Dummy-Based Anonymization for Voice-Controlled IoT Devices

Katrin Winkler and Erik Buchmann

Hochschule für Telekommunikation Leipzig, Germany
Email: {s111132|buchmann}@hft-leipzig.de

***Abstract*—Voice assistants like Amazon Alexa, Google Assistant or Siri are becoming increasingly popular. Such assistants allow for complex interactions with smart Internet-of-Things (IoT) devices that do not have a traditional user interface, such as monitor and keyboard. However, while voice assistants foster the proliferation of numerous convenient services from smart homes to connected cars, they are problematic from the perspective of user privacy. In many cases, IoT devices are permanently listening for keywords in sensitive areas such as living rooms or bed rooms. Once such a word is recognized, voice samples are sent to the voice-assistant provider into the cloud for further analyses. We explore how the users of IoT devices can anonymize the voice recordings sent to the voice-assistant provider. To this end, we identify categories of information sent to the provider, we describe an anonymization approach based on dummy voice commands, and we describe a prototypical anonymization device based on a Raspberry PI. Our device confirms that it is possible to anonymize some information sent to Alexa with limited inconveniences for the user.**

Keywords–User Privacy; Internet of Things; Voice Control.

I. Introduction

A major success factor for the Internet of Things (IoT) is the availability of reliable, user-friendly voice assistants. Without assistants like Alexa, Siri, Bixby or Cortana, it would be difficult to integrate IoT services with everyday appliances that do not possess graphical user interfaces. Today, a large number of voice-controlled services exist. Such services manage light bulbs, radio and video receivers, heating systems or alarm equipments, provide access to emails, text messages and calendar information, and activate vacuum cleaner robots.

A voice-controlled IoT device consists of one or more microphones, a small voice processor and an Internet link to a cloud service. The microphone records environmental sounds, which are locally processed. When the IoT device recognizes a preconfigured wake-up word ("‘Alexa, ..."’, "‘Ok Google, ..."’), it sends a few seconds of sound records to a cloud service. This cloud service does a more complex voice processing in order to extract spoken commands. Finally, the cloud service sends – depending on the IoT service invoked – control commands and/or information back to the IoT device (cf. Figure 1).

While this approach works very well from a technical point of view, and supports a plethora of useful and user-friendly IoT services, it is problematic from a privacy perspective. Typically, voice assistants are permanently listening, and placed in highly private areas, such as living rooms, kitchens or bed rooms. Experience has shown that the IoT devices react not only on the owner saying the wake word [15]. Thus, the service provider might overhear deeply private conversations.

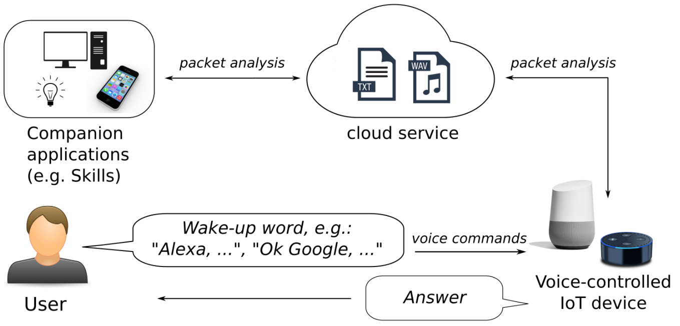

Figure 1. Ecosystem of voice-controlled IoT devices [26]

Many voice-controlled IoT devices integrate third-party services. For example, Amazon Alexa allows to control components of the Philips Hue lighting system. Thus, the processing of voice commands depends on a complex interaction of services from various parties. In consequence, the privacy policies of the service providers are also complex. Data breaches might occur at any place in this service architecture, revealing deeply sensitive personality profiles. Finally, guests of the device's owner might not be aware that their conversation can be transferred to a cloud service for further processing.

The purpose of this paper is to fuel the ongoing discussion about options on the user's side to mitigate the impact of voice assistants on privacy. To this end, we explore technical options to anonymize IoT devices using voice assistants from the user perspective. That is, we systematically analyze in which way a user of an IoT device can control or influence voice records sent to an IoT service in order to have certain sensitive information anonymized. As a prominent example, we focus on Amazon Echo. We explicitly leave aside legal questions, and we assume that the IoT service provider handles any user data as described in it's privacy policy. We also do not discuss the responsibility of the users, say, not to use voice assistants in public spaces [40]. In particular, we make the following contributions:

- We identify which private information can be potentially observed by the service provider.
- We analyze options for the user to control the voice records sent to the provider.
- We explore a dummy-based approach to anonymize the information sent to the provider.
- Finally, we describe what we have learned from implementing a prototypical anonymization device.

Our anonymization device demonstrates that it is possible to anonymize a part of the information sent to Amazon Echo, with little inconvenience for the user. We point out that some information cannot be anonymized without making the service useless. However, it is challenging to infer what the service provider might learn from the voice records sent to the cloud.

The rest of the paper is structured as follows. Section II reviews related work. Sections III and IV introduce Amazon Alexa and Amazon Echo. Section V outlines options to anonymize voice assistants. Section VI describes our anonymization device. The paper concludes with Section VII.

II. Related Work

In this section, we review related work on (a) voice assistants for IoT services, (b) data privacy in IoT and (c) dummy-based anonymization techniques.

A. Voice Assistants

The IoT is evolving rapidly. The worldwide IoT market is expected grow to $1.1 trillion in 2021 with a compound annual growth rate of 14.4% [16]. Similarly, the worldwide market for virtual personal assistant-enabled wireless speakers will reach $2.1 billion by 2020 [12]. Currently, the U.S. market is dominated by Amazon Echo with 73% and Google Home with 27% [13]. Thus, voice assistants like Amazon Echo, Google Home or Apple HomePod are popular devices for users to control their smart home appliances, adjust thermostats, activate home security systems, purchase items online, initiate phone calls, and many other tasks in their daily life. Voice assistants and smart speakers exist in many different types with varying level of capabilities. Some of them can be integrated into party products, such as connected cars [46]. For example, BMW customers can enable different vehicle functions via voice assistants like Google Home or Amazon Alexa since November 2017 [7]. Other examples are smart fridges like Samsung Family Hub 3.0, which communicates with its users via Bixby [14]. There are even voice-controlled toys for children (Mattels Hello Barbie [39]) or toilets with integrated voice assistants [9].

