Complex Systems Informatics and Modeling Quarterly (CSIMQ)
eISSN: 2255-9922
Published online by RTU Press, https://csimq-journals.rtu.lv
Article 139, Issue 24, September/October 2020, Pages 15–30
https://doi.org/10.7250/csimq.2020-24.02

Smart Home Voice Assistants: A Literature Survey of User Privacy and Security Vulnerabilities

Khairunisa Sharif and Bastian Tenbergen*

Department of Computer Science, State University of New York at Oswego,
Oswego, NY 13126, United States

ksharif@oswego.edu, bastian.tenbergen@oswego.edu

Abstract. Intelligent voice assistants are internet-connected devices, which listen to their environment and react to spoken user commands in order to retrieve information from the internet, control appliances in the household, or notify the user of incoming messages, reminders, and the like. With their increasing ubiquity in smart homes, their application seems only limited by the imagination of developers, who connect these off-the-shelf devices to existing apps, online services, or appliances. However, since their inherent nature is to observe the user in their home, their ubiquity also raises concern of security and user privacy. To justify the trust placed into the devices, the devices must be secure from unauthorized access and the back-end infrastructure tasked with speech-to-text analysis, command interpretation, and connection to other services and appliances must maintain confidentiality of data. To investigate existing possible vulnerabilities, approaches to mitigate them, as well as general considerations in this emerging field, we supplement the findings of a recent study with results from a systematic literature review. We were able to compile a list of six main types of user privacy vulnerabilities, partially confirming previous findings, but also finding additional issues. We discuss these vulnerabilities, their associated attack vectors, and possible mitigations users can take to protect themselves.

Keywords: Intelligent Voice Assistants, Virtual Assistants, Smart Home, Privacy, Security, Systematic Literature Review, Alexa, Cortana, Siri, Amazon Echo, Google Assistant, Apple Homepod.

1 Introduction

Intelligent voice assistants (IVAs) are conquering households. IVAs seemingly bridge the gap to the verbally responsive computers in works of science fiction, as they listen to the user's spoken command and do as commanded. IVAs are Internet of Things (IoT) devices [1], which much like embedded systems observe the world through sensors (e.g., external ones connected as home

* Corresponding author

Reference: K. Sharif and B. Tenbergen, "Smart Home Voice Assistants: A Literature Survey of User Privacy and Security Vulnerabilities," Complex Systems Informatics and Modeling Quarterly, CSIMQ, no. 24, pp. 15–30, 2020. Available: https://doi.org/10.7250/csimq.2020-24.02

Additional information. Author's ORCID iD: B. Tenbergen – https://orcid.org/0000-0002-0145-4800. PII S225599222000139X. Received: 2 February 2020. Accepted: 8 October 2020. Available online: 31 October 2020.

appliances, like IoT cameras, or through device-internal microphones). They transmit sensor data to some cloud infrastructure for analysis and command interpretation, and – depending on the user command – either communicate with other online services (such as calendars, message systems, weather information systems, etc.) or interact with actuators and smart appliances (e.g., smart light bulbs, or other IoT household devices). IVA devices are small and aesthetically pleasing to unobtrusively integrate into the user's home. Sometimes referred to as Smart Assistant, Intelligent Personal Assistant, Digital Assistant, Personal Virtual Assistant, Intelligent Agent, Virtual Assistant Bot, etc. [2], concrete examples include Apple's Siri, Microsoft's Cortana, Amazon's Alexa, and "OK, Google."

IVAs assist the user in mundane everyday tasks, increase comfort, and may even pose accessibility opportunities for special needs users. They are also deeply integrated into the vendors' respective business platforms to enable access to services or sales of other products. It is fair to say, IVAs are very personal devices with access to a wealth of potentially confidential user information. This warrants a close look at their potential security vulnerabilities, privacy concerns, and risks to the user. In a recent study [3], Edu, Such, and Suarez-Tangil investigated this issue. The authors found that a vast array of attack vectors and privacy concerns exist. Among the most severe impairments, the authors identified the physical access to IVAs, the technical platforms governing the IVA front-end, and insufficient encryption hardening of authentication and authorization. Edu et al.'s contribution takes a very technical approach and highlights the impacts for the user, but does not provide an account of what countermeasures users can take to protect themselves against privacy vulnerabilities.

In this article, we supplement Edu et al.'s work with findings of a systematic literature search and a review of the state of the art into this issue. We investigated the following research questions:

- *RQ1: What IVA vulnerabilities exist that threaten user privacy?*
- *RQ2: What technical or user-centric countermeasures can be taken to mitigate these vulnerabilities?*

We contrast and highlight our findings with those reported in [3] and supplement results and primary studies not previously considered. The remainder of this article is structured as follows. Section 2 presents background on IVAs and their security flaws as presented in IT press, industry accounts, and scholarly work. Section 3 discusses study selection approach. Section 4 discusses results from the review of the identified studies. Section 5 compares our findings against those in [3] and concludes this article.

2 Background and Related Work

Advances in natural-language processing and ubiquitous high-performance networks have paved the way for intelligent voice assistants (IVAs) to be the center of the smart home revolution ever since the early 2000s [4]. With applications ranging from eGovernment [5], to accessibility support for humans with disabilities [6], [7], to simple convenience, IVAs were present in more than 41% of homes in the United States alone; generating a projected revenue of up to $19 billion [8]. In fact, the average Google search interest for the term has quadrupled since 2013, after having experienced minor spikes in 2010 and 2012, respectively [9]. These spikes roughly coincide with the release of Apple's Siri as a standalone app in 2010 and its subsequent bundling into iOS in 2011. The increase in global interest further coincides with the release of Microsoft Cortana in 2013, Amazon's Alexa in 2014, and Google's Assistant in 2016.

IVAs operate through an agent/client software on a suitable device, typically provided by the same vendor as the IVA software. Voice commands and instructions are captured through internal microphones of the device and sent to a cloud natural language processing platform, which converts the voice recording to machine-interpretable data, performs linguistic analysis, and then retrieves additional information from other services or sensors connected to the IVA,

depending on the voice command entered by the user. IVA functionality is hence mainly driven by proprietary cloud infrastructure; the devices themselves are but a physical font-end for the software features.

