Data Alarm: Smart-Home Interaction Model for User Data Privacy

Sneha Das
Aalto University
Espoo, 02250, Finland
sneha.das@aalto.fi

William Turner
Aalto University
Espoo, 02250, Finland
william.turner@aalto.fi

Chandni Sharma
Aalto University
Espoo, 02250, Finland
chandni.sharma@aalto.fi

Abstract

The Internet of things (IoT) and ubiquitous computing are common in an average household today, and the growth trend is continuing. While such systems ease and aid our lives, they raise privacy concerns due to their pervasive nature. Specifically, the unregulated, non-transparent and sometimes unauthorized back-end data collection and exchange mechanisms are worrisome. In this work, we describe the creation of the Data Alarm, a data screen for a smart home which functions as a privacy-level indicator and is meant to work with smart and distributed systems. The objective of this work is to create an interaction model which empowers the users by keeping them aware of their privacy level at all times, and thus providing them with a tool to better manage and control their privacy levels and expectations.

Author Keywords

Personal data privacy; Permission-status check; Transparency, access-management, Smart-home; IoT

Introduction

Ubiquitous computing (ubicomp) has become a common commodity in households. Generally, ubicomp is the concept of integrating computing anytime, anywhere, in a seamless manner. Smart homes and smart health care are examples of systems which employ ubicomp. Smart home technology can ease the burden of everyday tasks, such as an autonomous vacuum, provide important information about the home, like a lightweight internet-connected security system, and improve a home's energy efficiency, through services like smart thermostats, all without much intervention from the user. Additionally, today home-care technologies can also provide care for the elderly and people with physical disabilities.

While smart homes and similar ubicomp systems aid lives to a great extent, most of these systems and services collect huge amounts of sensitive user data and metadata in order to transact with remote services and unauthorized external parties [9, 6, 7], in order to provide an optimal service. The data is often comprised of information corresponding to users' habits, preferences, and daily activities. In other words, the systems collect what amounts to intimate details of individuals. Even if the data is encrypted, the metadata can be used to learn a lot about an individual [3]. Data corresponding to an individual can generate profiles of people with decent accuracy. Besides the ethical implications of such data collection and transaction schemes, most often the users are unaware about this. Additionally, even if the user has some knowledge, the process to control or influence the access the services have to sensitive user information is complicated and obscure.

Such unregulated, non-transparent pervasive systems have the capacity to turn into surveillance machines and hence raise huge concerns about personal data privacy. While privacy is a broad entity, in this work we employ the definition of information privacy as being *the user's ability to control whether and how personal data can be gathered, stored, processed or selectively disseminated* [5]. An ideal system/service should be transparent to the users regarding the data-access, collection and transfer and provide ways to control it. The consent and permissions an individual grants a certain service or system indicates the legitimate access a service has to sensitive user resources. In contrast, current user consent protocols for internet-connected devices are usually a *Hobson's choice*, i.e. take it or leave it proposition, which is agreed to by the user at time of installation. This is often done for (i) simplicity; users must only read and agree to one long, legal document to access all features of a service, and (ii) lack of fidelity to dynamically request consent in a system. However, consent, permissions and thus privacy is hardly a constant entity; It is a variable, hugely dependent on the context of both, the requested permission and, the user [8]. Therefore, there is a need to review and modify the current permission systems and consent mechanisms, in order to make systems more privacy-abiding and user-empowering.

In this project, we propose a universal and inclusive interaction model to enable a transparent feedback mechanism to better empower users with control to their personal data. The system can be integrated with smart systems and require low computational resources, such that it can be implemented as a stand-alone system working out of a Raspberry-pi. Additionally, we focus on a non-intrusive system design, which seamlessly blends into the household without much disruption, yet it is there to always remind the users of their data privacy levels. Finally, we hope that this project lays the foundations of privacy in HCID and initiate a discussion surrounding future smart and distributed systems with respect for user-privacy embedded in their design. The report is organized as follows: in the next section we describe the proposed method and the motivation and reasons behind its design and implementation, following which we present the implementation of the system. We conclude with a discussion on the advantages and the limitations.

