Eliciting Privacy Concerns for Smart Home Devices from a User Centered Perspective

Chola Chhetri and Vivian Genaro Motti

George Mason University, Fairfax VA 22030, USA

{cchhetri, vmotti}@gmu.edu

Abstract. Smart homes are equipped with an ecosystem of devices that support humans in their everyday activities, ranging from entertainment, lighting and security systems. Although smart devices provide home automation features that are convenient, comfortable, and easy to control, they also pose critical privacy risks for users, especially considering their continuous ability to sense users' information and connect to web services. To elicit privacy concerns from a user-centric perspective, the authors performed a thorough analysis of 128 online reviews of consumers of smart home hubs –including Amazon Echo, Google Home, Wink and Insteon. The reviews expressed users' concerns about privacy. The reviews were coded and classified according to four information security principles and temporal dimensions ranging from data collection to information sharing. A discussion on how to improve the design of smart home devices with privacy-enhanced solutions is provided.

Keywords: User-centered Design, Security and Privacy Protection, Privacy, Privacy Concerns, Smart Home Devices.

1 Introduction

The Internet of Things (IoT) includes smart home devices that automate user tasks at home. According to a report from 2017, there exist 8.4 billion connected gadgets [1] and IoT field is expected to continue growing in the near future. Smart home technologies, a subset of IoT devices, are also expected to face significant growths. Business Insider Intelligence estimates that 1.8 billion smart home devices are expected to be sold by 2019, generating an estimated annual revenue of $490 billion [2]. In 2017, the estimated sale of smart home technologies in the United States (US) was 39 million units (35.9 million devices and 3.1 million hubs). Based on this data, one in every ten US households has at least one smart home device [3].

The primary motivator for users of smart home devices is the convenience that such devices offer. These technologies also promise comfort, control, safety and security [3]. However, the rise in adoption of smart home technologies presents new challenges to security and privacy [4, 5]. In 2016, the Mirai botnet affected hundreds of thousands of IoT devices —including Internet Protocol (IP) cameras, Digital Video Recorders (DVRs), routers, printers and Voice over Internet Protocol (VoIP) phones.

In this attack a large scale distributed denial of service (DDOS) against multiple targets was executed. Targets of this attack included Krebs on Security, Lonestar Cell and Dyn, a popular Domain Name System (DNS) provider. The DDoS on Dyn affected popular services such as Amazon, Github, Netflix, Paypal, Twitter and Reddit [6].

As the number of smart home devices grows, attacks of such nature are also likely to grow, not only in scale but also in sophistication [7]. Do users know about it? Are users concerned? What are they primarily concerned about? To address such questions, further investigation focusing on users concerns about smart home privacy is needed.

Previous studies about user adoption of smart speakers found privacy as the primary reason for non-adoption of these devices and adopters placed value on their privacy [8]. Privacy concerns of users are positively correlated to privacy importance [9]. In a study of fitness trackers, privacy concerns of users were found to be directly related to their valuation of data collected by the tracking devices [10]. Privacy behaviors also depend on context and evolve over time [11]. However, little work has been done to understand privacy concerns of smart home users [12, 13]. Understanding the privacy concerns of users of smart home devices helps stakeholders, investigators, and vendors to develop hardware and software solutions that are better suited to address privacy concerns of users. Improving smart home technologies with privacy-enhanced solutions has potential to boost user confidence and trust in such technologies [13, 14].

To elicit privacy concerns from a user-centric perspective, we analyzed thoroughly privacy-related online reviews of users of smart home hubs – including Amazon Echo [15], Google Home [16], Wink Hub 2 [17] and Insteon Hub [18]. One hundred twenty eight reviews expressing privacy concerns were retrieved and classified according to their contents, security principles involved and temporal dimensions regarding the automation process and information lifecycle, from data collection to sharing. A discussion on how to improve the design of smart home devices with privacy-enhanced solutions is provided.

