Analysing users´ privacy preferences in smart-home environments with situational contexts

Christopher Ruff [1], Benedict Benthien[2], Alexander Orlowski[3]

Abstract: Due to the increasing adoption of smart home devices and technologies, implications for privacy gain importance. In this paper, correlations between specific characteristics of people and their preferences regarding the activity status of components in smart home devices are investigated. In addition, said preferences are analysed for inherent patterns to assist people in their decisions by suggesting preferences, which often occur together. A special focus of this work is the differentiation of preferences according to situational contexts. An online survey was conducted, and the results were analysed. The results imply strong correlations within the preferences and differences in preferences across different contexts.

Keywords: smart home, Smart Environment, Privacy, Privacy Assistant, privacy preferences, contextual privacy

1 Introduction

The increasing adoption of smart devices, based on the Internet of Things (IoT), offers a range of opportunities. Most notably, these technologies have the potential, to increase efficiency and convenience of users' daily life, thus enhancing their overall quality of life [NLP20]. As smart home can be considered a sub-set of smart environments, where the home environment is enhanced by the various (often times automated) functions of IoT devices, serving different use cases and areas. [SKB10]. The automation of the smart home itself includes areas such as heating, electrics, lighting, security and monitoring of the smart home. Further service offerings and devices tailored to the home environment stem from the field of consumer electronics, such as smart TVs, smart speakers, and smart kitchen appliances or other household items (e.g. vacuums, coffee machines, refrigerators).

Due to the large number of technical components installed in those smart devices used for data collection (e.g. microphones, speakers, cameras, temperature or humidity sensors), a variety of information and personal data is collected and processed, which has implications for the privacy of all individuals present in the smart home. To act autonomously, the devices are continuously sending and receiving data via the internet, which is often not

[1] Fraunhofer Institute IAO, Nobelstreet 13, Stuttgart, 70569, christopher.ruff@iao.fraunhofer.de, https://orcid.org/0000-0003-0484-4131

[2] Fraunhofer Institute IAO, Nobelstreet 13, Stuttgart, 70569, Benedict.Benthien@iao.fraunhofer.de

[3] University of Tübingen, IZEW, Wilhelmstreet 19, 72074 Tübingen, alexander.orlowski@uni-tuebingen.de

cbe doi: 10.18420/OID2023_11

transparent to the users. Although the handling of personal data in Europe is regulated by the General Data Protection Regulation (GDPR[4]), the associated services are implemented on devices from various manufacturers that use proprietary and encrypted protocols, which leads to problems regarding data autonomy and transparency [OME19]. Through the analysis of consciously and unconsciously shared data on usage behaviour, manufacturers are able to derive data about users that go beyond the intended purpose of data sharing, which poses privacy violations and security risks [GHN16].

To mitigate these risks, an assistant system "DAMA"[5] was developed, that oversees all devices in the smart-home, transparently informs about their presence and activity and enhances the users' control over the collection of data during situations where privacy is more important than the functionality specific devices or sensors.

Acknowledgement

We thank the Baden-Württemberg Stiftung for supporting and financing the DAMA project. For more information, please visit: https://www.bwstiftung.de/.

2 Approach

The DAMA assistant allows the users to set specific configurations that are triggered (semi-)automatically or on specific user requests. Those situations or "contexts" are based on temporarily increased needs for privacy (e.g., a private telephone call, unfamiliar people visiting, preferences, etc.), of people living in the smart-home as well as visitors. The system will then automatically change the settings of the smart devices to restrict their data collection abilities or turn them on or off completely for the duration of the specified context.

We conducted an online survey to determine the following questions. 1) Are there correlations between specific personal characteristics (independent variables) and their decisions regarding activation status of smart devices in a smart home (dependant variables)? 2) Are there significant preference patterns as to the activity status of smart devices in a smart home across multiple situational context?

As a hypothetical smart-home environment we chose the following smart-devices: Smart Speakers with AI-Assistant; Smart TV; Smart Web-Cam (i.e. security cam); Smart doorbell.

For the hypothetical situational contexts, following scenarios were introduced: 1) Person alone at home; 2) Person at home with his/her partner; 3) Person at home with an unfamiliar visitor; 4) Person visiting a friend (in a smart home environment) 5) Person

[4] 1https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:32016R0679, abgerufen am 21.08.22

[5] https://www.dama-projekt.de

renting an unfamiliar smart home for a short period of time (e.g. AirBnB).