B. Privacy in the Internet of Things

As soon as appliances interact with their users in a natural way, such as per voice, humans tend to see them as partners instead of machines [30]. Thus, users are tempted to assign attributes regarding morale, attitude or responsibility to voice assistants [31]. This is problematic, as voice assistants often have access to data with a high impact on privacy and security [23]. This allows for numerous new threats and attacker models [27]. A large share of cell phone users face similar privacy risks as the users of voice-controlled IoT devices. However, in direct comparison cell phone users adjust their system settings in much more restrictive way [24] than users of voice-controlled IoT devices.

Since voice assistants typically do not distinguish different speakers – a feature that is helpful for anonymization – adversaries in microphone range might issue commands to the voice assistant. This has been already demonstrated in multiple ways [34], e.g., via television, via ultrasonic frequencies that are too high for the human ear to hear, or through closed windows. Furthermore, manipulated voice commands can lead to financial losses because the user's payment data is often accessed directly [34]. Currently, the best approach for consumer privacy is to unplug any IoT device when not in use. In addition, the user should frequently review the voice assistants history for unauthorized actions [34]. Typically, this history is accessible via the Web page of the service provider.

C. Anonymization Techniques

Anonymization means to protect the privacy of an individual regarding certain features, e.g., the presence of the individual in a data set, if the individual shares similar characteristics with a control group or if the individual can be assigned with certain attributes. There is a broad variety of anonymization approaches available (see [21] for an overview). State of the art is Differential Privacy [29], which ensures that the presence or absence of an individual record in a database has a very small impact on the result of a specific analysis. However, such privacy measures need data from multiple users as an input for the anonymization. From the perspective of a single user who wants to protect himself against a curious service provider, such approaches cannot be applied.

Dummy-based annonymization [35][36] has been extensively studied in the context of location based systems. Dummy-based anonymization means that a user does not only send its real position to a location-based system, but a number of made-up positions as well. From the set of answers provided by the system, the user considers only information regarding his real position. This approach has numerous benefits: (i) Anonymity can be obtained without the help of a trusted third party or other users. (ii) Each user can decide individually which properties to be hidden in the dummy requests. (iii) Finally, the approach scales linearly with the number of dummies sent. However, it is difficult to create a realistic set of dummies where the real information cannot be singled out by statistical means or other properties [32][33]. Dummy-based anonymization has been already studied in other contexts, e.g., social networks [25] or database tables [22].

III. Amazon Alexa

Without loss of generality, we will use Amazon Alexa as a prominent example of a voice assistant. Amazon operates the cloud service that extracts commands from voice records and provides an adequate reaction to this commands. Devices that allow to access Alexa are offered either by Amazon (e.g. Echo Family, Dash Wand, Fire Tablet, Fire TV), or by third-party-providers that provide devices based on a recent Android, iOS or Windows operating system. Other assistants operate in a similar way, e.g., Siri (Apple), Bixby (Samsung) or Cortana (Microsoft).

In this section, we will briefly describe Alexa's IT ecosystem. As Figure 2 shows, the cloud service plays the most prominent role. It allows Alexa to continually adapt to the speech patterns, vocabulary and personal preferences of the users [10]. Furthermore, it allows Alexa to integrate new functionality and to connect to other services by using the Alexa Skills Kit (ASK) and the Alexa Voice Service (AVS).

A. Alexa Skills Kit (ASK)

The Alexa Skills Kit is a collection of APIs, tools, documentations and source code samples. Initially, ASK has been designed for internal Amazon developers to build new features of Alexa's Automatic Speech Recognition (ASR) and Natural Language Understanding (NLU) systems. Right now, ASK is available for any third-party developer [37], and more than 25,000 skills have been built and deployed [1].

Figure 3 shows how ASK processes a request. In a first step, the IoT device waits the wake-up word from the user

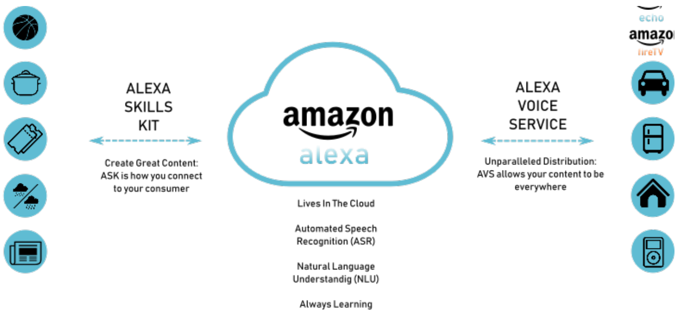

Figure 2. The Alexa Ecosystem [19]

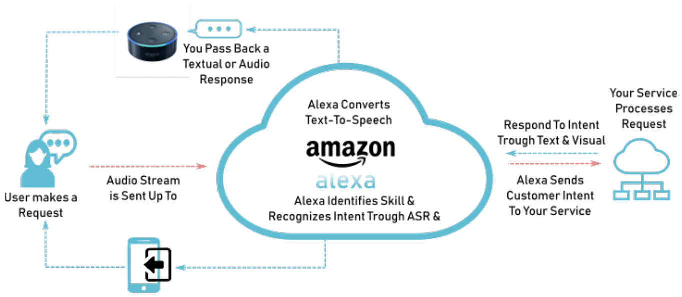

Figure 3. Request and response with the Alexa Skills Kit [19]

and sends a voice recording to the Amazon cloud service that provides Alexa. Alexa identifies the skill and recognizes the user's intent via ASR and NLU. This intent is sent to the service identified by the skill. This service responds to the intent. Depending on the user's device, the response can be a textual message, a verbal answer through Alexa's Text-to-Speech Synthesis (Figure 4) or a graphical response [19]. Any information flow is encrypted.