The rest of the paper is organized as follows: Section 2 describes the research methodology; Section 3 presents research results; and Section 4 includes conclusions, recommendations, limitations and next steps.

2 Methodology

To analyze privacy concerns of users, five common smart hubs were selected, namely Amazon Echo Dot 2, Samsung SmartThings Hub, Google Home, Wink Hub 2 and Insteon Hub [19], [20]. There was no single portal containing customer reviews of smart home devices. We selected amazon.com and bestbuy.com as our sources as they contained the largest number of user reviews on these selected products (n=66656). Table 1 shows the number of reviews and the source for the selected devices.

Table 1. Number of reviews for five smart hubs. The source of reviews for all devices, except Google Home (*), was amazon.com. Google Home reviews were obtained from bestbuy.com

Device	Number of Reviews (n=66656)
Amazon Echo Dot 2	57079
Google Home *	6902
Samsung SmartThings Hub	1856
Wink Hub 2	558
Insteon Hub	261

We filtered out the reviews that did not include the keyword 'privacy'. Our resulting dataset included 128 reviews: 120 for Amazon Echo, six for Google Home, one for Wink Hub 2, and one for Insteon Hub. No privacy-related reviews were found for Samsung SmartThings hub [21]. The reviews in our data set dated from October 2016 to October 2017 and included reviews that were classified as verified purchases by amazon.com and bestbuy.com.

We extracted and coded the reviews manually, and analyzed the qualitative data. For sentiment analysis, the reviews were manually read and then coded as positive, neutral or negative. For temporal analysis, the reviews were coded based on the life cycle of data in the smart home architecture, namely collection, transmission, storage and sharing. For security principle analysis, the reviews were coded based on the security principles confidentiality, integrity, availability and authentication. Table 2 illustrates the codebook showing the main themes of coding. All authors were in consensus with the codes.

Table 2. Codebook showing codes/themes for analyses performed in the study.

Concern	Sentiment	Temporal	Principle
	Positive	Collection	Confidentiality
Specific	Neutral	Transmission	Integrity
Non-specific	Negative	Storage	Availability
		Sharing	Authentication

Our methodology was inspired by [22], an analysis of privacy concerns in user comments on wearable devices involving exploratory and empirical methods.

3 Results

3.1 Specific Concerns

While 33% of the reviews were general, in which the users mentioned privacy concern but did not specify their privacy concern in detail, 67% of the users specified precisely what their privacy concern was. The top user concern was that these devices

were always listening to their conversations. Other five user concerns sorted per order of popularity were: (1) tracking of users, their actions and preferences, (2) storage of conversations and their transcripts (for audio conversations) in the cloud, (3) the lack of security of such content in the cloud, (4) the potential of private conversations to be hacked, and (5) the likelihood of such information to be subject to legal discovery by law enforcement and eventually disclosed publicly. Figure 1 depicts these concerns and their frequencies of occurrence.

The identified concerns resonate with recent studies that discuss and demonstrate various smart home vulnerabilities, such as network observing, tracking, eavesdropping, user behavior prediction, data leakage, data theft, identity theft, social engineering, disruption or denial of service and software exploitation [7], [23], [24].

Users in our analysis were concerned that microphone-enabled smart home devices were recording private conversations, background conversations and noise. Such content included personal conversations, such as family members speaking to one another, and conversations not directed to the device. For example, Amazon Echo is expected to record only conversations following the wake-up word 'Alexa'; however, users reported to find recordings of private conversations not including such wake-up word.

