IoT Service Store: A Web-based System for Privacy-aware IoT Service Discovery and Interaction

Hosub Lee
Informatics
UC Irvine
Irvine, USA
hosubl@uci.edu

Richard Chow
Intel Labs
Santa Clara, USA
richard.chow@intel.com

Mohammad R. Haghighat
Intel Software and Services Group
Santa Clara, USA
mohammad.r.haghighat@intel.com

Heather M. Patterson
Intel Labs
Santa Clara, USA
heather.m.patterson@intel.com

Alfred Kobsa
Informatics
UC Irvine
Irvine, USA
kobsa@uci.edu

Abstract—Internet of Things (IoT) services collect and analyze sensor data to provide users with intelligent functionality tailored to their needs. However, users are often unaware of privacy risks relating to sensor data collection and the inferences possible from this data. Even if aware of the data collection and possible inferences, users lack ways to manage the collection, processing, and transmission of the data. To address this problem, we designed and implemented a novel web-based system called IoT Service Store (ISS) that allows users to easily browse nearby IoT services, understand the privacy implications of these IoT services, and control the collection and usage of sensor data. To better inform users about the potential privacy risks in using IoT services, ISS displays detailed information on what personal information might be inferred from the sensor data being collected. ISS also allows each user to give a rating or to view other users’ ratings regarding the perceived utility-privacy tradeoff for each IoT service. ISS is designed to communicate with IoT services to modify those services’ data collection and usage practices, according to a user’s privacy preferences. Using the preferred privacy settings in the proposed system, users will be more confident in their decisions of whether to subscribe to IoT services and less concerned with privacy risks in using the services.

Keywords—privacy awareness system, collaborative privacy management, personal information inference, Internet of Things

I. INTRODUCTION

The ubiquity and density of the Internet of Things (IoT) are rapidly increasing. In the near future, we will be surrounded by numerous sensor devices that unobtrusively and collaboratively extract myriad types of sensor data from the user’s environment. Much of this data will be information involving people’s presence, behaviors, or states. For instance, an automated student attendance system might infer the presence of a particular user based on the Wi-Fi MAC address of the user’s device and/or video collected by cameras associated with the system. Rapid advancements in big data processing and analysis will additionally allow IoT service providers to infer more diverse user-related information from this sensor data. The inferred information may include personal information of users, such as their attendance patterns, emotions, health condition, or even sexual orientation. Our proposed system is predicated on the notion that these and other types of personal information should only be collected and processed with a user’s meaningful consent.

A more fundamental issue is that users generally do not know about the existence of nearby sensors (except possibly for sensors in the user’s residence), let alone the nature of services operating these sensors. Also, it is hard to make sensor devices transparent to users (e.g., by adopting a conventional notice-and-consent model) due to the lack of natural communication channels between the users and services. Therefore, users need to have a unified way to discover nearby IoT services and understand the privacy properties of these services. In this way, they can become aware of the privacy implications of using IoT services of interest [1]. To be specific, users not only need to be informed about the whole process of sensor data collection, but also to understand what types of personal information might be inferred from the collected sensor data. With this information, users will be better positioned to define their own privacy settings (i.e., preferences) more confidently and therefore possibly give IoT services a way to respect these preferences [2, 3].

With these aims in mind, we designed and implemented a web-based system called IoT Service Store (ISS) for privacy-aware IoT service discovery and interaction. ISS is a web server that manages various privacy-related information of multiple IoT services. Here is how it works: First, each IoT service registered to ISS broadcasts its unique identifier (Uniform Resource Locator; URL) through Bluetooth beacon(s). The beacon-generated URLs are automatically detectable by nearby users’ Bluetooth-enabled Android or iOS smartphones. When a user clicks on a specific URL, she will be redirected to the designated web page, hosted by ISS. This web page contains a visualization of the IoT service’s privacy-related information, including its data collection policies, inferable personal information, and users’ collaborative evaluations of the service’s utility-privacy tradeoffs. Most notably, we developed a novel information architecture of both sensor data and personal information, as well as their relationships (e.g., sensor data A implies personal information B), in order to better inform users of the privacy properties of the service. Also, we adopted a five-star rating system for letting users collaboratively evaluate the service in terms of the balance between utility benefits and privacy risks.

