THEME ARTICLE: The Spectre of Malicious Computing.

Personalized Privacy Assistants for the Internet of Things

An Infrastructure for Notice and Choice in the Internet of Things

Anupam Das
Carnegie Mellon University

Martin Degeling
Carnegie Mellon University

Daniel Smullen
Carnegie Mellon University

Norman Sadeh
Carnegie Mellon University

As we interact with an increasingly diverse set of sensing technologies, it becomes difficult to keep up with the many different ways in which data about ourselves is collected and used. Study after study has shown that while people generally care about their privacy, they feel they have little awareness of – let alone control over – the collection and use of their data. This article summarizes ongoing research to develop and field privacy assistants designed to empower people to regain control over their privacy in the Internet of Things. Specifically, we focus on the infrastructure we have developed and fielded to support IoT privacy assistants. The infrastructure enables the assistants to discover IoT resources (e.g., sensors, apps, services and devices) in the vicinity of their users, and selectively inform users about the data practices associated with these resources. It also supports the discovery of user-configurable settings for IoT resources (e.g., opt-in, opt-out, data erasure) if there are any, enabling privacy assistants to help users configure their IoT experience in accordance with their privacy expectations. We also discuss how, using machine learning to build and refine models of their users' privacy expectations and preferences, we plan to developed personalized versions of our privacy assistants capable of selectively informing their users about the data practices they actually care about, and capable of helping them configure associated privacy settings.

INTRODUCTION

Information privacy is about giving people meaningful choices when it comes to the collection and use of their data and about giving them sufficient details about these choices to make informed decisions. In practice, even when browsing the web from desktop or laptop computers, few people find the time to read privacy policies, or exercise choice options available to them. Research by the authors, as well as by others, shows that users only care to be informed about a small fraction of the statements found in privacy policies.[1—2] Over the past decade, the challenge of informing users about relevant data collection and use practices has further exacerbated by the proliferation of smartphones. Reading privacy policies and exercising choices are further hampered by the small form factor of these devices, as well as the added distractions associated with many mobile usage scenarios. Nevertheless, current mobile operating systems do have centralized permission management functionality that provides users with some control over the permissions requested by their mobile apps. In contrast, the Internet of Things (IoT) as we know it today does not offer any equivalent functionality. Instead, users are often unaware of the presence of IoT technologies, as there is no uniform mechanism for discovering them, let alone accessing any privacy settings these technologies might support. For instance, there is no standardized way of determining whether an area is under video surveillance and what algorithms might be applied to the footage captured by cameras (e.g., facial recognition, facial expression recognition, scene recognition, measuring engagement with products), or whether sensors are being used to track a user's whereabouts based on his or her unique device ID. There is typically no mechanism for users to opt in or opt out of these data collection practices. Given all the above, it is no surprise that a November 2014 Pew Internet survey reported 91% of adults "agree" or "strongly agree" that consumers have lost control over how their personal information is collected and used by companies.[3] In the United States, the FTC has identified notice and choice as one of the vital challenges associated with the widespread deployment of IoT.[4] In Europe, as of May 2018, the General Data Protection Regulation (GDPR) requires the adoption of transparent practices and affirmative user consent when it comes to collecting personal data, which includes IoT technologies.

People's general sense of hopelessness when it comes to controlling the collection and use of their data, as well as emerging regulatory requirements such as GDPR, call for a new scalable paradigm that empowers users to regain appropriate control over their data. As part of their work in this area, the authors have been working on the development and evaluation of Personalized Privacy Assistants (Patent pending) (or PPAs). Our IoT Privacy Assistants are intended to help their users discover the presence of nearby IoT resources as well as their data collection and use practices, and any privacy settings associated with these resources. Using machine learning to build models of people's privacy preferences and expectations, we are working on personalized versions of these Privacy Assistants. PPAs can use their learned models to selectively inform their users about the data practices they really care to be notified about, and also help them configure associated privacy settings. An early version of this technology has been demonstrated in the form of PPAs that help their users configure permissions required by the mobile apps on their Android smartphones.[5] These PPAs have been successfully piloted by actual Android users on their personal devices as part of their regular activities.

In this article, we discuss equivalent functionality being developed for the IoT and focus in particular on the underlying privacy infrastructure required to support PPAs for the IoT. Our work is informed by studies showing that people are not always comfortable with the data collection and use practices associated with IoT technologies, and that machine learning techniques can be used to help predict their privacy preferences and expectations.[6]

Our IoT Privacy Infrastructure revolves around the development of IoT Resource Registries (IRRs) where resource owners advertise the presence of IoT resources they deploy along with the data practices associated with these resources. The infrastructure includes web portals and resource templates intended to help resource owners populate entries in IRRs with minimal effort. IRRs advertise the data collection and use practices of registered resources, enabling PPAs to discover them, and to selectively inform their users about the practices and choices associated with them (e.g., opt in, opt out, erasure). The article further discusses deployment and management options associated with this infrastructure. In addition, we report on the deployment of our IoT privacy infrastructure on two university campuses in the United States.