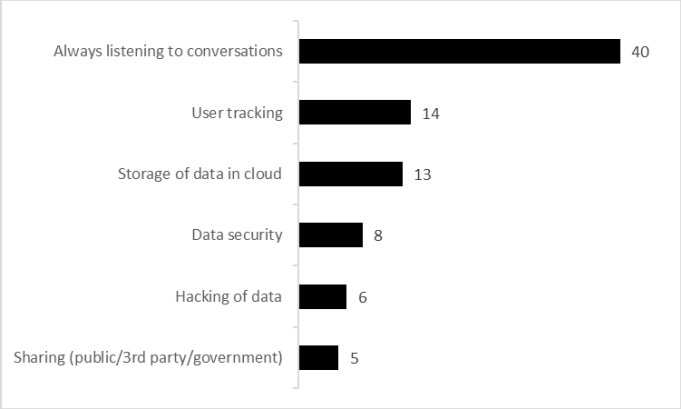

Fig. 1. Top six themes in users' privacy concerns and their frequencies

According to one user, "Echo thinks my TV is talking to it. Very often, without the word Alexa being said the Echo will start jabbering or playing a song while watching TV." Another user reported that when he/she checked the history of the requests made to Alexa, he/she found many recordings of strange people talking that did not live in the house. He/she mentioned that "It even randomly records things in my house like my dog barking, the TV audio and a regular conversation without requesting 'Alexa'… I'm either going to send it back or keep it turned off unless I want to use it. I don't want the world to know what's being said in the privacy of my home and I don't want to hear what's going on in their homes. Very scary."

Users expressed that they believed the features and convenience offered by smart home technologies outweighed their privacy concerns, which led them to utilize smart home technologies despite their privacy concerns.

3.2 User Sentiments

We analyzed the sentiments expressed by users in their privacy concerns, by coding them as negative, neutral or positive. We found that user sentiments associated with privacy concerns are mostly negative (74%). Among the remaining 26% reviews, three quarters (75%) expressed positive sentiments and the rest of them were neutral. Examples of reviews are shown in Table 3.

Table 3. Three examples of reviews with negative, positive and neutral sentiments.

Sentiment	Sample Reviews
Negative	"I am a bit paranoid that such a device will support furthering the evolution in the loss of privacy." "There's no privacy because every question that is asked is seen on the Alexa app which it should be private. I am disappointed with it."
Positive	"We love it even though we were initially concerned about it eavesdropping and related privacy issues."
Neutral	"For those concerned about privacy (like a few of my friends) just unplug when you don't want the device to listen in on your conversations or activities."

3.3 Temporal Analysis

Smart home devices collect information from users, transmit to a remote server on the cloud for storage, processing, and sharing. In a smart home, the life cycle of data ranges from collection and transmission to storage and sharing.

Based on our investigation, data collection stands out as the stage that concerns users the most. As Table 4 illustrates, about half (49%) of the 128 reviews analyzed mentioned the user concern was related to data collection, followed by storage (23%), sharing (9%), and transmission (2%).

Table 4. Four examples of privacy concerns quoted from end users' reviews considering the lifecycle of data from collection and transmission to storage and sharing.

Stage	Examples/Tasks	Sample Reviews	Percentage of Reviews (n=128)
Collection	Devices with microphones and sensors collect data	"It is connected to the internet and listens and records all of the time, not just when you talk to it." "Everything you say to it is recorded... even things you don't say to it"	49%
Transmission	Collected data is usually transmit-	"Records and uploads your conversations automatically"	2%

	ted to a remote server or cloud for storage	"I would prefer to remove the cloud or any data sent which is collected, retained and resold as I have no need to control outside my network and have ways I can connect to my local network remotely without their spying servers being involved."	
Storage	Data is stored in a remote server or cloud	"This device records and stores everything it hears offsite forever." "RECORDS everything you ask it. It keeps the text of what you ask, but it also actually keeps the audio recording of what you asked it"	23%
Sharing	Vendors may share this retained data to third parties or collected data may be requested by law enforcement	"I attempted to read all of the terms and conditions but soon found out that agreeing to them means that third parties can end up with my voice print!" "It is another window into your privacy so that they can build a more precise profile on you for marketing purposes."	9%

3.4 Security and Privacy Principles

The breach of the four main security principles—Confidentiality, Integrity, Availability [25], and Authentication [26]— results in critical consequences for users. Confidentiality deals with keeping data secret from unauthorized parties using techniques like encryption. Integrity deals with prevention of tampering of data. Availability ensures data is available to authorized parties when necessary and authentication checks credentials of parties trying to access data [27]. An insight into the principles that concern users the most can direct research and development focus towards those areas.