To activate, these devices perpetually listen for a "wake-word," (e.g., "Hey Siri," "OK, Google," or simply the name of the product). One natural concern becomes what else these devices listen to and what conversations and noises in the household are recorded in general. Recently, not only the IT press [10], [11], but also industrial vendors themselves [12], [13], and academia [14], [15] have reported how IVAs record entire conversations of anyone in the room, citing that up to 80% of IVA users are concerned about possible breaches of their privacy. This concern is wide-spread, as law enforcement agencies discourage users of certain models of smart TVs with bundled IVA functionality from discussing confidential information in the vicinity of the device [16]. This is particularly concerning in jurisdictions, where interacting with IVAs implies that the users legally relinquish their right to reasonable privacy (as is the case, for instance, due to the third-party doctrine exception to the 4th Amendment of the United States of America, see [17]).

IVAs do not only pose privacy issues for the user through their innate functionality, but also present physical attack vectors. For instance, lasers can be used to activate the devices and allow injecting simulated voice commands into the device's internal microphone [18]. Similarly, researchers have found ways to bypass device passcode restrictions on smart phones and inject voice commands into the voice assistant software using "guided ultrasonic waves," which allowed them to place unauthorized phone calls on the target device [19]. Other attacks may target the cloud ecosystem of the IVA [3].

While past studies, vendor investigations, and IT press seems to focus on hardware attacks, the consequences from a user perspective seem to be relegated to an implicit circumstance. In this article, we therefore place emphasis on the user perspective.

3 Study Selection Approach

To investigate the user privacy and security vulnerabilities of IVAs, we have conducted a systematic literature review (SLR) following the guidelines suggested in [20]. In contrast to [3], we present a detailed account on our search and study selection procedure to foster replication.

3.1 Study Purpose and Research Questions

The purpose in performing this SLR is to understand and become aware of the vulnerabilities that threaten the security and privacy of intelligent voice assistants and the strategies and techniques that are available in the field of cybersecurity engineering that can be used to mitigate the attacks. The aim is to answer the following research questions presented in Table 1.

Table 1. Research questions of this systematic literature review

ID	Research Question	Description and Motivation
RQ1	What IVA vulnerabilities exist that threaten user privacy?	This RQ identifies classes of vulnerabilities and common threats specific to users and the user experience, as reported in the literature. We hope that this information allows industry and academia to create a more trustworthy user experience.
RQ2	What technical or user-centric countermeasures can be taken to mitigate these vulnerabilities?	This RQ provides an overview of countermeasures that users can take right now to reduce their risks from exposure to the vulnerabilities from RQ1 and increase trustworthiness in their IVAs. We hope that this information is also useful for academia and industry to develop robust services and appliances that interact with IVAs.

3.2 Search Strategy and Search Strings

A manual search was conducted using the search engines of publishers and indexing services which typically include computer science literature. These include ACM Digital Library, Google Scholar, IEEE Xplore, and Science Direct. Since Apple's Siri was the first IVA product to come to market in 2010 and since the literature search was conducted in August 2019, we selected studies primarily from this time period.

The search strings are presented in the Appendix. We started out with SS1 (original search string), which was designed to accommodate both hardware and software product names. We selected the term "intelligent virtual assistant" and suitable synonyms as a boarder term in order to capture as many candidate studies as possible. The first part of the search string is comprised of key words related to RQ1, while the second part of the search string is comprised of key terms related to RQ2.

Character limits imposed by some of the search engines required modifying the search string. SS1 was broken down into different combinations while still making sure to include all the terms and also meeting the required character count. Different variations of the SS1 were used to produce modified search strings some of which are presented in the Appendix.

3.3 Inclusion and Exclusion Criteria

Due to the proprietary nature of many available IVAs, we anticipated that candidate studies would likely either consider generic aspects of IVAs, or target specific IVA products (e.g., Google Assistant). Since RQ1 and RQ2 aim at privacy concerns for IVAs at large, we specifically aimed to cast a wide net and select as many candidate studies as possible, notwithstanding the specific IVA or product that was the target of a specific investigation. We therefore define ***inclusion criteria*** as follows:

- Studies that consider specific IVA services, e.g., Amazon Alexa or Microsoft Cortana;
- Studies that consider specific IVA devices, e.g., Apple Homepod or Google Assistant;
- Studies that considered IVA hardware or software in general; and
- Studies that fall into the study selection period (see Section 3.2).

Nevertheless, to limit the results to studies pertaining specifically to IVAs, we exclude studies that pertain to ubiquitous computing at large. Among others, this includes smart TVs, wearable devices, and always-on devices not otherwise considered an IVA. Moreover, we excluded studies on chat-bot type "virtual assistants." Furthermore, in contrast to [3], we excluded studies pertaining to hardware-only attacks on the IVA devices (i.e., attacks which require physical access to the internal electronics of the device), as we considered this to be within the domain of electrical and computer engineering rather than cybersecurity engineering. The complete list of ***exclusion criteria*** is as follows:

- Studies on ubiquitous computing, wearable devices, or always-on devices that are not IVAs;
- Studies on virtual assistants accessed by online chat such as chatbots;
- Studies that pertain to IVA device hardware;
- Non-computer science literature;
- Non-peer reviewed literature; and
- Articles unavailable to the authors due to paywall restrictions, which could not be resolved using interlibrary loan or other means.

We aimed to keep the resulting set of studies small, refraining from including those already reported in [3], however also not strictly excluding these candidates, such that we are able to provide a complementary perspective.

3.4 Study Selection Procedure

The search was conducted in August 2019. For each search string, the first 100 result pages of each search engine (using the respective default setting for number of results per page) were considered. We discontinued the search using a search engine after 50 articles were encountered that were previously found. This yielded a total of 19,288 candidate articles (16,800 Google Scholar; 1,722 ACM Digital Library; 587 IEEE Xplorer; 179 SpringerLink).