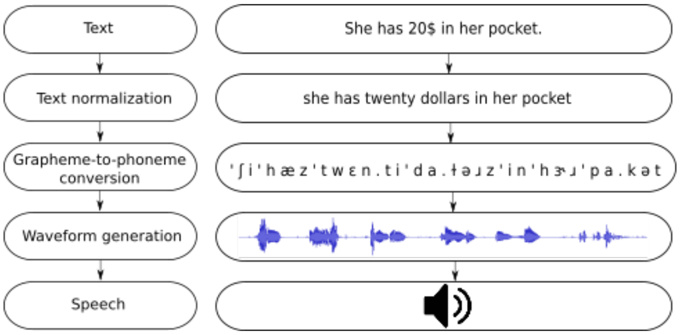
Text
She has 20$ in her pocket.
Text normalization
she has twenty dollars in her pocket
Grapheme-to-phoneme
conversion
'ʃi'hæz'twɛn.ti'da.ɫəɹz'in'hɝɹ'pa.kət
Waveform generation
Speech

Figure 4. Text-to-Speech Synthesis [4]

B. Alexa Voice Service (AVS)

The Alexa Voice Service allows to connect Alexa skills with third-party services and -products. Similarly to ASK, the AVS consists of a set of development tools and resources, e.g., technical documentation and development kits. While ASK is intended to build skills that reside in the Amazon cloud, AVS allows third-party vendors of IoT devices to connect their products to Alexa [2]. Typical examples of such devices are mobile phones, connected cars and smart home appliances.

C. Amazon Web Service (AWS) Lambda

Skill developers are able to use other services from the Amazon cloud [6]. A prominent example is AWS Lambda, a scalable, cloud-based data processing service that executes code in the backend, based on customer-specified events. Such events could be calendar entries, conditions in a stream database, push-services from external sources on the Internet, etc. To use AWS Lamda with Alexa, a developer creates an Alexa skill, uploads it to AWS Lambda and connects it with an event source that triggers its execution. Thus, Alexa is open to big-data management.

IV. Amazon Echo

One of the most prominent IoT devices accessing Alexa is the Amazon Echo and its smaller version, the Echo Dot. These devices are hands-free, voice-controlled speakers which connect to the Alexa Voice Service to play music, control smart home devices, make phone calls, etc. In this section, we will describe the hardware and the user interface of the Echo Dot, together with the respective privacy policy.

A. Hardware and User Interface

Figures 5 and 6 show the structure and internals of an Echo Dot. It comes with seven far-field microphones with beamforming technology and noise cancellation. Thus, verbal commands can be perceived from every direction in a noisy environment [10]. When the Dot recognizes the wake-up word, the light ring turns blue. This allows the user to see when a sound recording is sent into the Amazon cloud. The light also indicates the direction of the voice received. The mute button on the upper side allows to deactivate the microphones, indicated by a red light. Two other buttons adjust the speaker volume.

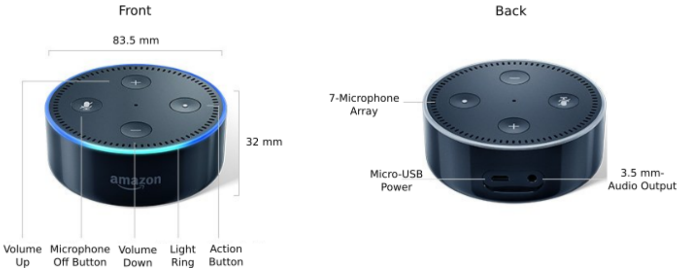

Figure 5. Basic structure of the Amazon Echo Dot [30]

To use the Dot, each user must create an Amazon account. This account is used to configure the Dot, either via the Alexa app or website [3]. The configuration includes network settings, the wake-up word and individual preferences, such as alarms, music, shopping lists and active skills. If the user registers multiple Dots, Alexa always responds from the closest Dot using Echo Spatial Perception (ESP) [20]. Many configuration options can be accessed via voice interface. For example, the default wake-up word can be modified by saying "Alexa, change the wake-up word to Echo". Similarly, the user can instruct Alexa to install a new skill by using the voice command "Alexa, activate <skill name>". Few information is stored locally at the Dot. In particular, this is the network configuration, and preferences, such as the volume and the wake-up word. Any other information is sent to and stored at

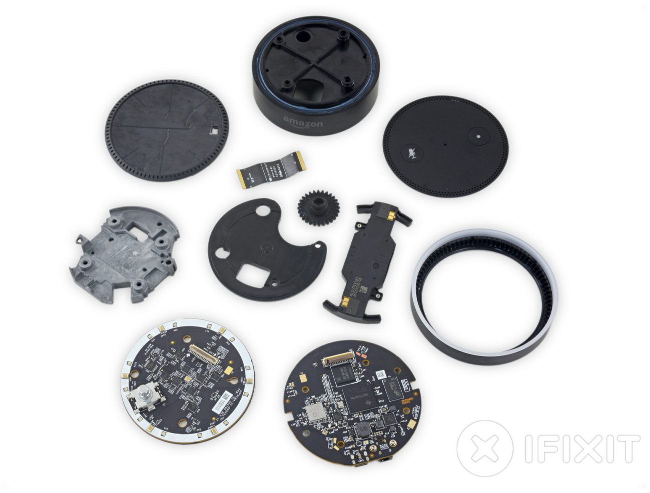

Figure 6. Teardown of the Echo Dot [5]

the Amazon cloud and linked with the user’s personal Amazon account. This is important, as the voice recordings transport much personal information (cf. Section V).