RELATED WORK

An early form of privacy assistants for ubiquitous computing was first described in the early 2000s by Langheinrich, who proposed the use of beacons and service discovery protocols to advertise the privacy practices of data collection services. In combination with privacy proxies and a centralized privacy-preserving database, this infrastructure was intended to tightly control the flows of personal information.[7] Around the same time, Sadeh et al. described their use of semantic web technologies to capture and enforce rich collections of privacy preferences in mobile and IoT contexts in their MyCampus project.[8] Sadeh and colleagues also reported on early work to learn people's privacy preferences to automatically or semi-automatically configure privacy settings such as location sharing settings.[9] Lilian Edwards also suggested the need for smart systems that can semi-automatically help users make privacy choices in IoT.[10]

Individual privacy preferences and expectations have been identified as factors that influence whether one will approve of sharing their personal information. Other factors include transmission principles and social norms.[11] Multiple studies have been conducted to identify individual factors, which include not only what data is shared, but more importantly with whom it is shared.[12—14] Other factors include the purpose of data sharing, how long the data will be accessible, and how it will be processed. Still, the availability of this information about important factors does not solve a fundamental problem; the amount of privacy decisions that need to be made increases with the diversity of new sensors, services, and apps that collect data. Therefore, a new paradigm in privacy research looks at how machine learning can be used to simplify privacy decision making through recommendations. Liu et al. have shown that recommendations based on clusters of like-minded users and predictive models of people's privacy preferences work to the users' satisfaction in the context of mobile app privacy.[5]

In a recent crowd-sourced vignette study, we asked participants to assess their comfort and interest in receiving notifications with respect to different hypothetical IoT-related scenarios.[6] These scenarios described up to eight different factors about what data is collected, from where, for what purpose, and the data retention period. Results of this study suggests the existence of some privacy norms in some contexts such as differentiation between private (at home) and public (in a library, or department store) contexts, with different contexts typically leading to very distinct privacy decisions by users. Other contexts however, such as the collection of sensitive data in the workplace for purposes such as saving energy, seem to give rise to fairly diverse privacy preferences, with some users feeling comfortable sharing their data and others not so much. In addition to modeling privacy preferences (when it comes to identifying when people are comfortable disclosing potentially sensitive data), we are also examining how to use machine learning techniques to build models of the types of practices people want to be notified about, how often, and in what manner.

OVERALL ARCHITECTURE

In the smartphone world, users control the apps they install on their devices and have access to unified permission management functionality, where they can review and control the permissions granted to apps. The situation in IoT is quite different. Here, users interact with technologies they often did not deploy and are seldom even aware of. This lack of awareness, as well as a dearth of settings available for users who do not own or manage these IoT resources, makes 'Notice' and 'Choice' a significantly more difficult proposition. IoT users generally do not know what devices are around them, what data they collect, and what happens to that data. To remedy this situation, we need an infrastructure that supports the discovery of nearby IoT resources and their data practices. By "nearby" IoT resources, we mean IoT resources that collect data in our physical vicinity. IoT resources may include IoT devices (e.g., smart home assistants, autonomous cars), IoT services (e.g., indoor location tracking systems, video analytics services connected to smart cameras) or IoT apps (e.g., smart TV remote apps) that collect and use data about us. Along with the discovery of these resources, the infrastructure also has to support the discovery of information about the data these resources collect, and how this data is used. Equally important are settings that these resources may expose to users, such as opt-out settings, opt-in settings, and more. Below, we introduce such an architecture, which we have implemented and deployed at two university campuses in the United States.

We highlight the three main components of our IoT privacy infrastructure:

- IoT Resource Registries (IRR)
- Personal Privacy Assistants for IoT (PPA)
- Policy Enforcement Points (PEP)

We first describe the functionality of each of these components. We then illustrate how these components interact with each other, to notify users of the existence of nearby sensors and privacy settings, and support the configuration of these settings.

Internet of Things Resource Registry (IRR)

We have developed an IoT privacy infrastructure that is intended to be open and scalable. The infrastructure is based on an open and distributed architecture in which any number of actors may be involved in the deployment and management of IoT resources. Resource owners include corporations deploying smart locks, HVAC systems, room presence systems, audio/video equipment, scheduling systems, and location tracking in office buildings. Cities may deploy public resources such as airborne or fixed traffic monitoring services, computer vision based crime reporting systems, and public health monitoring systems. Malls, stores, and restaurants may deploy IoT systems for security purposes, as well as marketing. Today, in many homes we see smart door locks, surveillance cameras, thermostats, and voice-enabled home assistants. These and other IoT environments all feature the deployment of connected devices designed to capture potentially sensitive data. Thus, there is a need for an infrastructure that can, at the very least, inform users about these data collection processes and ideally also offer them with some level of control over these processes.