As expected, users did not clearly utilize these terms in their reviews. Hence, we coded and mapped the reviews to these principles. Content of some reviews did not fall into any of the four principles. Such reviews could not be coded for this analysis but were still used for other analyses. Among the 73 reviews that were successfully mapped to a principle, 90% were related to confidentiality and 10% were related to authentication. None were directly related to integrity and availability. Thus, confidentiality of data was a leading privacy concern of users of smart home devices.

3.5 Privacy Protection Strategies

Users of smart home devices who expressed concern about privacy seemed to adopt individual controls to address their concerns. Our analysis revealed that users concerned about privacy adopt three major approaches. The first approach consists in

deleting the history of audio and text data when possible. According to users, a common drawback of such products is the inability to delete collected and stored data in bulk. Having to delete large collection of data one-by-one is a major annoyance for users.

The second approach consists in turning off the device when it is not in use. Devices with the ability to turn off the microphone seemed to be the favorite for concerned users. Switching the microphone off or even the entire device when not in use is a user practice to prevent disclosure and misuse of private information mainly by ensuring that the device is not listening and recording private conversations that are not directed to it.

The third approach was deciding to use a product from a vendor who has demonstrated to advocate for data privacy. Users expressed confidence with companies that stood up against providing data to law enforcement. For instance, one user expressed confidence in trusting the vendor to fight for the privacy of user data: "I also trust this vendor and they have shown to fight for users' data privacy so far."

4 Conclusions

4.1 Recommendations for Privacy Enhancing Solutions

Online reviews in the data set analyzed included six suggestions for enhancing privacy in smart home technologies. We discuss those suggestions and also add our own recommendations for privacy-enhanced smart home technology solutions. Author recommendations are based on team discussion and experience. Recommendations in both categories are in alphabetical order.

4.2 User-suggested Recommendations

Advocacy. Users asked vendors to advocate more for data protection. Such protection can be ensured by utilizing and informing users of techniques used to safeguard data from insider and outsider attackers. Users mentioned they trust vendors that stand up against the release of data collected to third parties.

Interface. Another user suggestion is that vendors provide a user-friendly interface with the ability to view, manage and delete data collected by the devices.

Local Control. Users suggested that vendors develop locally controlled hubs for smart home automation instead of using the cloud. A local storage hub will eliminate the need to store data in the cloud and elevate user trust by eliminating privacy concerns related to data stored in the cloud.

Policy. Users showed interest to understand what data is collected, how the data collected is or will be used, and who has access to it. Users suggested vendors clearly state the policies regarding collection, storage, sharing and protection of data.

Safeguarding. Users were concerned about protection of data in the cloud by the vendors. They suggested that vendors take steps, such as encryption, to protect their data from being hacked.

Trust. Users mentioned that they prefer the ability to control the collection of smart home data. Users expressed frustration over the inability to control the data collection and recordings. They suggested that vendors provide a way to manage and especially to delete the collected data.

4.3 Recommendations

Accuracy. Manufacturers of smart home technologies must address programming flaws and lack of accuracy, for example, accuracy in recognizing wake-up words (words that activate a voice enabled smart home device) and eliminating false positives.

Authentication. Smart home devices should provide mechanisms to authenticate the user to avoid unwanted users from using the devices or accessing their services in an unauthorized manner.

Data Protection. Vendors should employ data protection techniques, such as encryption, to safeguard data in all stages of its lifecycle —collection, transmission, storage, and sharing.

Opt In. Data collection is necessary for vendors to improve the services offered to users. However, users are concerned about excessive data collection and have questioned it. Vendors can address such concern by providing users with the option to opt-in for such data collection. Making data collection an opt-in rather than mandatory can bring in more users who would otherwise not utilize such devices due to privacy concerns.