B. *Alexa’s Privacy Policy*

To register an account, the user must consent to Amazon’s privacy policy (In our case, it is the privacy policy of Amazon Europe S.a.r.l., Luxembourg and its subsidiaries). Thus, Amazon does not provide a separate privacy policy for Alexa services that is different from the policy of other Amazon services. Amazon provides a detailed list of personal data, as follows:

- Any information that is sent directly to Amazon, e.g., personal data from the user account, shipping and order information, etc.
- Any information that is obtained automatically when interacting with one of the Amazon services, e.g., browser type, operating systems, time zones or location data of the device used.
- Information from other sources. For example, when a logistics service confirms a delivery, this information is sent back to Amazon.

Furthermore, Amazon’s privacy policy contains a section that explains the stored information with its purposes in detail. This section states that modes of usage, voice recordings and interactions from the past are essential to let Alexa provide meaningful recommendations and reasonable answers. This means that Amazon will not delete any information on its own accord, unless the user issues a request for deletion based on the EU-General Data Protection Regulation or decides to unregister and delete the personal Amazon account. Furthermore, the user might use the Amazon website to delete particular information, e.g., certain voice recordings or orders from the online shop that might result in misleading recommendations. In total, the user has to agree to the following documents [18]:

- Alexa Terms of Use
- Amazon Conditions of Use
- Amazon Privacy Notice
- Cookies & Internet advertising
- Amazon Prime Terms & Conditions
- Amazon Music Terms of Use
- Kindle Store Terms of Use
- Audible Service Conditions of Use
- Amazon Device Terms of Use

V. Privacy and Voice Control

In this section, we identify categories of private information a voice assistant might obtain or infer. Furthermore, we explore alternatives to anonymize such classes of information for voice-controlled IoT devices.

A. *Categories of Private Information*

To the best of our knowledge, there is no survey about the impact of voice assistants on the privacy of its users. However, we can learn from the ongoing discussion on the privacy aspects of smart cellphones [43]. Furthermore, some details can be inferred from the technical setup of voice assistants. We have identified three categories of information that can be obtained from the service provider in the cloud-based backend of a voice assistant:

1. Application Data: This category of data refers to any information a skill must access to perform a certain service. Application data can be information stored in a database at the site of the service provider. It can be also external information fetched via HTTP-request over the Internet. For cell phones, this kind of information corresponds to *content data* [45].

Example: Assume a user wants to manage calendar entries via commands like "‘Alexa, when is my next event?"’ or "‘Alexa, add an event to my calendar"’. In this case, the service provider needs access to the user’s personal calendar. Additional information, such as event priorities or alert times, emphasize the significance of calendar entries.

Privacy Issues: The impact of application data on the privacy of the user depends heavily on the service and the context the service is used. With our example, it makes a difference if the calendar is used personally to manage birthdays of friends, if it is used in a business environment to organize office meetings, or if it is used in a medical facility to fix patient appointments in an aseptic way. In general, the service provider learns from application data:

- The content of an interaction with the voice-controlled device, i.e., user interests, attitudes and personal data.
- The modes of use of a certain service, e.g., habits like asking for a certain stock price every day in the morning.

If the service is offered by an external provider, this information also goes to third parties.

If the voice assistant is used by individual persons, application data is a promising subject for anonymization techniques that aim for*hiding personal interests, attitudes and modes of use*. However, as application data are inherently needed to provide a service, there is a conflict between anonymity and user experience.

2. Technical Data: Data from this category is necessary to provide the service or stem from the domain of the service provider. It is needed or generated automatically when a service is executed and messages flow between the IoT device and the service provider. With cellphones, this is known as *call detail record* [45].

Example: With our personal calendar, the voice-assistant provider needs the user login to identify the personal calendar.

If the calendar is provided by third parties, say, on a Microsoft Exchange server, external login data is also needed. Furthermore, the service provider learns meta-data, such as date and time of use, IP addresses, time-zones etc.

Privacy Issues: Technical data allow to infer personal information that are related to the interaction with the voice assistant. For example, times of use correspond to the daily routine of the users. Furthermore, the IP address can be a user pseudonym and reveals the location of the user. However, any technical information is routed through the cloud of the voice-assistant provider. Thus, if the voice assistant accesses a service provided by a third party, the external provider might only "‘see"’ that a server from the cloud of the voice-assistant provider is communicating – the IP address or the time zone of the user’s IoT device can be hidden by network address translation.

Privacy-Enhancing Technologies, such as the TOR onion routing [28], I2P [44] or the JonDo Web Proxy [42] focus on technical data. Since voice-controlled IoT devices are required to send login information to the service provider anyway, it does not make sense to apply such technologies. However, technical data can be considered for anonymization techniques that *hide usage patterns, habits or activity periods*.

3. Adjunct Data: This category contains information that is neither needed to provide the service nor to communicate with the IoT device, but it is interwoven with other data. It corresponds to issues similar to location tracking [41] or activity recognition [38] of modern cellphones.

Example: The voice recordings sent to the service provider for speech recognition do not only transfer commands to the voice assistant. Instead, features like linguistic stress patterns, regional accents, the number of different speakers using the voice assistant or environmental sounds are part of the recording. Such information is independent from the service used, i.e., it is not bound to our ongoing example of using a personal calendar.

Privacy Issues: It is straightforward to infer sensible personal details from adjunct data. For example, a provider might be able to learn from environmental noises (perhaps accompanied by commands to the voice assistant) that the user performs a certain action, e.g., watching TV or making breakfast.

Information from this class *can be removed without loss of user experience* from the data sent to the service provider. However, this information comes as a byproduct of the sensory equipment or the information technology used to provide the service. Thus, it might be prohibitively expensive in terms of computational costs to filter adjunct data. Basically, this means to perform speech recognition locally instead of using a powerful cloud service. This calls for other options to anonymize voice assistants.