From an economic standpoint, IoT resource owners also have two major incentives to participate. First, compliance with existing and upcoming regulations. In particular, we believe that regulatory frameworks such as the EU's GDPR effectively demand the deployment of infrastructures such as the one we propose. Second, our infrastructure can also be used to advertise IoT resources. For instance, as a user enters a building, the building's IRR can be used to advertise the existence of an app that helps users navigate through the building or find different resources in the building. In other words, the infrastructure we propose to advertise the presence of IoT resources and their data collection and use practices, can also be used to advertise useful functionality, services, resources or even establishments.

IRRs allow IoT resource owners to publish and disseminate descriptions of their IoT resources. These descriptions include the data practices of these resources. An IoT resource, for example, can be an app, a service, a single sensor, a virtual sensor aggregating multiple sensors, as well as any infrastructure element that might collect and/or use user data. The IRR acts as a location-aware lookup service that supports the discovery of nearby IoT resources. Device owners and IRR administrators access IRRs through a secure web portal.

Typical resource entries include information about the party that collects data, the purpose of the data collection, retention period, granularity of data collection, and third-party data sharing (if any). Resource owners can also advertise control options that enable users to restrict how their data is used, such as the ability to opt in, opt out, erase data, restrict the retention period, define who the data can be shared with, restrict how it can be used, define whether it needs to be anonymized or aggregated, and more. These settings, where made available, are paired with specifications of APIs and control endpoints that users can configure through their privacy assistants.

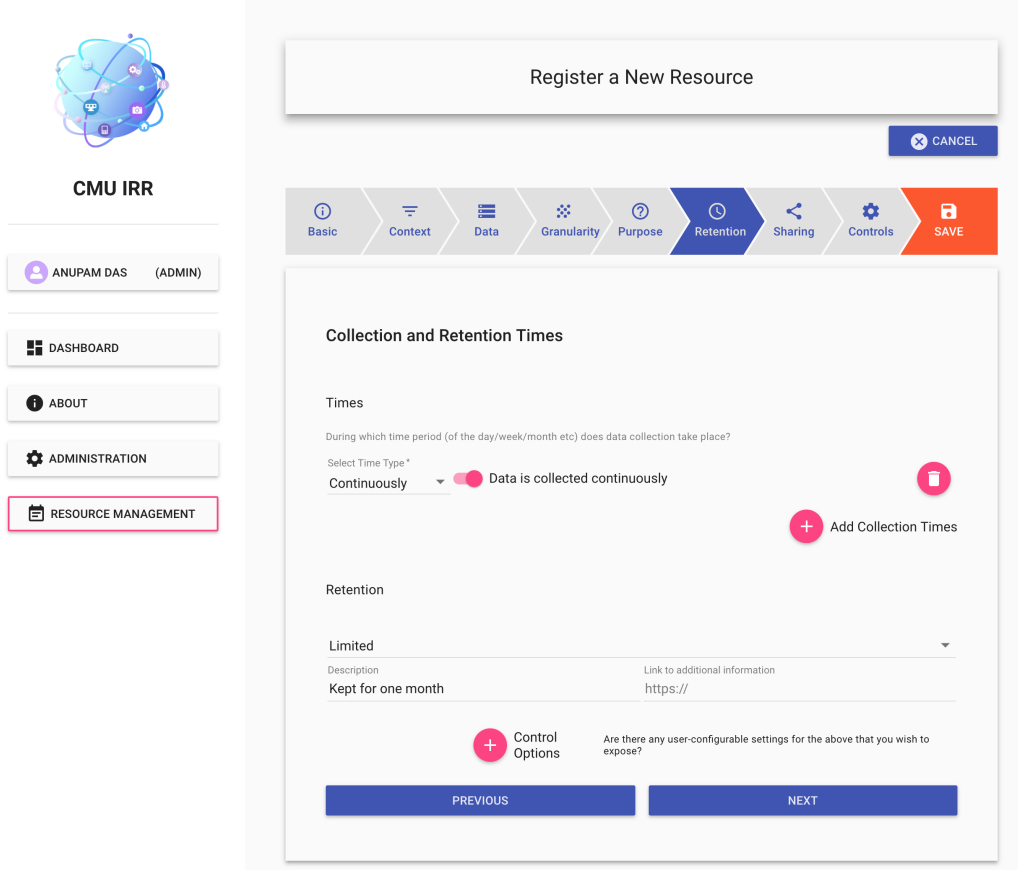

Figure 1: Screenshot of the IRR portal. It provides a wizard for specifying privacy practices. It allows resource owners to specify the location where a resource is deployed along with data practices such as the purpose(s) associated with a given data collection process, granularity of the data being collected, retention, the parties with whom the data is shared, and more. It also exposes any user-configurable privacy settings that may be associated with a given resource.