ISS is designed to allow IoT services to comply with user-defined privacy settings. For instance, a user may utilize the automated student attendance system while allowing his/her Wi-Fi MAC address to be gathered but at the same time disallowing the collection of face photos. We plan to integrate ISS with several IoT services running on an operational IoT framework called TIPPERS [4, 5]. Through TIPPERS’s open data APIs, we can programmatically make “opt-out” requests for
specific sensor data to be collectible by the IoT service. Using the proposed system, users can subscribe to specific IoT services that offer and honor their preferred privacy settings, thereby helping minimize user-perceived inappropriate information flows.

In summary, our work makes the following contributions to the field of usable privacy and pervasive computing:

- We designed and implemented a novel web-based IoT service management system called IoT Service Store for privacy-aware service discovery and interaction.
- We developed an information architecture describing the IoT service’s privacy properties, specifically the types/attributes of raw sensor data and the personal information that may be inferred. Our architecture clearly separates the data from the inferences, motivated by the ever-increasing capabilities of machine learning.
- We designed interfaces to allow users to inspect the privacy properties of the IoT service, to collaboratively evaluate its potential privacy risks, and to accordingly adjust data collection practices.

II. RELATED WORK

One of the first privacy awareness systems for ubiquitous computing environments was proposed by Langheinrich [6]. The author proposed a system called pawS. This system aimed not only to allow data collectors to announce data usage policies, but also to provide data subjects (i.e., users) with the technical means of managing how their personal information is stored and processed by the service. The author assumes that all entities in an environment have their own privacy proxies, continuously running services that handle privacy-related interactions between the entities. Each user has his/her own personal privacy proxy which contains privacy preferences with respect to the multiple service privacy proxies that codify data collection and usage processes. Service providers describe their data collection policies using a machine-readable XML format such as a P3P privacy policy [7]. Using such a policy, each service provider can describe, for example, who is collecting data, what data is being collected, and for what purpose, in each case. Correspondingly, end users can express their own privacy preferences via a machine-readable preference language such as APPEL [8], which consists of a set of (updatable) rules. On pawS, all data collection and usage are therefore performed in accordance with the user’s privacy preferences.

Even though privacy awareness systems like pawS provide standardized ways to safeguard user privacy, service providers are still required to change their service infrastructure and/or reveal internal data handling practices, which could be a significant barrier to adoption. To handle this issue, Kolter et al. prototyped a user-centric privacy architecture that enables provider-independent privacy awareness in using Internet services [9]. The core part of this architecture is an online privacy community that lets multiple users post and share privacy-related information regarding the service. Just like in Wikipedia, users can edit diverse information about a specific web service (e.g., Amazon.com), including required amounts of personal data, practices of data sharing with third parties, adherence to the stated privacy policies, and the subject evaluation of privacy risks. Users can also share their personal privacy preferences with others. Inexperienced users may import the pre-defined preferences of a trusted privacy expert and utilize the imported preferences as their baseline choices.

Recently, researchers are realizing privacy awareness in a real world IoT environment. Mehrotra et al. are developing a privacy-aware IoT framework named TIPPERS and deploying it in Donald Bren Hall (DBH) at the University of California, Irvine (UCI) [4, 5]. In order to transform DBH into a smart environment, TIPPERS captures raw data from various sensors installed in DBH and makes the collected data publicly accessible through open data APIs. Third-party developers are then able to create various IoT services (e.g., indoor location awareness apps) on the TIPPERS framework. Regarding user privacy, service providers (or building administrators) need to advertise building policies which give detailed information about data collection procedures regarding the service. Then, end users set their privacy preferences and ask TIPPERS to enforce these preferences while operating the service. Both building policies and privacy preferences are defined using a custom JSON-schema for supporting data request and access from devices outside TIPPERS (e.g., user’s smartphone). The research team is currently developing a remote storage called IoT Resource Registries (IRRs) for administering building policies. They are also developing a smartphone app called IoT Assistants (IoTA) that notifies users about available building policies, thereby configuring privacy preferences, whether via interactions with the users or automatically.