B. Options to Anonymize Voice-Controlled IoT Devices

From the perspective of the individual user, only a few options exist to anonymize a voice-controlled IoT device to some extent without abandoning the use of the device and without sacrificing user experience to a large extent. Firstly, recall that the data flow to and from the IoT device is encrypted. Thus, it is not an option to manipulate the data packets. Secondly, the IoT device is bound to a service provider hosting (a) the voice assistant and (b) the service infrastructure that allows to access a huge number of convenient services, such as calendar management, radio stations, etc. As a result, in many cases it is not an option to simply switch to a more privacy-friendly voice-assistant provider (if there were any). Third, for technical reasons it is not an option to do all voice processing locally at the IoT device and only send commands to the service provider that have been stripped from adjunct data.

However, is is possible to use a variant of the dummy-based anonymization (cf. Section II), i.e., to control what the IoT device is allowed to hear and to hide sensitive personal information in a number of dummy requests that are sent through the IoT device to the voice assistant.

Example: Assume a user wants to anonymize its personal calendar by using dummy requests. In the first step, the user identifies the information to be obscured. With our example, assume the user wants to conceal (i) which are the most sensitive calendar entries, (ii) how many people use the voice assistant and (iii) what are the typical daily activity times. For this purpose, the user asks a number of friends to provide voice samples, such as "‘Alexa, when is my next event?"’, "‘Alexa, add an event to my calendar."’, "‘Alexa, delete an event from my calendar"’ together with times and dates. A reasonable set of dummy requests would order Alexa to read, add and remove calendar entries at different times, by different voices, without having an impact on the correctness of the service. Similar dummy requests can be defined for other services, say, playing radio stations. We see three different options to realize such an anonymization:

External anonymization via speakers: This is the most simple option. An anonymization device with a speaker, e.g., a Raspberry Pi, is placed nearby the IoT device. Whenever the user is in the room, he or she uses the IoT device normally. When the user leaves the room, the anonymization device starts to play voice samples from an internal database containing dummy requests through its speaker to the IoT device. Our anonymization device implements this approach (cf. Section VI).

User Experience: Since the user does not have to manipulate the IoT device, this approach can be easily implemented. Furthermore, as long as the dummy requests are played only as long as the user is not in the room, user experience is high.

Anonymity: Assuming a good set of dummy requests for anonymization, and assuming further that the voice-assistant provider does the speech recognition automatically, it is possible to obscure private application data and adjunct data among dummy requests. Furthermore, it is possible to hide habits and activity times. However, a suspicious voice-assistant provider may sort out dummy requests that follow a different data distribution than the real requests, that are heard always from the same direction at the same sound volume or that are issued with the same accentuation. Furthermore, this approach assumes that the IoT device is listening only when the user deliberately says the wake-up word.

External anonymization via relay: A second option is to disassemble the IoT device and to connect the built-in microphones and the speaker to a relay that is controlled by the anonymization device. Thus, the anonymization device can protect against cases where the voice assistent is activated without intention of the user. Furthermore, the input and the output of the IoT device can be muted at any time with certainty. Thus, it is possible to, say, let the IoT device read dummy calendar entries or play dummy radio stations while the user is in the room.

User Experience: In comparison to the first variant, the implementation efforts of this approach are significantly increased. Because the input and the output of the IoT device can be externally controlled, the user experience is slightly higher.

Anonymity: The fact, that it is possible to mute the output while the user is in the room, allows for more options to issue dummy commands to the voice assistant. Ensuring that the IoT device is only listening when allowed increases the privacy of the user. This means that the anonymization device also needs to listen to the wake-up word. However, this can be realized locally, without having to send voice recordings into the cloud [17].

Re-wiring the IoT device: From the perspective of the IoT device, the most impacting approach is to re-wire the speakers and microphones directly to the anonymization device. Thus, any input or output is supervised. Verbal commands to the voice assistant are synthesized at the anonymization device and sent directly, without using speakers and microphones, to the IoT device. A similar approach would be to install the voice-assistant application on a standard Windows PC and to re-route the audio-drivers to a program that handles anonymization.

User Experience: In comparison to the other approaches, it is possible to control the input and output of the IoT device in a fine-grained way. However, the implementation effort of this variant is high. It requires expert knowledge to set up an anonymization device that sends synthesized voice commands to the voice assistant without loss of user experience.

Anonymity: This approach ensures that no information is sent unfiltered to the voice-assistant provider. It would be possible even to let the anonymization device synthesize verbal requests from the user and a database of dummy commands with the same artificial voice, i.e., the provider does not have an option to distinguish various speakers according to certain verbal characteristics. However, as with the other approaches, a fraction of the requests sent is still the real interest of the user. Thus, it remains challenging to create a set of dummy requests with exactly the same statistical characteristics as the real requests.

VI. Dummmy Requests for the Echo Dot

In this section, we describe our prototypical anonymization device. It is based on a Raspberry Pi-based that sends dummy requests via speaker to Alexa. Furthermore, we discuss what we have learned during its implementation.

To create a setting that is easily reproducible, we have decided to play various radio stations. In this setting, we have to consider only two voice commands: "‘Alexa, play <radio station> on Tune In"’ and "‘Alexa, stop"’. The skill "‘Tune in"’ provides 41 different genres of music. Every genre is associated with multiple radio stations. A typical voice command is "‘Alexa, play Radio Bob on Tune In"’. Our objective is to anonymize our (a) taste for music and our (b) activity times. Thus, we focus on anonymizing information from our category "‘Application Data"’.