We filtered the results following the procedure in [19] by applying the inclusion and exclusion criteria from Section 3.3. Search engine result pages were manually parsed and studies were selected based on their titles. This resulted in the inclusion of 61 candidate studies. Afterwards, papers were obtained and abstracts were read. This resulted in excluding 14 papers, yielding remaining 47 candidate studies. These 47 papers were then read entirely and further filtered based on inclusion and exclusion criteria. This led to the exclusion of 35 additional candidate studies. The remaining 12 papers were considered the final study amount and were retained for data extraction. After each filtering step, the first and second author discussed the inclusion/exclusion rationale for each individual candidate article in case of disagreement or uncertainty, the article was included for consideration in the next filter step. Selection procedure and the results after each step are summarized in Figure 1.

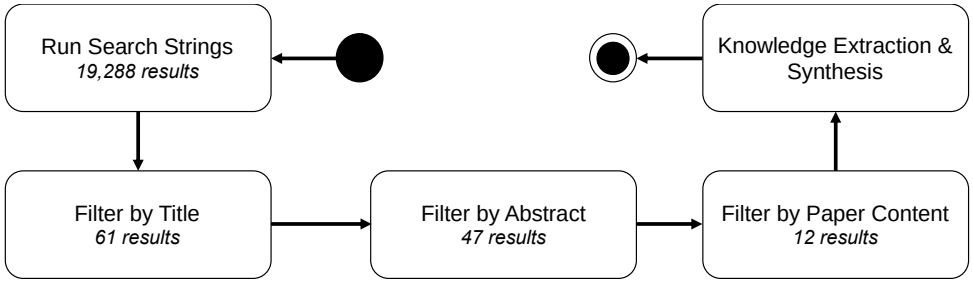

Figure 1. Study selection procedure and filter results

4 Results

In the following, we will discuss the findings from these papers on the basis of our research questions from Section 3.1 by contrasting them to the findings and studies reported in [3].

4.1 RQ1: What IVA Vulnerabilities Exist that Threaten User Privacy?

From the 12 included studies, we were able to identify the following main vulnerabilities of IVA: always listening, weak authentication, replay attacks, cloud infrastructure, aftermarket upgrades, and integration of IoT devices. In the following, we will discuss (1) the details of these vulnerabilities and (2) the attack scenarios outlined in the included studies.

4.1.1 Always Listening

In [2], it was discovered that permanently observing sound through the integrated microphones of the IVA device creates a possible infringement of user's personal privacy. The authors mention that although the IVA-enabled device only records the user's voice and transmits the recording to the cloud only when the wake-word is uttered, the device is still perpetually observing the conversations and typical noises around the device and some devices have been found to store their recording. If a malicious attacker gains access to a compromised enabled IVA device, all the recorded sounds or voices can be sent to the attacker in real-time.

Perpetually recording the sounds surrounding the IVA device also allows for non-attack-based intrusion. As reported in [21], even though Amazon, Apple, Google, and Microsoft claim that their devices only record when the users say the command to wake the assistant, there has been

at least one incident of the device recording and sending the recordings back to the vendor at times when the user did not use the wake word. In such cases, it becomes easy for the IVA vendor to analyze the user's conversations, and create a profile of the user's typical daily activities by means of analysis of household noises. It may even be possible to correlate the user's location with IP address information and geolocation.

Ford & Palmer present similar evidence in [17], where they found that Amazon Echo Dot devices were found to record and stream private conversations to Amazon for analysis without utilizing the wake-word. Using two identical devices, in one of which the authors disabled the internal microphone, the study identified inconsistencies between the Alexa application logs and identified Amazon Voice Service network traffic. Even when the wake-word has not been used, the authors found the presence of human voice triggered inadvertent recordings. Specifically, 61.5% of the recordings were triggered by TV audio and 38.4% by human conversation. From this data, it can be concluded that recording and transmission of possibly sensitive conversations within the user's home is possible even without the use of the wake word. The authors in [17] did not find evidence of private conversations having been recorded and transmitted in the device where the microphone was turned off.

This finding is confirmed in [3]. Therein, it is noted that the very nature of permanently surveilling the vicinity for the wake word may cause accidental utterances of said wake word to record the entire conversation. Edu et al. [3] remark the loss of control over voice data for the user.

4.1.2 Weak Authentication

A number of studies [2], [21] found that IVAs lack the proper ability to determine if it is in fact the owner or another authorized party that has uttered the wake word. Anyone who has access to a voice-activated device can ask it questions and gather information about accounts services affiliated with the device. A malicious attacker that comes in close range of a targeted IVA device can potentially fool the system into believing that the rightful owner is speaking [2], allowing the attacker to access calendar details, emails, and other personal information. An additional concern for Amazon Alexa is that it is built into Amazon's store interface which by default grants anyone with voice access to the device the ability to order items from the user's Amazon account (albeit safeguards such as shipment address confirmation and voice passcode are available, see [21]). Furthermore, even though the main user is able to designate certain access controls for secondary users, they still have the ability to modify the device set-up such as network connection, sound, and other features without the primary consent of the user.

Edu et al. [3] add to this finding that the core of this issue is two-fold: on the one hand, the device listens to any utterance of the wake word, regardless if the wake word was uttered by an authorized user. Moreover, the issue is emphasized by the fact that IVAs do not typically contain checks against synthesized speech and that the lack of appropriate functional role of separation inhibits users from correctly defining what and how resources should be accessed.

4.1.3 Replay Attacks

Some studies conclude that a consequence of IVA devices' weak authentication (see Section 4.1.2) means that synthesized speech imitating a legitimate user render the IVA device vulnerable to replay attacks [22]. Replay attacks can be achieved by recording authorized users or synthesizing a reasonable approximation of their voice fingerprint. Recent advances in voice synthesis techniques have made it easier for attackers to launch such attacks without being detected [9]. Moreover, inaudible signals can be embedded into the audio signal of a TV or radio broadcast to carry out an attack on numerous targets simultaneously, much akin to the aforementioned laser beam attack [18]. An example of such an attack taking place is the case of a fast food restaurant chain's TV ad which caused Google Home devices at the viewers' homes to read out loud information from Wikipedia about one of the restaurant chain's products [23]. Albeit this instance was merely a marketing gag, this incident shows that replay attacks can result in serious repercussions when interacting with appliances in the user's home, (e.g.,

opening a smart lock when no one is at home, sending messages in the user's name, or placing a call to premium numbers without one's knowledge, and using a voice-first device to start without the owner's knowledge, see [22]).