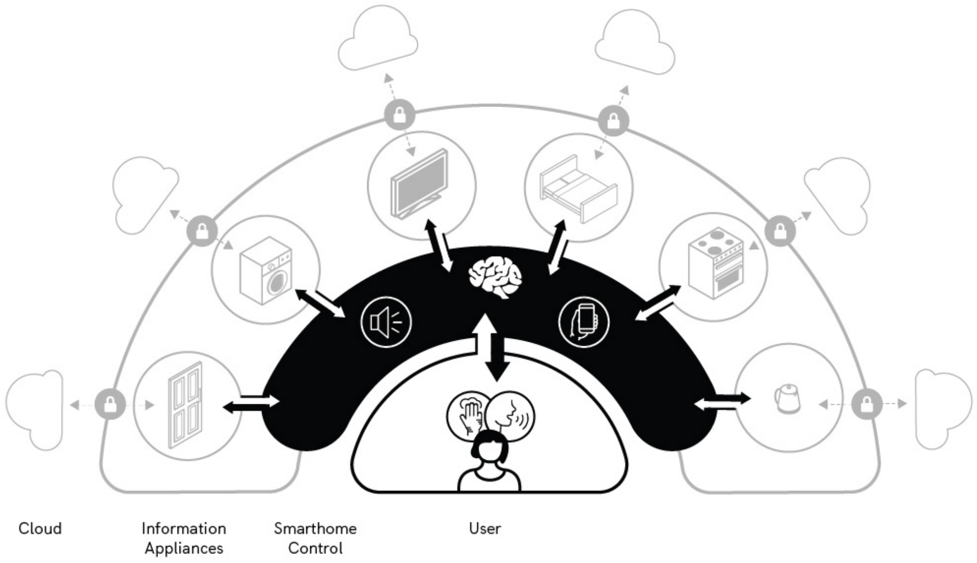

Figure 1: Conceptual representation of the proposed system.

Proposed Method

Most systems today restrict access of applications and services to sensitive user data through the permissions system [10]. For instance, the Android OS has employed the runtime install option (Android 6.0 onwards) to provide the users with the choice to grant or revoke application access to sensitive resources. Consequently, transparency and management of permissions is a first step towards restoring and managing personal data privacy.

Fig. 1 provides a conceptual illustration of our system. The proposed system poses as an additional layer, which is labeled as *Smarthome Control* in the figure. The layer can run via a raspberry-pi or any other physical computing device. In the course of this project, we created a proof-of-concept (POC) privacy and permission check system which alerts and updates the user about the access the applications, services and systems have to sensitive user data, hence the name *Data Alarm*. This system interacts with the various smart systems in the house and provides a multimodal feedback to the user. The core features of the system are explained in the following sections.

Multi-modal feedback system
The data alarm system uses both visual and audio output, and accommodates audio (voice) input in addition to the system input from a devices' permissions settings. In this

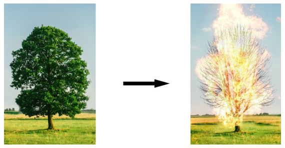

Figure 2: Changes in data privacy level(good → worst) represented through transition in tree image: tree in summer → tree on fire.

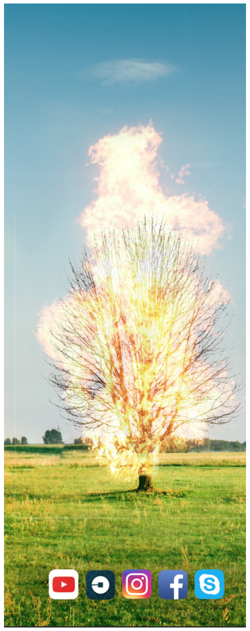

Figure 3: Display when user requests names of applications with access to an arbitrary data-resource.

manner, the system is multi-modal, and is designed inclusively to provide a sense of data security to a wide percentage of possible users. Also, the choice of the mode for feedback is situational; for instance, users may not want the system to use audio output to warn them about a data privacy issue while the user is engaging a room full of people. Hence, we incorporate more subtle visual and more specific audio modes to maintain a flexible feedback mechanism which is available to the user at all times in the mode they wish [2].