Figure 1 shows a screenshot of the different policy-related information captured through the IRR user interface. In particular, the top of the screen shows how the resource registration "wizard" guides a resource owner through a succession of steps (or workflow) to define the data practices associated with an IoT resource. For the sake of accommodating a wide range of resource owners and regulatory requirements, the wizard makes minimal assumptions about the particular fields the resource owner needs to fill to specify a valid resource. Most of the available fields can be optional, though IRRs can also be configured to require a more extensive collection of fields, as mandated by specific regulatory requirements or corporate policy. Many of the fields come with predefined options, designed to expose commonly accepted taxonomies used to characterize details of many data practices. For example, predefined options for data retention range from "ephemeral" to "limited" to a specific time period, all the way to "unspecified". This interface is designed to broadly facilitate the registration of resources in IRRs, but is primarily targeted towards professional users, such as system administrators, building managers, and the like. For casual or home users interested in deploying and advertising the presence of commercial off-the-shelf IoT resources in their personal spaces, our infrastructure supports the creation and consumption of vendor-generated "resource templates" that predefine the specific data practices of particular commercial products. Using these templates, vendors predefine the practices and capabilities of their products, reducing the burden on end users (e.g., a homeowner deploying a commercial off-the-shelf voice-based home assistant). When using templates, end users need only to enter deployment specifics, such as the place in their home where the resource is located. At the time of writing, we have created templates for a dozen popular IoT resources, including Amazon Echo (with Alexa),

Google Smart Home, and Nest Cameras. Our hope is that over time, vendors will develop product templates of their own.

Our infrastructure is designed to support the deployment of any number of IRRs. Different IRRs can be managed according to different policies, by different groups or organizations. Some IRRs may be designed to advertise the presence of IoT resources in corporate buildings, smart cities, malls and may be very tightly managed. Others may be managed by various communities and may have looser management policies and lighter vetting procedures when it comes to the registration of resources. We envision different users using their PPAs to filter out different IRRs and resources according to different possible criteria, e.g., based on the types of entities managing different IRRs or the policies of these IRRs. IRRs may have overlapping coverage areas, though some IRRs may be viewed as more authoritative over an area than others. For example, an official IRR for a university campus might be considered more authoritative than an IRR run by a hobby or student organization. Conceivably, some IRRs may charge users for advertising their resources as a way of generating revenue and/or as a possible approach to reducing spam.

Once an IRR is set up, the availability of the IRR can be discovered through centralized directories of registries covering different geographic areas. Our infrastructure also supports the local advertisement and discovery of resources, without relying on directories of IRRs. Such decentralized protocol (e.g., via Bluetooth beacons) can be used to support the discovery of mobile resources (e.g., an autonomous car, a drone flying overhead, or a nearby smartphone with an enabled microphone).

As discussed earlier, the IRR infrastructure itself can be managed at different levels. The central directories of IRRs can be curated by different parties to determine which IRRs become publicly available. This process is comparable to ICANN and authorities regulating domain names on the web. This design also allows multiple levels of directories, e.g., if the proliferation of IRRs for a given area warrants some reorganization.

Individual IRRs may have one or more administrators responsible for vetting new resource registrations. Others may have resource owners who publish their IoT resources in their own private IRRs. IRRs may also be completely unmoderated, allowing anyone to advertise IoT resources to potential users. The nature and governance of an IRR (e.g., whether it is strictly controlled, versus whether it is open) may call for different combinations of management functionality such as functionality to determine which resources get published and prevent abuse.

PPA for IoT

The PPA for IoT is an app running on a user's smartphone which aims to assist users in discovering IoT resources and services available in their vicinity. It retrieves resource listings from IRRs relevant to a user's current location, and uses their content to produce a privacy notice. The PPA lists resources registered (see Figure 2) in the IRRs relevant to the user's current location and informs its user about each resource's functionality, ownership and data practices.

Just as we have demonstrated for mobile app permission PPAs, we envision IoT PPAs capable of learning the privacy preferences of their users over time to ensure they are notified about practices they actually care about and to help them configure available privacy settings, if any. Armed with personalized notification preference models, PPAs will be able to selectively decide when, how often, and what to show to their users about nearby IoT resources. Modeling privacy preferences will also enable PPAs to detect mismatches between a user's privacy expectations and the practices of the resources with which they interact. Identifying such mismatches can in turn be used to selectively warn individual users, and/or semi-automatically adjust available privacy settings for them. We envision this to be similar to approaches developed to build models of the types of data collection and use practices different users expect,[1] as well as models of the data practices with which different users are comfortable. The latter has been demonstrated using both clustering techniques and other machine learning techniques to build models of people's privacy preferences when it comes to granting permissions to different mobile apps.[5—6] We would expect the availability of user-configurable privacy settings (e.g., opt-in settings) for IoT resources to become more prevalent over time, in part because they are mandated by new regulations such as GDPR. PPAs

could then also be used to semi-automatically configure privacy settings on their user's behalf, where such settings are made available by registered resources.