In [3], Dolphin attacks are highlighted as a variation of a replay attack. Dolphin attacks could occur when commands are delivered through ultrasonic frequencies, much like the SurfingAttack presented in [19]. Dolphin attacks exploit the vulnerability of the off-the-shelf microphone components used in some IVA devices, which are able to capture frequencies outside of the audible range, allowing an attacker to gain control of IVA devices. However, since these attacks make use of synthesized high-frequency sounds, the required close proximity of the attacker to the target device and the required special hardware make it difficult to perform in a real-world setting, according to [3]. Yet, as [19] shows, these attacks can be used to take control over the user's device and perform unauthorized actions, such as placing phone calls, and thereby possibly submit personal data to a hacker-controlled medium.

4.1.4 Cloud Infrastructure Vulnerabilities

Considering the strong dependence of IVAs on cloud infrastructure and third-party services, it is unsurprising that some studies found vulnerabilities [2], [24]. While typical attacks on service-oriented architectures (e.g., man-in-the-middle attacks or sniffing) are by nature also a vulnerability for IVAs, in contrast to [3], we consider such vectors to be part of the network infrastructure. However, specific to user privacy is the notion that this issue can be exacerbated by IVAs wireless connections. As presented in [2], HTTPS interception tools can be used to analyze requests and responses in order to understand which APIs were used for sending and receiving data to and from Amazon Alexa operating in the cloud. The analysis showed that although most network traffic is encrypted, not all information is sent over a secure protocol. Unencrypted connections including checking the current network connectivity status, transmitting the firmware image upgrades, etc. may still be present, allowing personal user data to be leaked or intercepted.

Apthorpe et. al [24] found privacy vulnerabilities in the exchange of information between the IVA device and the IVA cloud provider when they conducted a passive analysis of encrypted smart home traffic. User interactions with the Amazon Echo device were profiles by plotting send/receive rates of stream even with encrypted traffic. The study demonstrates that encryption alone does not provide all the necessary privacy protection requirements because an attacker can use the data obtained to infer a user's lifestyle and determine the best time to launch an attack that will successfully go undetected. The method used in the study may not be applicable to situations where different IoT devices communicate with the same domain due to the challenges in labeling streams by device type [3].

As Edu et al. point out, the IVA records, pools, and accesses large amounts of data and makes it centrally available in a single point, which would provide an attacker access to valuable and sensitive information [3]. Data can be accessed from multiple web-based or app-based platforms, which also broadens the attack surface. However, as we have shown above, the single-point access to cloud data accessible by weakly authenticated wake words allow attackers to create user profiles, potentially assuming their identity as well.

4.1.5 Aftermarket Upgrades

Many IVA devices allow users to add additional features. Sometimes referred to as "skills," it allows expanding IVA functionality by interfacing with other programs which users can invoke through voice commands. In fact, using some web services allow users to craft skills which automate certain task (e.g., posting to social media, or toggling appliances on and off, see [21]). This makes the IVA vulnerable to attacks. Aside from hence incorporating possible vulnerabilities in the web service used to create the skill, it also authorizes the skill to access confidential information which may result in accidentally disclosing sensitive information to third parties.

For example, Zhang et. al [25] conducted a study where they exploit adversarial NLP vulnerabilities to launch an attack by taking advantage of the way skills are invoked. Two basic threats in Amazon's Alexa and Google Assistant voice services including voice squatting and voice masquerading were analyzed. Voice squatting attacks exploit the weakness in the skill's invocation method. It permits an attacker to use a malicious skill with the longest matching skill name, similar phonemes, or paraphrased name to hijack the voice command of another skill. On the other hand, voice masquerading targets user's misconceptions about how the skill service interacts with the IVA, leading the user to believe in a false set of features, when the skill in fact is exploitative in nature. The authors successfully hijacked the skill for over 50% of the five randomly sampled vulnerable target skills. Alexa is more vulnerable to these types of attacks because multiple skills with the same invocation name are permitted. In the voice masquerading attack, a malicious skill is used to invoke another malicious skill. The malicious skill then proceeds to record the user's voice commands, allowing an attacker to eavesdrop on the user's conversations or leak sensitive information.

This major concern was also identified in [3]. The authors note that skill-based extensions of IVA functionality might pose an attack vector by exploiting poor enforcement of development policies and report on missing vetting processes regarding authorization to interface with the cloud infrastructure. One could argue that appropriate user education on the threats inherent to these skills may also assist in increasing trust in this IVA feature.

4.1.6 Integration of IoT Devices

Some studies consider vulnerabilities stemming from the integration of IoT devices with the IVA device [26], [27]. Such integration typically surrounds smart home devices (e.g., Nest, or Ecobee 4), smart security devices (e.g., Scout, or Abode), smart lighting devices (e.g., Philip Hue, or LIFX), household appliances (e.g., GE+ Geneva), and surveillance cameras (e.g., Cloud Cam, Netgear Arlo Q). Users are able to control their smart home devices with voice by speaking the instruction to the IVA, which in turn instructs the smart device to perform the task necessary. The instruction is delivered to the smart device through the IVA client software to the cloud provider, the skills services, and the smart device cloud. The integration unifies the smart home into a single verbally controlled system and allows the IVA to oversee the services of other connected smart devices. In doing so, the integration also creates a single-point vector which attackers can exploit to gain control or circumvent home security devices or even spy on users.

A study by Sivaraman et. al [26] examined the operation of smart home appliances in order to uncover the associated security and privacy concerns. The authors argue that the implementation of security and practices is going to be different for each device because it depends on factors such as device capabilities, mode of operation, and manufacturer. Five smart-home devices were examined including The Philips Hue connected bulb, The Belkin WeMo motion sensor and switch kit, the Nest smoke alarm, the Withings Smart Baby Monitor, and the Withings Smart Body Analyzer. The authors found that the aforementioned smart home devices lack the proper security implementation, and therefore are not only vulnerable to passive eavesdroppers. Italso becomes easier for an attacker to actively capture information, impersonate legitimate users, or launch man-in-the-middle attacks. They believe that such security/privacy issues are not exclusive to the five smart home devices, but in fact are prevalent across other IoT devices that are sold on the market.