2.1 Methods

The online survey was created using the Limesurvey[6] tool. We used a combination of quantitative and qualitative questions. For the quantitative questions a Likert scale was used. For the data analysis, processing and clustering, Jupyter Notebook[7] was used with the Scikit-Learn library[8] and the MLxtend Library[9] for the analysis of certain itemsets. The visualization of cluster heatmaps was achieved using Seaborn[10].

2.2 Data Analysis

To investigate the research questions, statistical methods were first applied to the collected data. Independent and dependent variables were described using simple statistical measures. In addition, correlations were examined within the independent and dependent variable groups, as well as between the two groups of variables. By analysing the distributions, general insights into the characteristics and privacy preferences were gained. The correlation analysis potentially revealed associations between the variables, which were further investigated using subsequent methods.

After conducting a statistical analysis of the variables for the various contexts, they were examined for the occurrence of clusters using the most suitable clustering algorithm. Initially, K-Means and K-Modes were compared as clustering algorithms for the dependent variables across all contexts. Subsequently, the impact of including the independent variables on the clustering results was investigated by comparing K-Means and K-Prototypes.

To gain further information on the mutual occurrence of variables and therefor strengthen the findings of the clustering, the variables were examined for itemsets. The calculated support values of these itemsets allow for quantitative comparison of their importance with each other and with themselves over different contexts.

3 Results

Out of 658 participants, we received 519 fully completed questionnaires. We removed questionnaires that were filled in under 3 minutes to weed out fake answers. Most participants were between 20 and 29 years of age as a lot of participants came from a

[6] https://www.limesurvey.org/

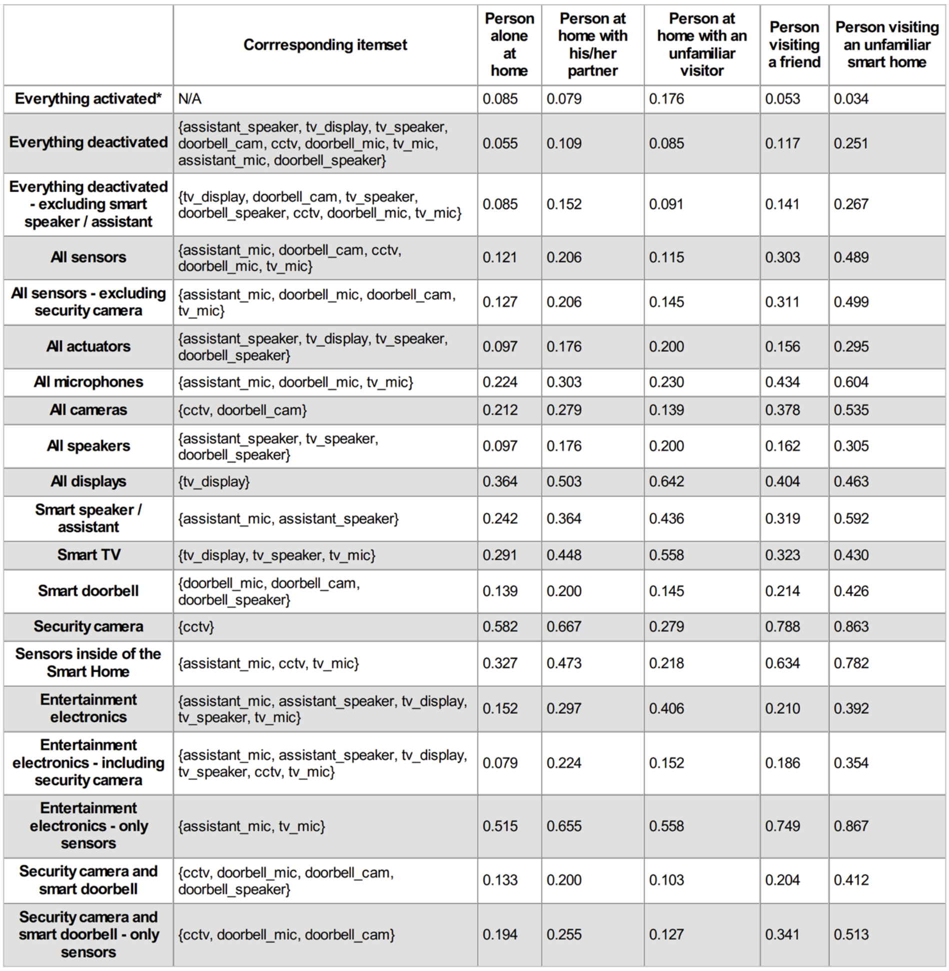
[7] https://jupyter.org/

[8] https://scikit-learn.org/

[9] https://rasbt.github.io/mlxtend/

[10] https://seaborn.pydata.org/

university background. This limitation should be taken into account when interpreting the results of this study.