Intuitive and non-intrusive Display

Data protection in a smart home scenario has additional context-related complexities. Displaying any sort of data, especially privacy data, in a home setting can be jarring and intrusive. Therefore, the data alarm blends a regular household fixture of a painting hanging on a wall with the display of privacy data. To achieve this, the data alarm uses unobtrusive and intentionally placating imagery of a line of trees in a field. Trees have natural stages of change, which have existing cultural understandings. A tree in spring and summer is viewed to be healthy, and the data alarm leverages this understanding to represent healthy data privacy. That perception of health changes as a trees' leaves change color and fall off, and especially when a dead tree catches fire, as is represented in the Figs. 2, 4.

Each tree represents the data permissions of one data feature (camera access, microphone access, location data, telephone access, and SMS), and each successive stage in the declining health of each tree represents a greater risk to the privacy of each data feature.

System Implementation

Overview

The block-diagram of the system is depicted in Fig. 4; the dashed area encapsulates the modules developed in this project. We observe from the figure that the Raspberry-Pi is the central processing device. Note that the raspberry-pi can be replaced by any other computing devices with the required memory and capability for wireless communication. The central processor runs a dedicated socket and server (Socket 1) to communicate with the smart devices. In this POC implementation, for simplicity we used a single smart device (Android tablet) which is a host to multiple applications and services. However, the current implementation can be easily extended to accommodate multiple smart devices. We created an application, which runs in the background of the device, which transmits the applications' access permissions to sensitive data. For the moment, we have restricted the bounds to 5 applications and 5 groups of sensitive resources (further details in the following section). On receiving the access permissions from the smart device, socket 1 calls the back-end algorithm. This algorithm processes the permissions and computes the image to be displayed. The display device in this work is a computer monitor, which is connected to the central processor. This could also be replaced by a wireless picture frame, if accessible.

For audio modality of the system, we employed the Google Assistant SDK [1]. This required the use of Dialogflow, which is a natural language processing (NLP)-based conversational interface which aides in building voice-based conversations with systems. While the NLP took place automatically in the Google cloud, the required back-end codes for our application were implemented in our local central processor. Thus, the central-processor required to run a second dedicated server to communicate with the

Informal user based evaluation:
1. Majority of the participants found the visual display of the data privacy levels intuitive and could instinctively associate the transition in tree images to changes in the data privacy levels.

2. Besides the meant purpose of the data alarm, multiple users found the multi-modal system fun and approachable.

3. A useful suggestion from an user was to extend the system by incorporating methods to inform the user about why and how data is being accessed by a certain application.

4. An user expressed concern over the static nature of the system and possibility to address the contextual nature of personal data privacy.