A study by Furfaro et. al [27] presented three IoT scenarios in order to investigate vulnerabilities of smart objects. The scenarios were created using SmallWorld which is a software platform that has been arranged to support the assessment, teaching, and learning of security related issues in numerous domains. The platform is based on what is described as state-of-the-art virtualization and cloud technologies for recreating in a realistic setting, a hybrid environment in which many distributed computer systems can be deployed and interact with real life users, software, and hardware. One of the target devices presented in the experiment was the Video Surveillance System, where the goal was to access sensitive information. The aim of the presented scenario was to demonstrate not only how a user's privacy can be violated, but also

illustrate how an attacker can acquire access to sensitive information and potentially harm the victim. An infected smartphone connected to the home network was used to scan the network in order to find other IoT devices and gather information of other devices connected such as the device model and firmware version. This information is then processed by a malicious Command and Control (c&c) server in order to find vulnerabilities which can be exploited. When the network is attacked, the attacker is able to direct the infected device to send spoofed Address Resolution Protocol (ARP) messages. The goal is to associate the smartphone MAC address with the IP address of the default gateway and direct the traffic on the network to be sent to the attacker; this way the attacker is able to inspect packets and gather information without being detected by sending the traffic to the actual default gateway. Once the software agent is implanted on the personal computer to access the surveillance system from the web interface, the credentials are sent over the network without https encryption, thus permitting the bad gateway to access the credentials and granting the attacker access to the surveillance system. The configuration of the surveillance system can be changed by the attacker, so that it can be accessed from the Internet. The authors conclude that security threats related to IoT in a smart home scenario is due to the malware ability to access the home network which the devices are connected to.

This concern was also mentioned in [3]. The authors list a variety of additional issues, which include data acquisition, accumulation, and profiling, privilege escalation on connected devices and services, as well as intercepting data exchanged between devices.

4.2 RQ2: What Technical or User-Centric Countermeasures can be Taken to Mitigate these Vulnerabilities?

In this section, we discuss possible countermeasures to the vulnerabilities discussed in Section 4.1 as outlined by the papers included in our study.

4.2.1 Always Listening

An obvious solution to prevent unwanted recording from intelligent voice assistants is to disable the microphone of the IVA device. Ford & Palmer [17] in their study found that when Echo's microphone was disabled, Alexa would not record and send the recording to Amazon Services for processing (see Section 4.1.1). However, this procedure obviously limits the user from using the IVA's hand free features such as playing music, controlling the home environment or other features that it was purchased for. The authors suggest that users need to find their own balance between privacy and usability. Ammari et. al [11] suggest that the voice assistant device should also provide a better signifier that would let the user know when the device is muted. The device can display a significantly different color or icon that would better inform the user that the microphone of the device is indeed muted. In addition, the device should also provide the user with some indication of when the device is interacting with the cloud service.

4.2.2 Weak Authentication

A countermeasure that has been presented to help with weak authentication is voice authentication [22]. Google and Amazon have implemented speaker verification using voice authentication known as Voice Match and Voice Profiles. The IVA devices Google Home and Amazon Echo are not readily equipped with the mechanisms and problematically leave it to the user to install, configure, and use the mechanism. Apple is also training Siri to become familiar with the user's voice in order to correctly identify it. However, the authors in [22] could not conclude whether the feature has already been released. While these mechanisms can be helpful, an attacker can still potentially use as synthesized voice sample of the legitimate user [28]. In addition, since human voice is open to the public, it is easy for an attacker to collect voice samples [28] and a human voice cannot be replaced or changed when compromised unlike passwords [3].

Feng et. al [29] conducted a study in which they proposed another method for voice authentication. The authors presented continuous authentication VAuth system which seeks to ensure that the intelligent voice assistant works only on commands uttered by legitimate users. The system is comprised of a wearable security token that repeatedly correlates the utterances received by the voice assistant with the acquired body-surface vibrations of the legitimate user. The solution was reported to achieve 97% detection accuracy and close 0.1% false positive. It also works regardless of differences in accents, languages, and mobility [3].

Another solution is suggested by Xinyu et. al [30], where the router Wi-Fi technology is used to detect human motions through the channel states of information from the router. The advantage of the proposed solution is that it does not require the user to wear any devices. However, the performance of the system depends on the location of the Wi-Fi devices and the specified parameters for detection. The system is most effective if there is no structural change to the location of where the intelligent voice assistant devices are deployed [3].

4.2.3 Replay Attacks

Pradhan et. al [22] propose a voice replay detection system as a solution to counter replay attacks. The system is wearable-free, privacy preserving, and supports room scale detection. The system is able to detect different types of replay attacks using voice and WiFi features. Inherent differences between live vs. replayed voice and human breathing pattern during speech detected through WiFi are leveraged by the system. Although the system was found to be effective in detecting replay attacks, the authors found that their system can still use some improvements such as increasing the detection range which was currently found to be two meters, diversify the datasets used for training in order to generalize the voice and WiFi models, and also incorporate other physiological signals including heart rate and other biometric measures to better assist in replay detection.

Lavrentyeva et al. [31] used a reduced version of Light Convolutional Neural Network architecture (LCNN) based on the method of the Max-Feature-Map activation (MFM) to investigate different countermeasures to defend against voice replay attacks. The LCNN with Fast Fourier-based features acquired an equal error rate of 7.34% on ASV spoof 2017 dataset in comparison to the spoofing detection method used in [32] which had an error rate of 30.74%. To further evaluate the effectiveness of their approach, the Support Vector Machine classifier was used. They found that their primary system based on systems scores fusion of LCNN (with FFT based features), SVM (i-vector approach), recurrent neural network and conventional neural network (with FFT based features) displayed a lower error rate of 6.73% on their evaluation dataset. Their approach was implemented in the cloud due to the required extensive computational power [3].

4.2.4 Cloud Infrastructure Vulnerabilities

Security threats in a cloud computing environment range from network level to application level threats. In addition, the data that is housed in the cloud is also vulnerable to numerous threats, therefore certain factors such as confidentiality should be considered when purchasing storage services from cloud service providers. A suggested solution to protecting the cloud against external threats is to frequently examine the cloud [33].