Policy. Providers of smart home technologies can gain higher consumer confidence by clearly stating what data they collect from the smart home device, how they transmit the data collected, how they handle smart home data and what measures they take to safeguard the data for ensuring its confidentiality and privacy.

Regulatory Framework. The data collected from smart homes should be subject to data protection law or regulation. A legal framework for the protection of data collected by smart home devices is necessary. Industry-level guidelines and best practices in smart home data protection are needed to make the smart home domain more secure and more private. The user-vendor-government trio needs to work collectively to preserve privacy in the age of smart homes. Past research [28] has also shown that users expect strong legal protection of their data.

Stop Technique. When a device is recording conversations, it is essential for the device to know when to stop recording. If a device does not know when to stop recording the conversation, it can record private conversations not directed to it. We recommend introducing a stop word (for example, 'thank you', 'bye bye', bye [name of device], etc.) or a stop technique. Such a method can be beneficial in informing the device of the end of the conversation, indicating it to stop recording or collecting data. An indicator light is recommended as a method of informing users of the recording action.

Visibility. Placing the on/off switch in a visible location and an indicator light depicting the recording action can elevate the comfort level of concerned users. Live status may also be helpful concerning collection of data.

4.4 Limitations

The extraction, reading and coding of reviews was performed manually. The resulting data set contained more reviews provided by Echo consumers than other devices. Thus, there may be a bias towards Echo. The study focuses on analyzing privacy concerns of only consumers that chose to write an online review for the product. Demographic analysis was not feasible as the online reviews lacked such information.

Privacy may be addressed without explicitly mentioning it. This study was limited to the content analysis of reviews that explicitly mentioned the word ‘privacy’. Another limitation is that users who post comments tend to have more extreme opinions about those devices and belong to a subset of users who are tech savvy, young, and literate.

4.5 Next Steps

In this paper, we analyzed privacy concerns of smart home device users to shed light into privacy concerns of actual consumers and we discuss recommendations for making the devices more privacy preserving. We expect this paper to motivate researchers, developers and manufacturers to develop privacy-enhanced smart home solutions.

We will further explore smart home privacy concerns through complementary research methods including responses from an online survey, and we will extend the data set analyzing comments of additional smart home devices. It is likely that people who are concerned about privacy are choosing not to use such technologies [29]. We

will also analyze concerns of such non-users of smart home devices to address their privacy concerns and smart home technology adoption. Future work will include research into educational tools and training to promote awareness on smart home privacy. The smart home domain can benefit from an investigation into the legal framework for smart home data protection.