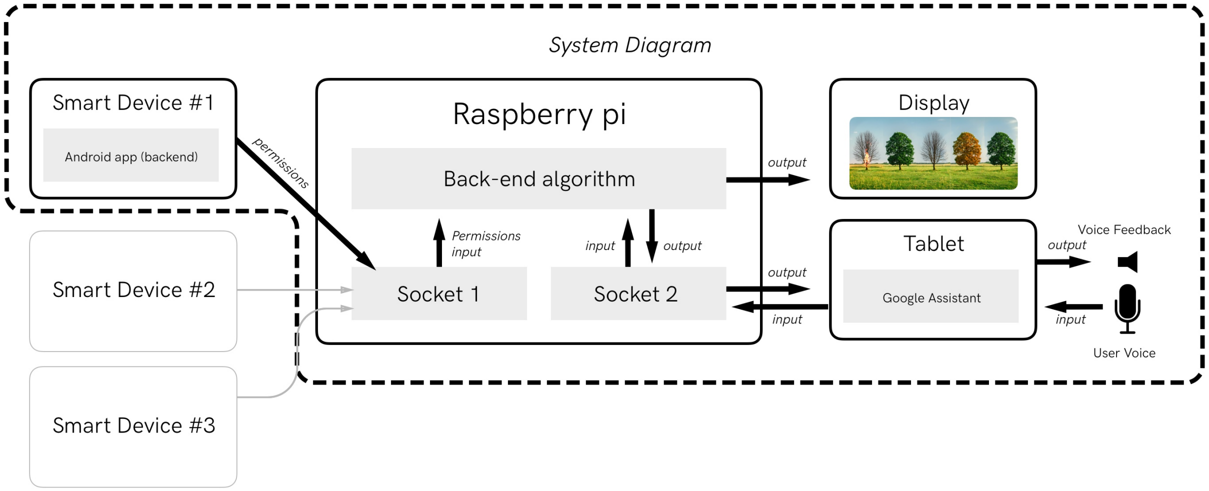

Figure 4: Block diagram of the Data Alarm

Google assistant. Additionally, to communicate with the assistant, we employed a tunneling application which provided a public URL to our local server (socket 2) and re-routed requests between the local server and the cloud-based Google assistant through the public URL. The host for the Google assistant was a tablet; when requested for privacy level as follows: *"Data alarm, give me an overview of privacy level"*, the assistant first recognizes the words spoken (in the Google cloud). Following this, the recognized words are transmitted to the local server, which calls the back-end algorithm to create the appropriate response. The response is transmitted to the cloud, which then uses TTS (text-to-speech) to generate the speech response, which is then generated through the tablet.

Feedback system
In this work, we restricted the POC to 5 applications: Facebook, Instagram, Skype, Uber, and Youtube. Additionally, we considered 5 sensitive data resources (in our opinion): camera, microphone, location, SMS, and telephone. Human senses are unequal in terms of cognitive bandwidth requirements [4]. In other words, humans can retain more information via visual inspection than through audio inspection. Therefore, to design a practical and useful feedback system, specifically a conversation based system, we should optimally balance between the amount of information released and maintaining a useful a privacy-management system which can provide the user with any information requested. To this end, we designed

the data alarm to operate in 3 levels of granularity, as explained below:

1. **Status overview:** This level provides the user with a general idea of their current privacy level, i.e. an intuitive measure for how many applications have access to a data resource. For instance, if the user says: *Give me an overview of my data privacy level.*, the back-end algorithm computes the status, S as follows: $S = \frac{1}{M}\sum_{i=1}^{N}\sum_{j=1}^{M} P_{ij}$, where P is the permissions matrix, $P \in \mathbb{Z}_2^{NXM}$, N is the total number of applications/services and M is the number of sensitive data resources being considered. The resulting voice response could be *Your data permissions and privacy is xxxx. Do you want to hear more?* ; *xxxx* is based on the value of S.

2. **Status details:** This level presents the number of applications having access to each data resource. E.g. User: *I want details of my privacy level.* Response: *Certainly! here goes your detailed status! x1 applications have access to your camera, x2 applications have access to your microphone....* x1, x2 depends on the vector $d = \sum_{i=1}^{N} P_{ij}$

3. **Status of specific resource:** This is the highest level of information a user can retrieve from the system, wherein, the system spells out all applications which have access to the specific resource requested by the user. E.g. User: *Give me list of applications with camera access.* Response: *The apps with access to your camera are a, b, c ....The App icons are displayed in the picture frame.* To aid information absorption, we also use the visual display when this level of feedback is invoked by the user. We do this by displaying the logos of applications with access to the hardware resource underneath the tree corresponding to the resource, as shown in Fig. 3.