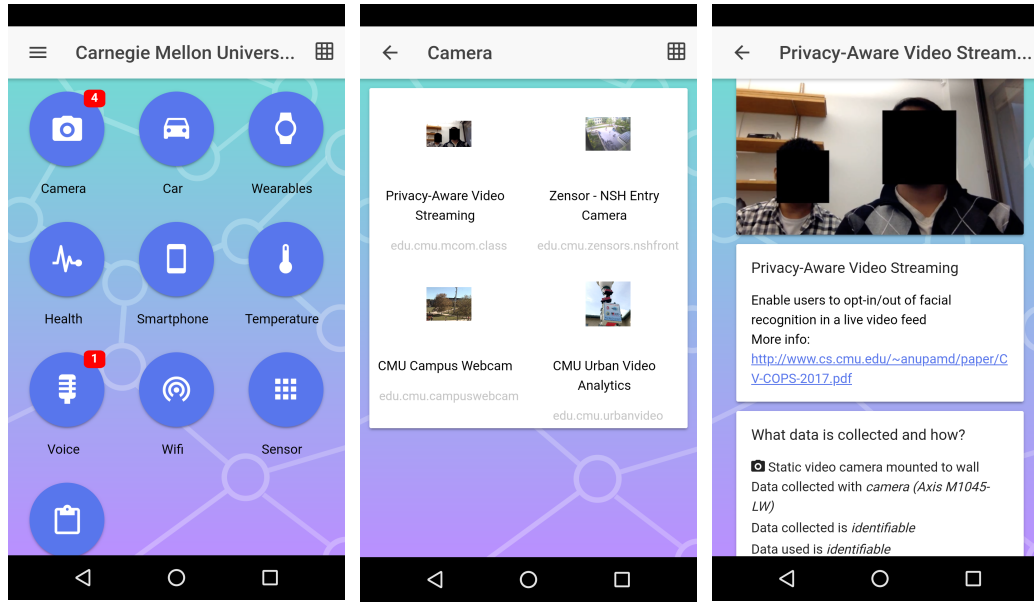

Figure 2: PPA for IoT app. It lists the resources available in the user's vicinity (left and middle), as well as details about the data collection practices of a particular resource (right), including available options for the user (not shown here).

It goes without saying that one would expect PPAs to only use the data they collect for the purpose of notifying users of the presence of IoT resources, and helping users adjust available privacy settings. In other words, PPAs should be provided by trusted third parties that are committed to not using data they collect for any secondary purpose. Optionally, PPAs could also provide users with regular summaries of what data has been collected about them, and for what purposes. Eventually, they could also engage in dialogues with their users to ensure that they remain aware of how much data they have leaked and to motivate them to possibly rethink their privacy preferences. It is important to carefully design such technology to prevent information overload or notification fatigue. Accordingly, we are conducting human subject studies aimed at informing the design and refinement of this technology.

Policy Enforcement Point (PEP)

New regulations such as GDPR, COPPA, GLBA, and CalOPPA (at least under some interpretations) require IoT resource owners or data collectors to expose different privacy settings to their users. In such contexts, there is a need for policy enforcement functionality responsible for both storing users' preferred privacy settings and enforcing those settings. For example, in the context of deploying cameras equipped with facial recognition, one possible user-configurable privacy setting would be to allow individual users to opt out of facial recognition during specific times of the day or at a specific location. We envision some IoT resources coming with their own policy enforcement functionality. At the same time, we also expect other IoT resources to require external policy enforcement functionality, for instance in the form of a PEP. The PEP would be responsible for controlling the collection and/or use of user data accordingly to user-configurable privacy settings advertised by IoT resources via IRRs. These settings would typically come in the form of APIs (e.g., opt-in, opt-out APIs).

Our PPAs are designed to allow users to configure privacy settings for different IoT resources when such settings are available. Privacy settings selected by users via their PPAs for a given IoT resource can be sent to the corresponding privacy enforcement API, as advertised by an IoT resource in its IRR. The policy enforcement functionality, whether embedded in the IoT resource or

whether implemented as policy enforcement proxy, is in turn responsible for enforcing the privacy settings selected by users via their PPAs. Our infrastructure includes configurable policy enforcement functionality that includes simple RESTful APIs to enforce privacy settings. The URLs and availability of these RESTful APIs can be advertised in a resource's IRR entry, and the advertised API can be directly accessed by a user's PPA (e.g., to opt in or out of some data practices).

Interaction Among Components

The interaction among the different infrastructural components is shown in Figure 3. As shown in the figure, IRR resource owners first register their IoT resources with a given IRR (the IRR directory, in this example, lists public IRRs). Access to the portal and administrator privileges are controlled through an authentication system. An IRR resource owner can use predefined templates to describe their IoT resources. Once IoT resources are registered with an IRR, users can rely on their PPA to discover the resources in their vicinity. PPAs can also help users configure any available privacy settings by brokering access to APIs that interface with the PEP enforcing settings for a resource. All of these parameters are advertised in the IRR entry for that resource. For example, the PPA can expose a facial recognition opt-out API, advertised in the IRR entry of a smart camera system. Perhaps this resource is in a mall, and used for marketing. When a user in the mall opts out, the smart camera resource's PEP ensures that each user's privacy settings are properly applied to the data streams coming out of the camera system, for example, preventing their face from being recognized.