A. Hardware Setup

Our anonymization device is shown in Figure 7. It consists of three components:

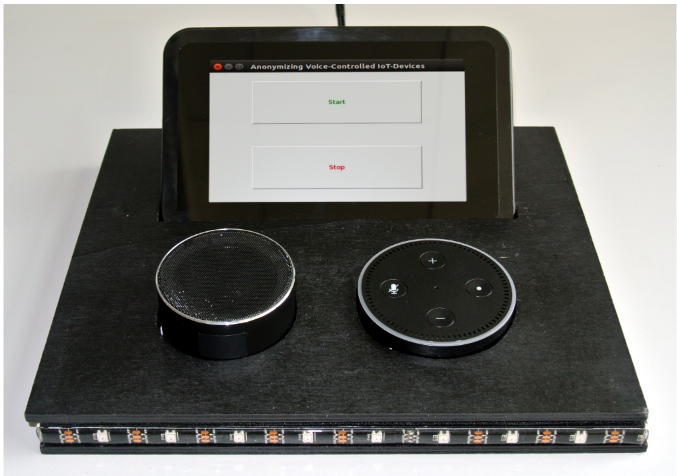

Figure 7. Our anonymization device

1) *Amazon Echo Dot 2nd generation:* This is the lower right device in Figure 7. The Echo hardware complement includes a 64-bit quad-core MEDIATEK ARM MT8163V 1636-KBCAH CCMKYRHS processor, 512 MB of LPDDR3 SDRAM and 4GB of storage space. It connects to the Internet via WiFi 802.11a/b/g/n.
2) *Raspberry Pi 3 Model B Rev 1.2 with 7 inch touch display:* This is the upper device in Figure 7. The Raspberrys operating systems is Raspbian GNU/Linux 8 (Jessie). For this project, we’ve used a Raspberry Pi 3 Model B, which uses a Broadcom BCM2837 SoC with a 1.2 GHz 64-bit quad-core ARM Cortex-A53 processor, 1 GB of LPDDR2-SDRAM and can be connected to the WiFi through 802.11a/b/g/n.
3) *Nubwo Wireless Speaker:* This is the device in the lower left position in Figure 7. This is a wireless bluetooth speaker with integrated microphone. It is compatible with different devices, such as tablets, laptops or smartphones.

B. Software Setup

The Raspbian Linux already comes with all the tools needed to play dummy commands that follow a certain distribution: With the scripting language Python and its toolkits, we have realized a graphical user interface to control the order and the distribution of the dummy voice commands. Python allows to start linux commands to play voice samples in various audio file formats that have been recorded in advance and stored on the local SD card of the Raspberry Pi. Alternatively, we have tested Google’s text-to-speech synthesis "‘Simple Google TTS"’ to generate voice commands from text, that is, without having to record voice samples beforehand. This synthesis can be accessed via Python library "‘gTTS"’.

C. Dummy-based Anonymization

Figure 8 shows our process to generate dummy commands. We have written a Python script for the Raspberry Pi, which plays dummy voice commands. When the user is about to leave the room, he starts this script via command line or a graphical user interface. Subsequently, the script randomly selects a radio station, plays "‘Alexa, play <radio station> on Tune In"’ via speaker to the Dot, waits a random time interval ranging from

10 seconds to 60 minutes, and plays "‘Alexa, stop"’. This procedure recurs, until it is terminated by the user.

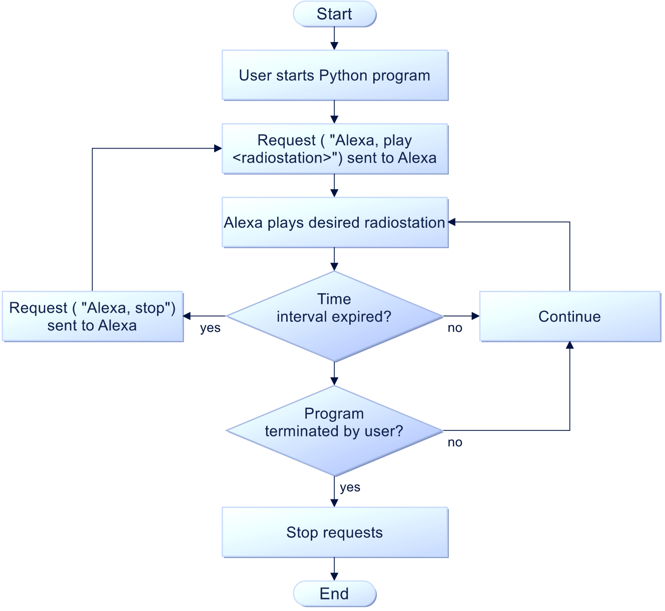
Start
User starts Python program
Request ("Alexa, play <radiostation>") sent to Alexa
Alexa plays desired radiostation
Time interval expired?
yes
no
Request ("Alexa, stop") sent to Alexa
Continue
Program terminated by user?
no
yes
Stop requests
End

Figure 8. Flowchart of the dummy-based anonymization

Switching radio stations after a few seconds corresponds to a user browsing for a station that fits to his mood at the moment. Switching stations randomly with a frequency between some seconds and 60 minutes corresponds to a user having heard songs he dislikes. After 60 minutes, the user might have lost interest in listening to the station. We assume, that this procedure creates realistic dummy commands, which cannot be distinguished from user commands with the algorithms currently used by the service provider.

D. Discussion

We have implemented a scenario that is understandable yet realistic. By anonymizing music taste and activity times when listening radio stations via Amazon's Echo Dot, we have made a number of observations.

It was surprisingly simple to issue dummy voice commands with a certain frequency and distribution by using off-the-shelf hardware and software components. With some experience in Linux, it did not took long to realize our anonymization device. It was not necessary to open the IoT device. Except for the touch-sensitive display used for debugging purposes, we have spent less than 50 EUR for hardware components. However, it needs a profound technical understanding to configure the generation of dummies according to a person's individual privacy requirements.