An additional advantage of the proposed system is it can be extended to enable user adaptation and customization. This accounts for the fact that data privacy is a subjective concept and is perceived differently by different groups of individuals. In other words, while a certain level of access to sensitive resources might be acceptable by a certain group of individuals, the same level can be perceived as an abysmal breach in personal data privacy by another group. Therefore, there is a need to create systems which can adapt to the user requirements. The proposed system can attain this by changing the weighting function in the back-end algorithm, which can be further controlled by the user through voice input.

Discussion and Conclusion

In this work, we implemented the Data Alarm, which functions as a *privacy and permission status check machine* in a smart-home and a distributed-system setup, to empower users to better manage their privacy levels. Through an informal demo session, we received constructive comments on the feasibility of the approach, which are summarized on page 5. The advantages and limitations of the current approach are listed on page 6.

Contributors

1. Sneha Das: Implemented the prototype which included: android application, back-end algorithm, server-socket programs, Google assistant algorithm.

2. William Turner: Responsible for the entire design aspect of prototype which included: designing the

Advantages of the current approach:

(i) Multimodal user interaction.

(ii) Non-intrusive and intuitive feedback system.

(iii) Provision to customize and adapt to the users definition and expectation of privacy.

(iv) Easily extensible and can be added to existing smart-home infrastructure.

Limitation of the current approach:

The decision to use Google assistant for voice-user interaction has been limiting in the course of this work. While the decision was based on the high quality of Google NLP modules, this came at the cost of simplistic implementation, wherein we had to employ cloud-based processing, which not only increased the delay in the responses but also, the accuracy of responses suffered.

visual display, the voice conversations, and other aesthetic and interaction aspects.

3. Chandni Sharma: Executed the prototype in the Raspberry-Pi and took care of all hardware requirements of the team.

REFERENCES

1. Google Assistant SDK. https://developers.google.com/assistant/sdk/. (????). Accessed: 2018-05-20.
2. What is inclusive design? http://www.inclusivedesigntoolkit.com/whatis/whatis.html. (????). Accessed: 2018-05-20.
3. Kashmir Hill and Surya Mattu. The House That Spied on Me. https://gizmodo.com/the-house-that-spied-on-me-1822429852. (????).
4. Joseph Galen Lindley, Paul Coulton, Haider Akmal, and Brandin Hanson Knowles. 2017. Anticipating GDPR in Smart Homes Through Fictional Conversational Objects. (2017).
5. Gustavo López, Gabriela Marín, and Marta Calderón. 2017. Human aspects of ubiquitous computing: a study addressing willingness to use it and privacy issues. *Journal of Ambient Intelligence and Humanized Computing* 8, 4 (2017), 497–511.
6. Ewa Luger and Tom Rodden. 2013. An informed view on consent for UbiComp. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM, 529–538.
7. William Seymour. 2018. How loyal is your Alexa?: Imagining a Respectful Smart Assistant. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, SRC20.
8. Primal Wijesekera, Arjun Baokar, Lynn Tsai, Joel Reardon, Serge Egelman, David Wagner, and Konstantin Beznosov. 2017. The feasibility of dynamically granted permissions: Aligning mobile privacy with user preferences. In *Security and Privacy (SP), 2017 IEEE Symposium on*. IEEE, 1077–1093.
9. Primal Wijesekera, Arjun Baokar, Lynn Tsai, Joel Reardon, Serge Egelman, David Wagner, and Konstantin Beznosov. 2018a. Dynamically Regulating Mobile Application Permissions. *IEEE Security & Privacy* 16, 1 (2018), 64–71.
10. Primal Wijesekera, Joel Reardon, Irwin Reyes, Lynn Tsai, Jung-Wei Chen, Nathan Good, David Wagner, Konstantin Beznosov, and Serge Egelman. 2018b. Contextualizing Privacy Decisions for Better Prediction (and Protection). In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 268.