	Corrresponding itemset	Person alone at home	Person at home with his/her partner	Person at home with an unfamiliar visitor	Person visiting a friend	Person visiting an unfamiliar smart home
Everything activated*	N/A	0.085	0.079	0.176	0.053	0.034
Everything deactivated	{assistant_speaker, tv_display, tv_speaker, doorbell_cam, cctv, doorbell_mic, tv_mic, assistant_mic, doorbell_speaker}	0.055	0.109	0.085	0.117	0.251
Everything deactivated - excluding smart speaker / assistant	{tv_display, doorbell_cam, tv_speaker, doorbell_speaker, cctv, doorbell_mic, tv_mic}	0.085	0.152	0.091	0.141	0.267
All sensors	{assistant_mic, doorbell_cam, cctv, doorbell_mic, tv_mic}	0.121	0.206	0.115	0.303	0.489
All sensors - excluding security camera	{assistant_mic, doorbell_mic, doorbell_cam, tv_mic}	0.127	0.206	0.145	0.311	0.499
All actuators	{assistant_speaker, tv_display, tv_speaker, doorbell_speaker}	0.097	0.176	0.200	0.156	0.295
All microphones	{assistant_mic, doorbell_mic, tv_mic}	0.224	0.303	0.230	0.434	0.604
All cameras	{cctv, doorbell_cam}	0.212	0.279	0.139	0.378	0.535
All speakers	{assistant_speaker, tv_speaker, doorbell_speaker}	0.097	0.176	0.200	0.162	0.305
All displays	{tv_display}	0.364	0.503	0.642	0.404	0.463
Smart speaker / assistant	{assistant_mic, assistant_speaker}	0.242	0.364	0.436	0.319	0.592
Smart TV	{tv_display, tv_speaker, tv_mic}	0.291	0.448	0.558	0.323	0.430
Smart doorbell	{doorbell_mic, doorbell_cam, doorbell_speaker}	0.139	0.200	0.145	0.214	0.426
Security camera	{cctv}	0.582	0.667	0.279	0.788	0.863
Sensors inside of the Smart Home	{assistant_mic, cctv, tv_mic}	0.327	0.473	0.218	0.634	0.782
Entertainment electronics	{assistant_mic, assistant_speaker, tv_display, tv_speaker, tv_mic}	0.152	0.297	0.406	0.210	0.392
Entertainment electronics - including security camera	{assistant_mic, assistant_speaker, tv_display, tv_speaker, cctv, tv_mic}	0.079	0.224	0.152	0.186	0.354
Entertainment electronics - only sensors	{assistant_mic, tv_mic}	0.515	0.655	0.558	0.749	0.867
Security camera and smart doorbell	{cctv, doorbell_mic, doorbell_cam, doorbell_speaker}	0.133	0.200	0.103	0.204	0.412
Security camera and smart doorbell - only sensors	{cctv, doorbell_mic, doorbell_cam}	0.194	0.255	0.127	0.341	0.513

Fig. 1: Support values for itemsets of components that participants want to be mutually deactivated across situational contexts. “Everything activated” would have resulted in an empty itemset, so the case in which every component would be mutually activated was observed