Amazon and Google both provide history logs to their users in order to enhance their experience when interacting with their IVAs. Alexa History and the Google Activity dashboard enable users to view transcripts of audio clips, listen to the audio, as well as see Alexa or Google's response, and delete items [11]. However, it is left to the user to check the activity logs to ensure that no malicious activity has occurred. Moreover, this solution only provides post-facto security. In order for this approach to be effective, the information in the activity log must be logged correctly, which may not always be the case. Ford & Palmer [17] concluded that the device Echo Dot does not do so, leaving this vulnerability unmitigated. Both the customer and provider are responsible for securing the cloud. Providers should make sure that they are providing a secure infrastructure which protects the data and applications of their customers. The

customers should also do their part in taking the appropriate measures to protect their own applications by using secure passwords and appropriate authentication measures [34].

4.2.5 Aftermarket Upgrades

Zang et al. present a system in [25] that analyzes the skill's response and user's command in order to detect voice masquerading attacks. The system uses User Intention Classifier (UIC) and a Skills Response Checker (SRC). The SRC semantically analyzes the response from the skill and compares it to the commands from a black-list of malicious skill responses in order to abate malicious responses. The user UIC protects the user matching the meaning of what the user says to the context of the skill that the user is interacting with at the time of the system commands. The skill that is being used by the user and the relation to what the command that the user utters is also considered. Although, the system was found to report an overall detection precision rate of 95.60%, it is difficult to implement a generic UIC due to variation in Natural language-based command and the legitimate user commands [3]. However, this discussion shows that from the perspective of the user, there is very little that can be done to counteract possible privacy vulnerabilities, other than simply refraining to use third-party skills.

4.2.6 Integration of IoT Devices

Sivaraman et. al [26] propose a network level solution for securing an IoT environment such as a smart home. They suggest the use of a "Security Management Provider" (SMP) to exert limited configuration control over the ISP network and/or home router on behalf of the consumer, but also without getting in the way of the data path. The provider can accomplish this at the ISP access switch or in the user's home gateway that control and monitor network operations for each IoT device. To evaluate the effectiveness of their approach, the authors added the service to the Philips Hue light-bulb and the Nest smoke-alarm in a lab setting. The authors wrote a python script which used captured white-list information to construct attack packets that can be played from the Internet and pretend to be an authorized user and gain access to the bulb. At this point, the SMP responded and invoked the network API in order to employ the appropriate access control rules that allow only authorized users to access the bulb which in this case would be the residents of that house. A mobile app which was installed on the user's smartphone notified the SMP of the public IP address which then dynamically programmed it into the home/edge-router's ACL. The same method was applied to improve the privacy of the Nest smoke-alarm installed in the lab.

Furfaro et. al [27] present an approach for securing a smart home environment where they identify the devices that need protection, group them into logical groups, identify critical and non-critical groups, and isolate each group in a separate subnetwork to better monitor the activities among them. In their proposed scenario, the video surveillance system was identified as a critical device and for this reason, it was determined as the device that should have the least interaction with other devices. In addition, other devices that were within the home (included the multimedia systems, tablet, or PCs) were determined to be confined to the home network only. Mobile devices that can connect to other networks such as smartphones were isolated from the home network because they present a potential attack vector. If they are separated from the home network, then they cannot be used to perform a scanning in the whole network in order to find what devices are inside the smart home. The authors suggest that confining intelligent voice assistant devices into a specific isolated network can help secure the devices against malicious attacks. Since IVAs are able to assist in performing tasks such as making purchases, turning on devices in the home, and other tasks that require access to sensitive information. Perhaps one way to ensure the security of such devices is to ensure that they go undetected when an attacker performs a scan for devices [3], [21].

One provision which can be taken to secure the home network is to install a suitable configured firewall [27]. This would prevent external connection to connect to the network and as a result this would help contain the malware inside the network. According to the authors, the issue arises when dealing with smartphones as they typically do not operate within a delimited

zone because they can also connect to the internet through mobile networks. The firewall can therefore be bypassed and a malicious application can send data to the malicious server and get commands that will subsequently execute inside the LAN.

In addition, IVAs usually have an accompanying application through which they can control the IVA. This application can also be vulnerable to attacks [2]. Keeping devices that can be used as an attack vector such as smartphones in a separate network or isolating these devices from IVAs in a network's demilitarized zone may be a potential solution to counteract their security vulnerabilities [27].

5 Conclusion and Future Work

In this article, we presented a systematic literature search according to [20] and a review of the state of the art on user-centered privacy vulnerabilities of intelligent voice assistants (IVAs). IVAs are personal IoT devices located in the user's home which analyze verbal commands and interact with online services of other IoT devices to deliver some value-added service to the user.

From 61 candidate papers, we selected 12 primary studies, and analyzed them with regard to which vulnerabilities exist for IVAs from the perspective of the user (RQ1) and which countermeasures the users of IVAs can take right now to mitigate these vulnerabilities (RQ2). Table 2 shows the six main vulnerabilities we found, summarizes the impact on the user, and compares this finding to similar concerns reported in a previous study [3]. Table 2 furthermore highlights countermeasures users can take right now to limit the impact on their privacy.

Table 2. Identified user privacy vulnerabilities and user-achievable countermeasures

Vulnerability	Impact on User Privacy	Privacy Impact Reported in [3]	Countermeasure
Always Listening	Spontaneous recording of conversations and transmission to vendor, even without uttering wake word allowing access to private conversations.	Loss of voice data control with accidental utterance; leaked confidential conversations.	Deactivate micro-phone and/or wake word feature to a user-acceptable degree.
Weak Authentication	No protection against un-authorized users, allowing access to personal information stored on device or associated with online accounts.	Circumvent security measures; privilege escalation.	-
Replay Attacks	Recorded or synthesized voice allows attackers to assume user identity.	-	Using replay detection mech-anisms; adjusting distance threshold to user.
Cloud Infrastructure	Single-point access to user information allows inter-cepting, leaking, profiling, and identity theft.	Access to sensitive information.	-
Aftermarket Upgrades	Poor authentication of third-party services with compre-hensive data access allow siphoning sensitive user data, and executing malicious code.	Over-privileged skills obscure user-data flow.	Refrain from using skills.
IoT Integration	Single-point access to IoT devices allows attackers to circumvent home security systems or spy on users.	Data acquisition, accumulation, and integration, snooping, and privilege escalation.	Proper protection of home network using encryption and firewalls; network DMZ for IVAs, IoT devices.

