user account locally on the VCDA. Additional end users would have to register in the same way. After the registration is done both primary and secondary end users are able to choose their privacy settings, i.e. whether they want to create audio recordings, automatically delete recordings, or create no history at all. According to the chosen setting end users are able to view past interactions on their SD and decide to delete or keep them on the VCDA.

IV. Proposed Approach

With our new approach, we want to help potential end users protect their privacy in multi-user environments. We specifically address the setup process for VCDAs and describe what changes to the current process for off-the-shelf VCDA are required. In the following sections, we describe in more detail how we would like to implement accountless authentication and how an offline voice command history could be achieved.

A. Accountless Authentication

Fig. 2 shows the current setup process as well as our proposed process. Currently, when setting up the VCDA, users must download a companion app for their smart phone and set up a user account with the respective manufacturer. Without

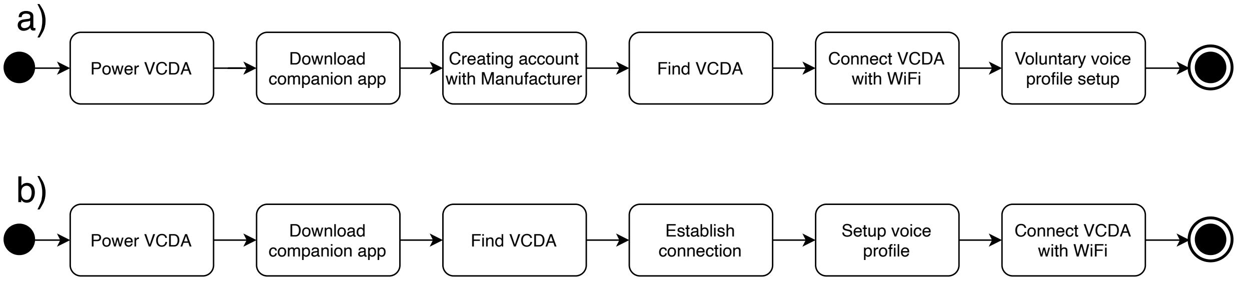

Fig. 2. Comparison of the current setup process a) and the process planned to be used in our potential prototype b)

a user account, the speech processing service cannot be used. After the VCDA has been connected to the WiFi, a voice profile can then be optionally created. This voice profile is not only used for a more personal user experience, but also ensures that voice recordings are linked to the respective end users.

In our proposed setup process Fig. 2 b) we still need a companion app for an SD, but an online user account is no longer needed. Instead, we would like to introduce a different form of authentication. In our approach, the end-user utilizes his/her SD to perform mutual authentication with the VCDA based on the station-to-station protocol that implements authenticated Diffie and Hellman key exchange by combining signature algorithms (e.g., DSA, RSA) [16]. In Fig. 3 we briefly highlight the initial steps in order to provide a secure connection between the VCDA and an end user's SD. In a first step the VCDA as well as the SD have to generate their public and private key pairs that are necessary to sign all exchanged messages. The public keys are then exchanged and the Diffie-Hellman key exchange is initiated. For the sake of simplicity we do not cover the Diffie Hellman approach in detail but want to state that the approach is used to generate a symmetric key that can later be used for a secure communication between the VCDA and the SD.

When the station-to-station protocol was successful, the voice of an end user is recorded via the SD and forwarded to the VCDA in order to create a voice profile. After the voice profile has been created, the authenticated SD from which the voice recording was sent is linked to the newly generated voice profile, so that future interactions can be assigned not only to a specific voice but also to the SD of the end user, so that access to the respective audio recordings can be obtained afterwards. Since end users may not want to access the VCDA only with one SD, they would need to repeat the accountless authentication steps for each additional device, which will then be associated with the end user's voice profile.

We would like to state that the processing of the voice and the steps of Natural Language Processing (NLP) for speech recognition are still done online, so that possible inferences about personal traits are possible via features in the voice. To address this issue, the prototype could be equipped with obfuscation in a next step, as already described in [9], [10].

In summary, our solution will result in the creation of a local user account linked to the end users' devices. As a result, multiple accounts could be easily set up without any dependencies to other end users.