Alexa's Automatic Speech Recognition technology works very well, even with voice commands that are synthesized from text. For our tests, we have used Simple Google TTS, which is an online service of Google. However, we assume that offline text-to-speech engines like Cepstral [8] or eSpeak [11] would be applicable as well. This way, no external party might learn which dummy commands are synthesized and sent to Alexa.

We have observed that Amazon's recommendations for radio stations follow our dummy commands. This indicates that our anonymization approach is working as intended at the moment. On the other hand, a voice assistant that is based on garbled recommendations might reduce the user experience. Furthermore, it is impossible for the user to find out what the service provider learns indeed. If the provider implements machine-learning algorithms that distinguish different voices, only the rewiring approach from Section V might be able to provide some degree of privacy. Amazon says that voice samples are used to improve speech recognition. However, we did not observe that the the real user's voice was recognized less accurately due to learning from a synthesized voice.

In conclusion, at this moment, even a simple dummy-based anonymization approach allows for some more privacy regarding application data, than the voice-assistant provider is offering the user. We have focused on the user's activity times and his taste for music. In the same way, many other aspects and many other services could be anonymized. For example, it would be possible to anonymize the places of interest that are sent to a service for local weather reports. Similarly, events and holidays could be anonymized with calendar services. However, some services do not allow for this approach. For example, it would not make sense to turn on and off smart home components, such as cleaner robots, or to anonymize the wake-up alarm. Finally, it is impossible for the user to find out or influence which algorithms are implemented by the provider. In the worst case, the provider might implement machine learning to distinguish different personality profiles from voices and statistical properties of the commands, without telling in the privacy policy. Thus, it is impossible to give anonymity guarantees like differential privacy [29].

VII. Conclusion

Currently, voice assistants are integrated into a plethora of different IoT devices and used every day in numerous situations. Voice assistants on IoT devices make users' life more comfortable. On the other hand, such assistants send voice samples into the cloud, which contain private information from private places. Even more, the voice samples do not only transport the verbal commands required, but also adjunct information, such as verbal stress patterns or background noises indicating certain activities. Thus, the service provider is in a position to create deeply impacting personality profiles.

We have identified three categories of information that can be observed by the service provider. Furthermore, we have analyzed from the user perspective at which points it is possible to realize anonymization. We have described a dummy-based anonymization approach, and we have explored it's properties by implementing a prototypical anonymization device based on a Raspberry Pi that feeds dummy commands to an Amazon Echo Dot. Our device confirms that it is possible to anonymize a part of the information sent to Alexa with limited inconvenience for the user.

References

[1] Alexa Skills Kit. https://developer.amazon.com/de/alexa-skills-kit, retrieved: Aug. 2018.

[2] Alexa Voice Service. https://developer.amazon.com/de/alexa-voice-service, retrieved: Aug. 2018.

[3] Amazon Alexa. https://alexa.amazon.com, retrieved: Aug. 2018.

[4] Amazon Alexa Technologies, AWS Stockholm Summit 2017. https://de.slideshare.net/AmazonWebServices/amazon-alexa-technologies, retrieved: Aug. 2018.

[5] Amazon Echo Dot Teardown. https://de.ifixit.com/Teardown/Amazon+Echo+Dot+Teardown/61304, retrieved: Aug. 2018.

[6] AWS Lambda. https://aws.amazon.com/de/lambda/features/, retrieved: Aug. 2018.

[7] BMW is now integrated with the Google Assistant. http://www.bmwblog.com/2017/11/07/bmw-now-integrated-google-assistant/, retrieved: Aug. 2018.

[8] Cepstral. https://www.cepstral.com, retrieved: Aug. 2018.

[9] CES 2018: voice-controlled showers, non-compliant robots and smart toilets. https://www.theguardian.com/technology/2018/jan/12/ces-2018-voice-controlled-showers-robots-smart-toilets-ai, retrieved: Aug. 2018.

[10] Echo Dot (2nd Generation) - Smart speaker with Alexa. https://www.amazon.com/Amazon-Echo-Dot-Portable-Bluetooth-Speaker-with-Alexa-Black/dp/B01DFKC2SO/ref=sr/_1/_1, retrieved: Aug. 2018.

[11] eSpeak. https://espeak.sourceforge.net, retrieved: Aug. 2018.

[12] Gartner Says Worldwide Spending on VPA-Enabled Wireless Speakers Will Top 2 Billion by 2020. https://www.gartner.com/newsroom/id/3464317, retrieved: Aug. 2018.

[13] Home Automation Device Market Grows Briskly, to 27 Million. https://www.voicebot.ai/wp-content/uploads/2017/11/cirp-news-release-2017-11-06-echo-home.pdf, retrieved: Aug. 2018.

[14] Home has a new hub. https://www.samsung.com/us/explore/family-hub-refrigerator/connected-hub/, retrieved: Aug. 2018.

[15] How to Keep Amazon Echo and Google Home From Responding to Your TV. https://www.wired.com/2017/02/keep-amazon-echo-google-home-responding-tv, retrieved: Aug. 2018.

[16] IDC Forecasts Worldwide Spending on the Internet of Things to Reach 772 Billion in 2018. https://www.idc.com/getdoc.jsp?containerId=prUS43295217, retrieved: Aug. 2018.

[17] Jasper. https://jasperproject.github.io, retrieved: Aug. 2018.

[18] Nutzungsbedingungen für Alexa und Alexa-Geräte. https://www.amazon.de/gp/help/customer/display.html?nodeId=201566380, retrieved: Aug. 2018.

[19] Please meet Amazon Alexa and the Alexa Skills Kit. https://de.slideshare.net/AmazonWebServices/please-meet-amazon-alexa-and-the-alexa-skills-kit, retrieved: Aug. 2018.

[20] Using Multiple Alexa Devices. https://www.amazon.com/gp/help/customer/display.html? nodeId=202013740, retrieved: Aug. 2018.