References

1. Gartner, Inc., https://www.gartner.com/technology/research.jsp.
2. BI Intelligence, http://www.businessinsider.com/research.
3. Smart Home Ecosystem: IoT and Consumers. Park Associates and Consumer Electronics Assoiciation, USA (2014).
4. Ye, M., Jiang, N., Yang, H., Yan, Q.: Security Analysis of Internet-of-Things: A Case Study of August Smart Lock. In: MobiSec 2017: Security, Privacy, and Digital Forensics of Mobile Systems and Networks, pp. 499–504, Georgia (2017).
5. Fernandes, E., Jung, J., Prakash, A.: Security Analysis of Emerging Smart Home Applications. In: 2016 IEEE Symposium on Security and Privacy (SP), pp. 636–654. IEEE California (2016).
6. Antonakakis, M., April, T., Bailey, M., Bursztein, E., Cochran, J., Durumeric, Z., Alex Halderman, J., Menscher, D., Seaman, C., Sullivan, N., Thomas, K., Zhou, Y.: Understanding the Mirai Botnet. In: Proceedings of the 26th USENIX Security Symposium, pp. 1093–1110. Vancouver, Canada (2017).
7. Arabo, A., Brown, I., El-Moussa, F.: Privacy in the age of mobility and smart devices in smart homes. In: ASE/IEEE International Conference on Privacy, Security, Risk and Trust, and ASE/IEEE International Conference on Social Computing, SocialCom/PASSAT, pp. 819–826. IEEE (2012).
8. Lau, J., Zimmerman, B., Schaub, F.: Alexa , Are You Listening ? Privacy Perceptions, Concerns and Privacy-seeking Behaviors with Smart Speakers. In: Proceedings of ACM Human-Computer Interaction, p. 102:1-31 (2018).
9. Acquisti, A., Grossklags, J.: Privacy and rationality in individual decision making. IEEE Secur. Priv, 3, 26–33 (2005).
10. Vitak J., Liao Y., Kumar P., Zimmer M., Kritikos K.: Privacy Attitudes and Data Valuation Among Fitness Tracker Users. In: Chowdhury G., McLeod J., Gillet V., Willett P. (eds) Transforming Digital Worlds. iConference 2018. LNCS, vol 10766. Springer, Cham (2018).
11. Nissenbaum, H.: A contextual approach to privacy online. Digit. Enlight. Yearb. 2012. 219–234 (2012).
12. Fruchter, N., Liccardi, I.: Consumer Attitudes Towards Privacy and Security in Home Assistants. In: CHI'18 Extended Abstracts. ACM, Canada (2018).
13. Zeng, E., Mare, S., Roesner, F.: End User Security & Privacy Concerns with Smart Homes. In: Symposium on Usable Privacy and Security (SOUPS). (2017).
14. Kugler, L.: The war over the value of personal data. Communications of the ACM. 61, 17–19 (2018).
15. Echo Dot (2nd Generation), https://www.amazon.com/dp/B01DFKC2SO.
16. Google Home, https://store.google.com/us/product/google_home.
17. Wink Hub 2, https://www.wink.com/products/wink-hub-2/.
18. Insteon Hub, https://www.insteon.com/insteon-hub/.

19. The Best Smart Home Devices of 2017, https://www.pcmag.com/article2/0,2817,2410889,00.asp.
20. Best Smart Home Devices of 2017, https://www.cnet.com/topics/smart-home/best-smart-home-devices/.
21. SmartThings Hub, https://www.smartthings.com/products/smartthings-hub.
22. Motti, V.G., Caine, K.: Users' Privacy Concerns About Wearables: Impact of Form Factor, Sensors and Type of Data Collected. In: Brenner, M., Christin, N., Johnson, B.., Rohloff, K. (eds.) Financial Cryptography and Data Security, LNCS, vol 8976, pp. 1-14. Springer, Heidelberg (2015).
23. Geneiatakis, D., Kounelis, I., Neisse, R., Nai-Fovino, I., Steri, G., Baldini, G.: Security and Privacy Issues for an IoT based Smart Home. In: Information and Communication Technology, Electronics and Microelectronics (MIPRO). pp. 1292–1297. Opatija, Croatia (2017).
24. Apthorpe, N., Reisman, D., Feamster, N.: A Smart Home is No Castle: Privacy Vulnerabilities of Encrypted IoT Traffic. In: Data and Algorithmic Transparency Workshop (DAT). New York (2016).
25. Jacobsson, A., Davidsson, P.: Towards a model of privacy and security for smart homes. In IEEE World Forum Internet Things, WF-IoT 2015 - Proc., 727–732. IEEE (2016).
26. Fisher, B.: Identity and Access Management for Smart Home Devices. National Institute of Standards and Technology, (2016).
27. Gibson, D.: Managing Risk in Information Systems. Jones and Bartlett Learning, Burlington, MA (2011).
28. Malkin, N., Bernd, J., Johnson, M., Egelman, S.: " What Can't Data Be Used For?" Privacy Expectations about Smart TVs in the U.S. In: Proceedings of the European Workshop on Usable Security (EuroUSEC). IEEE, London (2018).
29. Zheng, S., Apthorpe, N., Chetty, M., Feamster, N.: User Perceptions of Smart Home IoT Privacy. In: Proceedings of ACM Human-Computer Interaction, p. 200:1-20. ACM (2018).