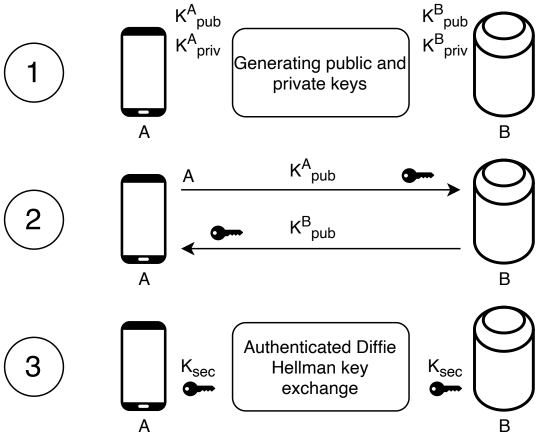

Fig. 3. Steps of the proposed accountless authentication scheme

B. *Offline Voice Command History*

VCDAs currently store voice recordings on manufacturers' cloud servers, where they are linked to the respective user accounts. For most commercially available VCDAs, these recordings are by default associated with the account that the device was originally set up with. Therefore, recordings that may have been triggered by other, unregistered end users are still linked to the account that was used for the initial setup. In a multi-user environment, this means an increased risk to privacy for all unregistered end users, as their commands can be also accessed by others.

As described in IV-A our proposed system requires end users to be registered from their first use, leading to account registration for every end user locally on the VCDA in a multi-user environment. Since there are no online accounts in our approach, the voice recordings are stored offline on the VCDA, where they can be directly associated with the respective local user account. However, end users will still have the option to choose in their mobile application whether or not to store a history of recordings. If a history of audio recordings should be stored, those recordings can then be viewed, deleted, or automatically deleted accordingly. Moreover the corresponding application keeps track of the history by frequently being updated from the VCDA in the local network.

In comparison to the current state of the art, we hence propose to store voice recordings offline in order to provide increased transparency and control over the data shared with a VCDA.

V. Discussion and Limitations

For the realization of our approach, the use of an SD is mandatory to perform the setup process. Without an SD, an end user cannot authenticate with the VCDA and thus cannot use it. In a household where there may be children who also want to interact with the VCDA, there is a risk that operation is not possible due to the lack of an SD. Thus, user groups without SDs would be excluded from using the VCDA from the outset. One conceivable approach would be to continue to guarantee operability for people without SDs, but to protect their privacy by not storing any recordings by default. In this way, users would give up some control over their interactions, but would not be excluded from using VCDAs, and their privacy would be respected because no voice recordings would be stored. As the new approach is different to the current state of the art when setting up a smart speaker it has to be investigated whether users are comfortable with this new approach or if the new approach hinders end users to setup the VCDA. During the development of the prototype, we will also investigate whether it is possible to create voice profiles locally on the SD and share them with the VCDA. Since there are already offline platforms for voice recognition, we do not want to exclude offline voice profiling, but we would also like to point out that the current offline approaches impose a non-negligible additional effort on the end user and could thus have a negative impact on its adoption.

VI. Future Research

After the prototype will be successfully implemented and tested, we will evaluate the prototype in a user study. Firstly, we would like to capture the mental model of our participants in relation to the new methodologies and thus determine whether individual understanding of the system differs in real-world implementation. Secondly, we will examine its usability in more detail. For example, when using the new setup process, participants may have difficulty using this new feature because it is not based on email address registration like other processes. We will also investigate whether there are differences in the performance of the methods offered and whether participants prefer a particular method.

VII. Conclusion

Common methods of data control in the use of VCDA provide only limited privacy protection for end users. Solutions that are designed for primary end users are often not directly available to secondary end users. While primary end users already have the ability to view and delete the history of their interactions, in multi-user environments, these control mechanisms are often not available to secondary end users because they do not have their own account registered on the VCDA. When using VCDA in multi-user environments, there is often a privacy conflict between primary and secondary end users, as primary end users have more possibilities to protect their privacy and moreover control over interactions made by secondary end users. In this paper, we present a first concept on how secondary end users can protect their privacy in multi-user environments and regain control over their data. The proposed setup process includes the need to create local user accounts with voice profiles for each end user before using the VCDA and to store voice recordings local on the the VCDA. By following this approach it is possible to identify the respective end user and to associate interactions made, so that other end users do not get access to these interactions anymore.