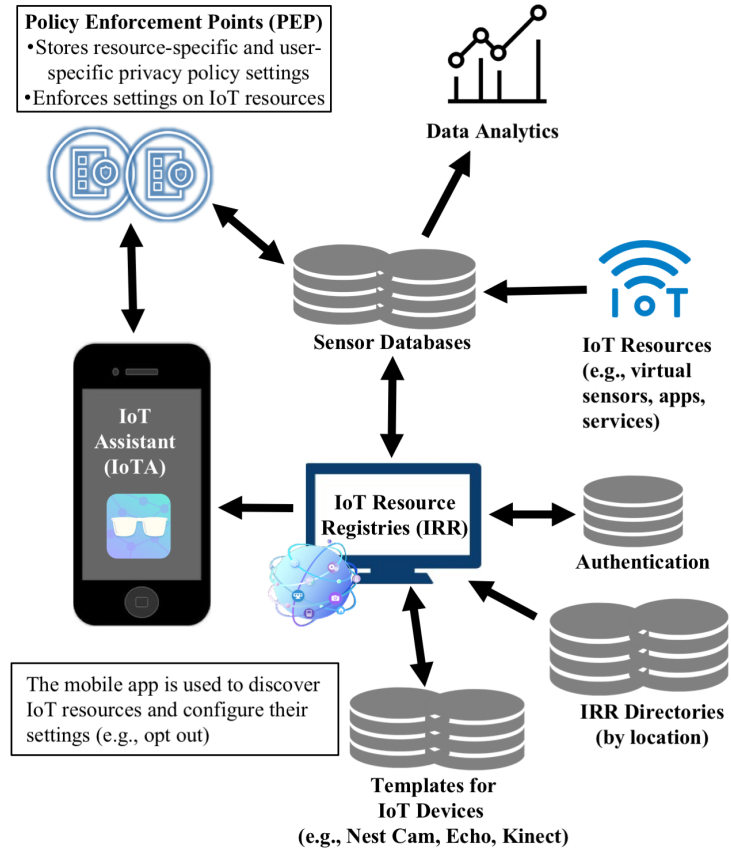

Figure 3: Interaction among components of our proposed system. The privacy assistant discovers IoT resources through IoT Resource Registries (IRR), and preferences are enforced through Policy Enforcement Points (PEP).

It is worth noting, while we envision an extensible collection of privacy controls, a simple generic version of a policy enforcement point which only supports opt-in and opt-out functionality can go a long way in providing users with a modicum of control over the collection and use of their data by IoT technologies – something that is not available today.

DEPLOYMENT SCENARIOS

We envision our infrastructure to be ubiquitous and easily deployable to a variety of public and private settings. In this section, we describe the deployment process that different IoT resource owners would typically face to register their resources with an IRR.

Retail Setting

Suppose Jessica, a small business owner, runs a high-fashion clothing shop. In order to track her customers for marketing purposes, she has installed a smart camera system in her store. The system uses facial recognition and behavioral tracking to determine what items her customers linger around, indicating their potential interest. The system contains a database of known faces and associated contact information. When the system recognizes a customer interested in a particular item, it sends him/her a promotional message. Jessica's shop is located in a mall which already has an IRR covering the entire location, so she registers her smart cameras in the mall's IRR. This enables her to notify customers about the use of smart cameras and also allows her to expose an opt-in privacy setting for facial recognition as the camera system is configurable.

Corporate Setting

Jim is an IT administrator for an enterprise that employs several hundred employees at several sites across the country. He is situated in a shared office building, housing other companies on other floors. Jim is responsible for overseeing security for this office branch. The enterprise decided to upgrade its security by installing new security devices around the office. Jim deploys facial recognition cameras, magnetic door locks with smart card and biometric authentication, and alarm buttons with two-way audio recorder intercoms that connect with security guards. Company policy mandates informing employees about the presence of devices that may collect personal information. The company uses strictly curated IRRs. He requests a new IRR, overseeing the floors where Jim's branch office is situated. After Jim's credentials are verified, he is allowed to open the IRR link and enters the details for the new IoT resources that have been deployed around the office.

CAMPUS DEPLOYMENTS

Thus far, we have developed three mobile applications and several underlying IoT services that are modeled as IoT resources in IRRs. Two IoT applications are available on Carnegie Mellon University's campus (CMU Friend Finder, Automated Class Attendance Tracker), and one on the University of California Irvine's campus (Indoor Navigation Concierge). Both campuses are equipped with indoor location tracking services using WiFi access points and Bluetooth beacons, each modeled as IoT resources in IRRs deployed on each campus. WiFi access points offer somewhat coarse grained location data (e.g., location distinguished by building, wing, or hallway). Finer-grained location data is based on Bluetooth beacons. Depending on the number and density of beacons that are deployed in a given area, Bluetooth beacons can be used for location detection precise enough to distinguish between individual rooms. In our deployment, PPAs are able to discover these services and apps and notify their users about their presence and data practices. To simplify user interaction with the location tracking services and the apps built on top, the IRRs advertise associated privacy controls (in this case opt-in and opt-out settings) that can be discovered by users via their PPAs. When users configure these options, their settings are automatically sent to APIs associated with policy enforcement functionality as advertised in the IRRs.