Note that while [3] also mentions replay attacks as a possible security vulnerability of IVAs, the authors do not discuss the impact on user privacy for this concern. Missing countermeasures indicate that existing studies do not discuss possible solution for the given vulnerability, suggesting that there is very little the user can do.

As can be seen, the very nature of the functionality IVAs present poses a security and privacy risk to the user. While everyone must answer the question for themselves whether the risks described above outweigh the benefit and convenience, there are some ways that users can take control over their privacy protection for most of the six major vulnerabilities. Nevertheless, for some vulnerabilities, i.e., those pertaining to weak authentication and cloud infrastructure, the user is forced to entrust the mitigation of security vulnerabilities to the IVA vendor. Albeit mechanisms such as voice indexing and fingerprinting, continuous user authentication, or two-step verification may be means that impact the user, these mechanisms must be implemented by the vendor, possibly with additional operating cost of decreased user convenience.

We consider our study to be complementing the comprehensive investigation by Edu et al. [3]. We were able to confirm many of the findings from [3]. However, the main difference to [3] is that our study took a specific look at the repercussions for the user. We therefore do not claim completeness of our findings, particularly due to the low number of primary studies that we considered. We attribute this to the proprietary nature of available IVAs on the market combined with the relative novelty of intelligent virtual agents on the market (which have not seen wide-spread adoption until after 2016), which may have impaired internal validity of our study. To combat this issue, we specifically selected our period of study inclusion and designed our search strings to capture as many related studies as possible. Nevertheless, generalizability may be limited, as this study cannot by definition give a complete picture of possible vulnerabilities. Yet, we believe our conclusions to be sound and believe that these results complement previous reports in a way that helps practitioners and academics in their quest to propose solutions for secure IVAs. Furthermore, in contrast to [3], we specifically outline our literature search method, allowing for replication of our results and enable future investigations into the field.

The central theme of our research is on requirements engineering for autonomous systems. Therefore, our future work will be concerned with investigating the impact of and similarity between security threats in IVAs and other autonomous systems with the aim to provide techniques to ensure security and safety of cloud-connected autonomous systems in early stages of development.

References

[1] D. Bastos, M. Shackleton, and F. El-Moussa, “Internet of things: A survey of technologies and security risks in smart home and city environments,” *Proceedings of Living in the Internet of Things: Cybersecurity of the IoT*, pp. 1–7, 2018. Available: https://doi.org/10.1049/cp.2018.0030

[2] H. Chung, M. Iorga, J. Voas, and S. Lee, “Alexa, can I trust you?” *Computer*, vol. 50, no. 9, pp. 100–104, 2017. Available: https://doi.org/10.1109/MC.2017.3571053

[3] J. S. Edu, J. M. Such, and G. Suarez-Tangil, “Smart Home Personal Assistants: A Security and Privacy Review,” *arXiv preprint arXiv:1903.05593*, 2019.

[4] Citius Minds. The Evolution of Smart Homes. Available: https://www.citiusminds.com/blog/the-evolution-of-smart-homes/. Accessed on Sept. 10, 2020.

[5] E. C. Paraiso and J.-P. Barthès, “A Voice-Enabled Assistant in Multi-Agent System for e-Government Service,” *Proceedings of the Intl. Symposium and School on Advanced Distributed Systems*, pp. 495–503, 2005. Available: https://doi.org/10.1007/11533962_45

[6] M. Jefferson, “Usability of Automatic Speech Recognition Systems for Individuals with Speech Disorders: Past, Present, Future, and a Proposed Model,” *University of Minnesota Digital Conservancy*, 2019. Available: http://hdl.handle.net/11299/202757

[7] N. Friedman, A. Cuadra, R. Patel, S. Azenkot, J. Stein, and W. Ju, “Voice Assistant Strategies and Opportunities for People with Tetraplegia,” *Proceedings of the 21st. Intl. ACM SIGACCESS Conf. on Computers and Accessibility*, pp. 575–577, 2019. Available: https://doi.org/10.1145/3308561.3354605

[8] S. Perez, Smart Speakers Hit Critical Mass in 2018. Available: https://techcrunch.com/2018/12/28/smart-speakers-hit-critical-mass-in-2018/. Accessed on Sept. 10, 2020.