References

[1] S. Han and H. Yang, "Understanding Adoption of Intelligent Personal Assistants," *Industrial Management & Data Systems*, 2018.

[2] M. B. Hoy, "Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants," *Medical Reference Services Quarterly*, 2018.

[3] M. Porcheron, J. E. Fischer, S. Reeves, and S. Sharples, "Voice Interfaces in Everyday Life," in *Proc. of the 18th Conf. on Human Factors in Computing Systems (CHI)*, 2018.

[4] Z. Y. Chan and P. Shum, "Smart Office: A Voice-Controlled Workplace for Everyone," in *Proc. of the 2nd International Symposium on Computer Science and Intelligent Control (ISCSIC)*, 2018.

[5] L. Schönherr, M. Golla, T. Eisenhofer, J. Wiele, D. Kolossa, and T. Holz, "Unacceptable, Where Is My Privacy? Exploring Accidental Triggers of Smart Speakers," *arXiv preprint arXiv:2008.00508*, 2020.

[6] J. Lau, B. Zimmerman, and F. Schaub, "Alexa, Are You Listening? Privacy Perceptions, Concerns and Privacy-Seeking Behaviors with Smart Speakers," *Proc. of the ACM on Hum.-Comp. Interact.*, 2018.

[7] N. Malkin, J. Deatrick, A. Tong, P. Wijesekera, S. Egelman, and D. Wagner, "Privacy Attitudes of Smart Speaker Users," *Proc. on Privacy Enhancing Technologies (PoPETs)*, 2019.

[8] J. L. Kröger, O. H.-M. Lutz, and P. Raschke, "Privacy Implications of Voice and Speech Analysis–Information Disclosure by Inference," in *Proc. of the 14th IFIP Int. Summer School on Privacy and Identity Management (IFIP SC)*, 2019.

[9] J. Qian, H. Du, J. Hou, L. Chen, T. Jung, X.-Y. Li, Y. Wang, and Y. Deng, "Voicemask: Anonymize and Sanitize Voice Input on Mobile Devices," *arXiv preprint arXiv:1711.11460*, 2017.

[10] J. Qian, H. Du, J. Hou, L. Chen, T. Jung, and X.-Y. Li, "Hidebehind: Enjoy Voice Input with Voiceprint Unclonability and Anonymity," in *Proc. of the 16th ACM Conf. on Embedded Networked Sensor Systems*, 2018.

[11] A. Coucke, A. Saade, A. Ball, T. Bluche, A. Caulier, D. Leroy, C. Doumouro, T. Gisselbrecht, F. Caltagirone, T. Lavril *et al.*, "Snips Voice Platform: An Embedded Spoken Language Understanding System for Private-by-Design Voice Interfaces," *arXiv preprint arXiv:1805.10190*, 2018.

[12] Rhasspy. (2019) Rhasspy Voice Assistant. [Online]. https://rhasspy.readthedocs.io/en/latest/, accessed in 2021-12-12.

[13] E. Zeng and F. Roesner, "Understanding and Improving Security and Privacy in Multi-User Smart Homes: A Design Exploration and In-Home User Study," in *Proc. of the 28th USENIX Security Symposium (USENIX Security)*, 2019.

[14] N. Abdi, X. Zhan, K. M. Ramokapane, and J. Such, "Privacy Norms for Smart Home Personal Assistants," in *Proc. of the 48th Conf. on Human Factors in Computing Systems*, 2021.

[15] J. S. Edu, J. M. Such, and G. Suarez-Tangil, "Smart Home Personal Assistants: A Security and Privacy Review," *ACM Comput. Surv.*, 2020.

[16] W. Diffie, P. C. Van Oorschot, and M. J. Wiener, "Authentication and Authenticated Key Exchanges," *Designs, Codes and cryptography*, 1992.