A second application we have implemented uses facial recognition technology to automatically detect and record attendance for university lectures.[15] Participants register their face with the application using their phone. Once registered, as they walk past a camera when entering the lecture room, their attendance is recorded. Lecturers and students may use these records to keep track of who attended the class. Similar to applications that use the location tracking service, users can use the PPA to change their privacy settings for the attendance tracking system. This allows users to opt in or out of the tracking, during the course of the semester. The application uses the same policy enforcement server as the location tracking service, which controls the facial detection processing service that the attendance tracking relies on.

The deployments helped us improve our infrastructure as well as refine the design of our IRR and PPA components. For instance, we introduced the template functionality after realizing that it would be unrealistic to expect a lay user to register commercial off-the-shelf IoT devices from scratch. We also determined that, while some organizations will wish to deploy IRRs on their own servers, many others will just want their IRRs to be hosted for them. Through our early interactions with IoT developers at both CMU and UCI we have also found that many developers lack sufficient understanding of privacy issues and struggle to articulate key disclosures about the collection and use of personal data by their technologies. The wizard was designed to help mitigate this situation by providing a framework that helps developers organize their thinking and identify key data collection and use practices they will want to disclose.

CONCLUSION

In this article, we have introduced a novel privacy infrastructure for the IoT and illustrated how this infrastructure can help support PPAs, designed to selectively inform their users about the presence of IoT resources and their data practices. While early deployments of this infrastructure suggest that it offers the promise of overcoming challenges associated with notice and choice in IoT environments, our current IoT Privacy Assistants are simple clients that help users discover IoT resources and their data practices, and help them configure available privacy settings. Over time, we plan to leverage machine learning models to reduce user burden, selectively determining what practices to inform users about, and help them configure available privacy settings.

The single most significant challenge is, without a doubt, to get a critical mass of technology providers (e.g., device manufacturers, app developers, virtual sensor providers) to agree on a common taxonomy to describe their data collection and use practices, and to adopt protocols such as the ones we have developed to support the advertisement and discovery of IoT Resource Registries and IoT resources. We hope that new regulations such as GDPR, rising consumer concerns about privacy, and the desire by at least some technology providers to differentiate themselves from their competitors based on privacy, will all contribute to creating the incentives necessary for this to happen. It would be unrealistic to assume that this will take place over night, or, even if such an approach is eventually adopted by key players that all IoT resource owners will suddenly disclose the presence, data collection and use practices of their resources. For this reason, in its most generic form, our infrastructure is agnostic when it comes to specific regulatory regimes and assumes that IRR administrators will decide for themselves which collection of fields they may want to mandate.

It also goes without saying that configuring privacy assistants to ensure they are as usable as possible will require significantly more research and empirical evaluation with users. While we have been fairly successful at modeling people's mobile app privacy preferences and have been able to accurately recommend many mobile app permissions settings,[5] IoT presents a significantly broader set of scenarios and contexts. Recognizing relevant contextual attributes, such as automatically turning off Alexa when kids visit your house and doing so in a privacy-preserving manner, will require more work. Aiming to develop privacy models which capture all possible scenarios is unlikely to be a realistic objective. Instead, PPAs will need to be able to entertain meaningful dialogues with their users to make up for the inevitable shortcomings of their models.

ACKNOWLEDGMENTS

This research has been supported in part by DARPA and the Air Force Research Laboratory under agreement number FA8750-15-2-0277 and by the National Science Foundation under grant SBE-1513957. The US Government is authorized to reproduce and distribute reprints for Governmental purposes not withstanding any copyright notation thereon. Additional support has also been provided by Google. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA, the Air Force Research Laboratory, the NSF, Google, or the US Government.

REFERENCES

1. A. Rao, F. Schaub, N. Sadeh, A. Acquisti and R. Kang, "Expecting the Unexpected: Understanding Mismatched Privacy Expectations Online," *Proc. 12th Symposium on Usable Privacy and Security (SOUPS)*, 2016, pp. 77—96.

2. J. Gluck, F. Schaub, A. Friedman, H. Habib, N. Sadeh, L. F. Cranor and Y. Agarwal, "How Short Is Too Short? Implications of Length and Framing on the Effectiveness of Privacy Notices," *Proc. 12th Symposium on Usable Privacy and Security (SOUPS)*, 2016, pp. 321—340.

3. M. Madden, "Public Perceptions of Privacy and Security in the Post-Snowden Era," Online Report, 2014, www.pewinternet.org/2014/11/12/public-privacy-perceptions/.

4. Privacy & Security in a Connected World, FTC Staff Report, 2015, www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf.