[9] Google Trends 2020. Worldwide relative interest in the search term "voice assistant." Available: https://trends.google.com/trends/explore?date=all&q=voice%20assistant. Accessed on Sept. 10, 2020.
[10] S. Perez, 41% of Voice Assistant Users Have Concerns About Trust and Privacy, Report Finds, 2019. Available: https://techcrunch.com/2018/12/28/smart-speakers-hit-critical-mass-in-2018/. Accessed on Sept. 10, 2020.
[11] T. Ammari, J. Kaye, J. Y. Tsai, and F. Bentley, "Music, Search, and IoT: How People (Really) Use Voice Assistants," *ACM Transactions on Computer-Human Interaction*, vol. 26, no. 3, pp. 1–28, 2019. Available: https://doi.org/10.1145/3311956
[12] Microsoft, Inc. (2019). Voice Report. From Answers to Action: Customer Adoption of Voice Technology and Digital Assistants. White Paper. Available: https://about.ads.microsoft.com/en-us/insights/2019-voice-report. Accessed on Sept. 10, 2020.
[13] C. Fisher, Amazon Enlists 30 Companies to Improve How Voice Assistants Works Together. Web resource, 2019. Available: https://www.engadget.com/2019-09-24-amazon-voice-interoperability-initiative.html. Accessed on Sept. 10, 2020.
[14] S. Lafia, J. Xiao, T. Hervey, and W. Kuhn, "Talk of the Town: Discovering Open Public Data via Voice Assistants," *Proceedings of the 14th Intl. Conf on Spatial Information Theory*, pp. 10:1–10:7, 2019.
[15] E. Alepis and C. Patsakis, "Monkey says, monkey does: security and privacy on voice assistants," *IEEE Access*, vol. 5, pp. 17841–17851, 2017. Available: https://doi.org/10.1109/ACCESS.2017.2747626
[16] A. Holmes, The FBI just issued a warning about the risks of owning a smart TV — here are its suggestions for protecting your privacy, 2019. Available: https://www.businessinsider.com/smart-tv-security-fbi-warning-2019-12. Accessed on Sept. 10, 2020.
[17] M. Ford and W. Palmer, "Alexa, are you listening to me? An analysis of Alexa voice service network traffic," *Personal and Ubiquitous Computing*, vol. 23, no. 1, pp. 67–79, 2019. Available: https://doi.org/10.1007/s00779-018-1174-x
[18] L. Armasu, L. Alexa, Google Assistant, Siri Vulnerable to Laser Beam Hacking, 2019. Available: https://www.tomshardware.com/news/light-commands-laser-beam-hack-alexa-google-assistant-siri. Accessed on Sept. 10, 2020.
[19] W. Yan, K. Liu, Q. Zhou, H. Gui, and N. Zhang, "SurfingAttack: Interactive Hidden Attack on Voice Assistants Using Ultrasonic Guided Waves," *Proceedings of the Network and Distributed Systems Security Symposium*, 2020. Available: https://doi.org/10.14722/ndss.2020.24068
[20] S. Keele, "Guidelines for performing systematic literature reviews in software engineering," Technical report, Ver. 2.3 EBSE, vol. 5, 2007.
[21] M. B. Hoy, "Alexa, Siri, Cortana, and more: an introduction to voice assistants," *Medical reference services quarterly*, vol. 37, no. 1, pp. 81–88, 2018. Available: https://doi.org/10.1080/02763869.2018.1404391
[22] S. Pradhan, W. Sun, G. Baig, and L. Qiu, "Combating Replay Attacks Against Voice Assistants," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, pp. 100:1–100:26, 2019. Available: https://doi.org/10.1145/3351258
[23] V. Wong, Burger King's New Ad Will Hijack Your Google Home, 2017. Available at https://www.cnbc.com/2017/04/12/burger-kings-new-ad-will-hijack-your-google-home.html. Accessed on Dec. 25, 2019.
[24] N. Apthorpe, D. Reisman and N. Feamster, "A Smart Home is No Castle: Privacy Vulnerabilities of Encrypted IoT Traffic," *arXiv preprint arXiv:1705.06805*, 2017.
[25] N. Zhang, X. Mi, X. Feng, X. Wang, Y. Tian, and F. Qian, "Understanding and Mitigating the Security Risks of Voice-Controlled Third-Party Skills on Amazon Alexa and Google Home," *arXiv preprint arXiv:1805.01525*, 2018.
[26] V. Sivaraman, H. H. Gharakheili, A. Vishwanath, R. Boreli, and O. Mehani, "Network-level security and privacy control for smart-home IoT devices," *Proceedings of the IEEE 11th International conference on wireless and mobile computing, networking and communications*, pp. 163–167, 2015. Available: https://doi.org/10.1109/WiMOB.2015.7347956
[27] A. Furfaro, L. Argento, A. Parise, and A. Piccolo, "Using virtual environments for the assessment of cybersecurity issues in IoT scenarios," *Simulation Modelling Practice and Theory*, vol. 73, pp. 43–54, 2017. Available: https://doi.org/10.1016/j.simpat.2016.09.007
[28] Y. Meng, Z. Wang, W. Zhang, P. Wu, H. Zhu, X. Liang, and Y. Liu, "WiVo: Enhancing the security of voice control system via wireless signal in IoT environment," *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 81–90, 2018. Available: https://doi.org/10.1145/3209582.3209591
[29] H. Feng, K. Fawaz, and G. S. Kang, "Continuous authentication for voice assistants," *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, 2017. Available: https://doi.org/10.1145/3117811.3117823

[30] L. Xinyu, T. Guan-Hua, X. L. Alex, A. Kamran, L. Chi-Yu, and T. Xie, "The insecurity of home digital voice assistants: Amazon Alexa as a case study," *arXiv:1712.03327*, 2017.
[31] G. Lavrentyeva, S. Novoselov, E. Malykh, A. Kozlov, O. Kudashev, and V. Shchemelinin, "Audio-replay attack detection countermeasures," *Proceedings of the International Conference on Speech and Computer*, pp. 171–181, 2017. Available: https://doi.org/10.1007/978-3-319-66429-3_16
[32] M. Todisco, H. Delgado, and N. Evans, "Constant q cepstral coefficients," *Computer Speech & Language*, vol. 45, pp. 516–535, 2017. Available: https://doi.org/10.1016/j.csl.2017.01.001
[33] R. Bhadauria, R. Chaki, N. Chaki, and S. Sanyal, "A survey on security issues in cloud computing," *arXiv preprint arXiv:1109.5388*, pp. 1–15, 2011.
[34] C. Stergiou, K. E. Psannis, B. G. Kim, and B. Gupta, "Secure integration of IoT and cloud computing," *Future Generation Computer Systems*, vol. 78, pp. 964–975, 2018. Available: https://doi.org/10.1016/j.future.2016.11.031

Appendix

Original and Modified Search Strings

ID	Description	Search String
SS1	Original Search String	(intelligent virtual assistants OR intelligent personal assistant OR virtual personal assistant OR voice assistant OR Amazon Echo OR Alexa OR Google Home OR Microsoft Cortana OR Apple Siri) AND (cybersecurity OR information security OR electronic security OR internet safety OR vulnerability threat OR privacy OR confidentiality OR hack OR cyberattack OR penetration test) AND (internet of things OR system of computing devices OR network of computing devices)
SS2	Modified Search String to accommodate character limit	(intelligent virtual assistants OR intelligent personal) AND (Cybersecurity OR Information security) AND (internet of things OR system of computing devices)
SS3		(intelligent virtual assistants OR intelligent personal) AND (cyberattack OR penetration test) AND (network of computing devices OR system of computing devices)
SS4		(intelligent virtual assistants OR virtual personal assistant) AND (Cybersecurity OR Electronic security) AND (internet of things OR network of computing devices)
SS5		(intelligent virtual assistants OR Amazon Echo) AND (Cybersecurity OR Vulnerability threat) AND (internet of things OR network of computing devices)