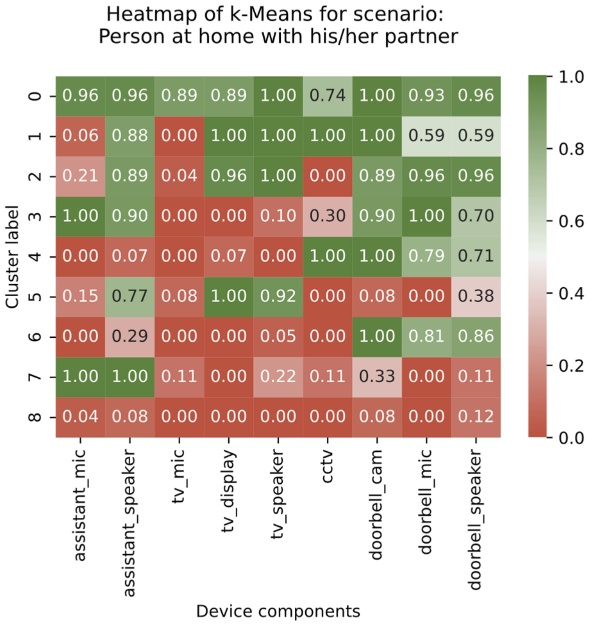


Fig. 2: Heatmap of k-Means for scenario 2). Values near 0 or 1 indicate a strong similarity of the decisions to (de-)activate a component within a cluster.

The independent variables often show expected correlations among themselves. Moreover, mostly weak correlations between the independent and dependent variables were detected. However, the correlation coefficient values are partly too small to confirm these results quantitatively. One reason for this is the different scales used for the variables, which makes correlations difficult to estimate and sometimes undefined. This reduces the answerability of the first research question in this study. The distributions of the dependent variables were examined for patterns using frequency distributions and correlation coefficients, which led to various logical groupings of components as itemsets. The clustering results (exemplary visualization seen in Fig. 2) were not of sufficient quality to derive quantitative conclusions from them, but they provided a qualitative contribution to the search for additional itemsets to be investigated. The inclusion of the independent variables in clustering negatively affected the quality of the results. The quantitative analysis of itemsets as seen in Fig. 1 showed that clear patterns can be found in decisions to deactivate certain components together. Grouping the components by type yielded clearer patterns and tendencies than grouping the components that are installed in the same devices, which makes the possibility of granular control of smart home devices even more relevant. The fact that support values for identical itemsets differ significantly across contexts underscores the importance of context in the decision-making process for device component activity. Therefore, the second research question can be affirmed based on these results. In general, it was found that as participants' familiarity with the smart home decreased, the desire to deactivate device components increased. The desire to deactivate sensors (data collectors) was expressed more strongly than to deactivate actuators. This effect was observed more strongly for components of entertainment electronics than for components of security electronics, as security was rated as more important, especially in the third context. For the fourth and fifth contexts, where study participants themselves were guests in other people's smart homes, components were most frequently deactivated.

4 Conclusion

This paper highlights and investigates specific users' privacy needs in smart home environments. A survey was conducted, referencing selected smart home devices and privacy contexts, although both areas can take on virtually unlimited dimensions in the real world, thus confirming the need for further research. Examination of other methods of data analysis, such as calculating additional metrics for itemsets, could provide further insights into the survey data. A need for the possibility to regulate individual device components independently of the overall devices is identified. Furthermore, choices related to component activity could be clustered among the participants, implying the possibility of creating justified privacy pre-sets for situational contexts. By requiring a small portion of participants to not have to indicate any preferences for device components, this not only expresses a lack of interest in data protection, but also highlights the importance of said pre-sets to facilitate the choice of an appropriate solution. In addition to analysing existing preferences, educating about the importance of this issue is another important step.

Bibliography

[NLP20] S. Nižetić, P. Šolić, D. López und L. Patrono, „Internet of Things (IoT): Opportunities, issues and challenges towards a smart and sustainable future," en, Journal of Cleaner Production, Jg. 274, S. 122 877, Nov. 2020. doi: 10.1016/j.jclepro.2020.122877.

[SKB10] H. Strese, U. Seidel, T. Knape und A. Botthof, „smart home in Deutschland," Institut für Innovation und Technik (iit), Jg. 46, S. 13, 2010. /www.iit-berlin.de/iit-docs/ee5de919903845599e54482177694873_Smart_Home_in_Deutschland.pdf.

[OME19] T. O'Connor, R. Mohamed, M. Miettinen, W. Enck, B. Reaves und A.-R. Sadeghi, „HomeSnitch: behavior transparency and control for smart home IoT devices," en, in Proceedings of the 12th Conference on Security and Privacy in Wireless and Mobile

[GHN16] M. Ghiglieri, M. Hansen, M. Nebel, J. V. Pörschke und H. S. Fhom, Smart-TV und Privatheit, Ser. Forum Privatheit und selbstbestimmtes Leben in der digitalen Welt. Karlsruhe: Fraunhofer ISI, 2016. https://publica.fraunhofer.de/bitstreams/d1af15f4-2c7a-4f28-ac0a-dcfa0308be5b/download.