5. B. Liu, M. S. Andersen, F. Schaub, H. Almuhimedi, S. Zhang, N. Sadeh, Y. Agarwal and A. Acquisti, "Follow My Recommendations: A Personalized Privacy Assistant for Mobile App Permissions," *Proc. 12th Symposium on Usable Privacy and Security (SOUPS)*, 2016, pp. 27—41.

6. P. E. Naeini, S. Bhagavatula, H. Habib, M. Degeling, L. Bauer, L. F. Cranor and N. Sadeh, "Privacy Expectations and Preferences in an IoT World," *Proc. 13th Symposium on Usable Privacy and Security (SOUPS)*, 2017, pp. 399—412.

7. M. Langheinrich, "A Privacy Awareness System for Ubiquitous Computing Environments," *Proc. 4th International Conference on Ubiquitous Computing (UbiComp)*, 2002, pp. 237—245.

8. N. Sadeh, E. Chan and L. Van, "MyCampus: an agent-based environment for context-aware mobile services," *Proc. 1st International Joint Conference on Autonomous Agents & Multiagent Systems*, 2002, pp. 34—39.

9. N. Sadeh, J. Hong, L. F. Cranor, I. Fette, P. Kelley and J. Rao, "Understanding and capturing people's privacy policies in a mobile social networking application," *Personal and Ubiquitous Computing*, vol. 13, no. 6, 2009, pp. 401—412.

10. L. Edwards, "Privacy, Security and Data Protection in Smart Cities: A Critical EU Law Perspective," *European Data Protection Law Review*, vol. 2, no. 1, 2016.

11. H. Nissenbaum, "Privacy in Context: Technology, Policy, and the Integrity of Social Life," *Stanford University Press*, 2009.

12. S. Lederer, J. Mankoffand and A. K. Dey, "Who Wants to Know What when? Privacy Preference Determinants in Ubiquitous Computing," Proc. CHI '03: Extended Abstracts on Human Factors in Computing Systems, 2003, pp. 724—725.

13. M. Benisch, P. Kelley, N. Sadeh and L. F. Cranor, “Capturing location-privacy preferences: quantifying accuracy and user-burden tradeoffs,” *Personal and Ubiquitous Computing*, vol. 15, no. 7, 2011, pp. 679—694.

14. J. Lin, S. Amini, J. Hong, N. Sadeh, J. Lindqvist and J. Zhang, “Expectation and Purpose: Understanding Users’ Mental Models of Mobile App Privacy Through Crowdsourcing,” *Proc. 14th Conference on Ubiquitous Computing*, 2012, pp. 501—510.

15. A. Das, M. Degeling, X. Wang, J. Wang, N. Sadeh and M. Satyanarayanan, “Assisting Users in a World Full of Cameras: A Privacy-aware Infrastructure for Computer Vision Applications,” *Proc. 1st Workshop on The Bright and Dark Sides of Computer Vision: Challenges and Opportunities for Privacy and Security*, 2017.

ABOUT THE AUTHORS

Anupam Das is a postdoctoral fellow in the School of Computer Science at Carnegie Mellon University. His research interests lie in the domain of security and privacy with a focus on understanding and mitigating the security and privacy risks of emerging technologies. Anupam received his Ph.D. in Computer Science from the University of Illinois at Urbana-Champaign where he was a Fulbright Science and Technology fellow. Contact him at anupam009@gmail.com.

Martin Degeling is a postdoctoral fellow at the Horst Goertz Institute for IT Security at Ruhr-University Bochum. He received his Ph.D. in 2016 for his work on Privacy and Online Profiling from the University of Duisburg-Essen in Germany and worked as a postdoc on the Personalized Privacy Assistant project at Carnegie Mellon University. Contact him at martin.degeling@ruhr-uni-bochum.de.

Daniel Smullen is a Software Engineering Ph.D. candidate at the Institute for Software Research, part of the School of Computer Science at Carnegie Mellon University. His research focus is cybersecurity and privacy, with an emphasis on applied artificial intelligence, usability, pervasive computing, and software architecture for the Internet of Things. Daniel holds a Bachelor of Engineering (B. Eng., Honours with Distinction) in Software Engineering from the University of Ontario Institute of Technology in Oshawa, Canada. Contact him at dsmullen@cs.cmu.edu.

Norman Sadeh is a Professor in the School of Computer Science at Carnegie Mellon University, where he is affiliated with the Institute for Software Research, the Human Computer Interaction Institute and the CyLab Security and Privacy Institute. Norman is also co-founder and co-director of the Master’s Program in Privacy Engineering, director of the School of Computer Science’s Mobile Commerce Lab and co-founder of the School’s PhD program in Societal Computing. His research interests span mobile and IoT, cybersecurity, privacy, machine learning, artificial intelligence and related public policy issues. Dr. Sadeh received his Ph.D. in Computer Science from Carnegie Mellon University. Contact him at sadeh@cs.cmu.edu.