[21] C. C. Aggarwal and S. Y. Philip. A General Survey of Privacy-preserving Data Mining Models and Algorithms. In *Privacy-preserving data mining*, pages 11–52. Springer, 2008.

[22] G. Aggarwal, T. Feder, K. Kenthapadi, R. Motwani, R. Panigrahy, D. Thomas, and A. Zhu. Anonymizing Tables. In *International Conference on Database Theory*, pages 246–258. Springer, 2005.

[23] E. Alepis and C. Patsakis. Monkey Says, Monkey Does: Security and Privacy on Voice Assistants. *IEEE Access*, 5:17841–17851, 2017.

[24] J. L. Boyles, A. Smith, and M. Madden. Privacy and Data Management on Mobile Devices. *Pew Internet & American Life Project*, 4, 2012.

[25] S. Chester, B. Kapron, G. Ramesh, G. Srivastava, A. Thomo, and S. Venkatesh. Why Waldo befriended the Dummy? k-Anonymization of Social Networks with Pseudo-Nodes. *Social Network Analysis and Mining*, 3(3):381–399, 2013.

[26] H. Chung, M. Iorga, J. Voas, and S. Lee. Alexa, Can I Trust You? *IEEE COMPUTER SOCIETY*, 50(9):100–104, 2017.

[27] W. Diao, X. Liu, Z. Zhou, and K. Zhang. Your Voice Assistant is Mine: How to Abuse Speakers to Steal Information and Control Your Phone. In *Proceedings of the 4th ACM Workshop on Security and Privacy in Smartphones & Mobile Devices*, pages 63–74. ACM, 2014.

[28] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The Second-Generation Onion Router. In *Usenix Security*, 2004.

[29] C. Dwork. Differential Privacy: A Survey of Results. In *International Conference on Theory and Applications of Models of Computation*, pages 1–19. Springer, 2008.

[30] M. Ebling. Can Cognitive Assistants Disappear? *IEEE Pervasive Computing*, 15(3):4–6, 2016.

[31] L. Gong and C. Nass. When a Talking-Face Computer Agent is Half-Human and Half-Humanoid: Human Identity and Consistency Preference. *Human Communication Research*, 33(2):163–193, 2007.

[32] Q. Han, H. Zhao, Z. Ma, K. Zhang, and H. Pan. Protecting Location Privacy Based on Historical Users over Road Networks. In *International Conference on Wireless Algorithms, Systems, and Applications*, pages 347–355. Springer, 2014.

[33] T. Hara, A. Suzuki, M. Iwata, Y. Arase, and X. Xie. Dummy-based User Location Anonymization Under Real-world Constraints. *IEEE Access*, 4:673–687, 2016.

[34] C. Jackson and A. Orebaugh. A study of Security and Privacy Issues Associated with the Amazon Echo. *International Journal of Internet of Things and Cyber-Assurance*, 1(1), 2018.

[35] R. Kato, M. Iwata, T. Hara, A. Suzuki, X. Xie, Y. Arase, and S. Nishio. A dummy-based Anonymization Method Based on User Trajectory with Pauses. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, pages 249–258. ACM, 2012.

[36] H. Kido, Y. Yanagisawa, and T. Satoh. An Anonymous Communication Technique Using Dummies for Location-based Services. In *Pervasive Services, 2005. ICPS'05. Proceedings. International Conference on*, pages 88–97. IEEE, 2005.

[37] A. Kumar, A. Gupta, J. Chan, S. Tucker, B. Hoffmeister, M. Dreyer, S. Peshterliev, A. Gandhe, D. Filiminov, A. Rastrow, C. Monson, and A. Kumar. Just ASK: Building an Architecture for Extensible Self-Service Spoken Language Understanding. *arXiv preprint arXiv:1711.00549*, 2017.

[38] J. Kwapisz, G. Weiss, and S. Moore. Activity Recognition Using Cell Phone Accelerometers. *ACM SigKDD Explorations Newsletter*, 12(2):74–82, 2011.

[39] T. Lackorzynski and S. Koepsell. Hello Barbie - Hacker Toys in a World of Linked Devices. *Broadband Coverage in Germany; 11. ITG-Symposium*, 2017.

[40] A. E. Moorthy and K. L. Vu. Voice Activated Personal Assistant: Acceptability of Use in the Public Space. In *International Conference on Human Interface and the Management of Information*, pages 324–334. Springer, 2014.

[41] C. Ratti, D. Frenchman, R. M. Pulselli, and S. Williams. Mobile Landscapes: Using Location Data from Cell Phones for Urban Analysis. *Environment and Planning B: Planning and Design*, 33(5):727–748, 2006.

[42] S. Shakila and G. Ganapathy. Privacy for Interactive Web Browsing: A Study on Anonymous Communication Protocols. *International Journal*, 2(5), 2014.

[43] C. Spensky, J. Stewart, A. Yerukhimovich, R. Shay, A. Trachtenberg, R. Housley, and R. Cunningham. SoK: Privacy on Mobile Devices - Its Complicated. *Proceedings on Privacy Enhancing Technologies*, 2016(3):96–116, 2016.

[44] J. P. Timpanaro, T. Cholez, I. Chrisment, and O. Festor. Evaluation of the Anonymous I2P Network's Design Choices Against Performance and Security. In *Information Systems Security and Privacy (ICISSP), 2015 International Conference on*, pages 1–10. IEEE, 2015.

[45] H. Wang, F. Calabrese, G. Di Lorenzo, and C. Ratti. Transportation Mode Inference from Anonymized and Aggregated Mobile Phone Call Detail Records. In *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, pages 318–323. IEEE, 2010.

[46] C. Wueest. A Guide to the Security of Voice-activated Smart Speakers - An ISTR Special Report (2017). https://www.symantec.com/content/dam/symantec/docs/security-center/white-papers/istr-security-voice-activated-smart-speakers-en.pdf, retrieved: Aug. 2018.