The common goal of this body of work is making personal data collection as transparent as possible to users, thereby helping them make an informed privacy decision. However, it is still unclear how well users understand the implications of some data collection. For instance, not all users understand that the Wi-Fi MAC address is an identifier that may be used to track the location of the corresponding user device or its owner. Service providers can explicitly describe the implications of various sorts of sensor data collection; however, they may not cover all inferences of personal information based on the sensor data. For instance, it’s obvious that images from a camera might reveal users’ identity, but it’s less well-known that these images may also reveal their sexual orientation [13]. The IoT service may not rely on these other inferences and in fact the service provider may be oblivious to them. Nevertheless, for the sake of user privacy, the user should know about these inferences. Thus, our aim is the development not only of an information architecture for describing the privacy properties of IoT services, but also of user interfaces that efficiently convey possible inferences of data collected. To the best of our knowledge, our platform is the first to address user understanding of inferences and to define an architecture that distinguishes inferences from raw sensor data collection. In addition, this body of work has not considered collaborative privacy management strategy (e.g., crowdsourced evaluations of privacy risks) in improving users’ privacy awareness in IoT. Since Kolter et al. showed its feasibility in Web environments [9], we are also applying this approach in our work.
III. WEB-BASED PRIVACY AWARENESS SYSTEM FOR IoT

In this section, we discuss the design and implementation of our system in detail. We first describe an overall system architecture including software/hardware specifications, and then explain the functional details of the system.

A. System Architecture

Our system is designed as a client-server model (Fig. 1). Users can query diverse privacy-related information (e.g., privacy properties) of IoT services of interest through web browsers installed on their smartphones. The web server (IoT Service Store; ISS) then retrieves requested information from the database (DB) and returns the results back to the users. ISS is regarded as a trustworthy entity. Users can also send requests for other types of operations (e.g., giving a rating or setting up privacy preferences for the target service) to the server. With the intent of making the system as easily accessible as possible, we followed standard Web protocols in the implementation of functionalities for communication and interactions between the entities under the system. In addition, all information sent between the user and ISS is secured through use of the HTTPS protocol.

In order to let IoT service providers uniformly inform users about service descriptions along with privacy implications, we adopted the Physical Web as an underlying communication mechanism. Physical Web is an open source project that aims to transform all physical objects (e.g., parking meters) into smart objects by allowing the objects to interact with the Web via Bluetooth Low Energy beacon profile called Eddystone [10]. Eddystone beacons are capable of broadcasting object-specific identifiers like URLs, which are automatically searchable by nearby users’ Bluetooth-enabled Android or iOS smartphones. Using these URLs, users are then able to browse web pages containing relevant information about physical objects (e.g., parking rates), and also to perform additional actions through their smartphones (e.g., payment of parking fees). We chose Physical Web since it enables users to easily find and interact with resources in their physical environments, without first downloading an additional app. In the current implementation, ISS assigns a unique URL to each of the registered IoT services. Service providers need to deploy Eddystone beacon(s) broadcasting the assigned URLs to advertise their services.

ISS is a standalone web server running on a virtual private cloud. We installed a standard web service stack composed of Linux, Apache, MySQL, and PHP (LAMP) on an Amazon Elastic Compute Cloud (EC2) instance and deployed a server program on this virtual machine. We assume that each IoT service provider registers its service(s) to ISS, while providing the detailed data collection practices in an honest manner. Based on this information (stored in a MySQL database), ISS systematically composes a specific web page with user interfaces displaying privacy-related information of each IoT service. After that, ISS assigns a unique URL for the service. We designed and implemented all user interfaces for ISS using HTML5 and CSS. We also added some JavaScript functions (e.g., jQuery for changing the content of a web page without reloading) to make the user interfaces more interactive and responsive for the users. We will further elaborate on ways to formulate the web page in the following sections. As discussed above, ISS is also designed to communicate with the TIPPERS IoT framework so as to ask IoT services to follow user-defined privacy settings (see the lower left of Fig. 1).

B. Workflow

We now explain the functional workflow of the proposed system (see Fig. 2).

(1) As discussed, IoT service providers broadcast their service URLs through Physical Web-compatible beacons. These URLs, mapped with web pages summarizing privacy-related information of the IoT services, are automatically detectable by
nearby users’ Bluetooth-enabled smartphones. Android users might become aware of available services (URLs) via OS-level notifications (i.e., Google Nearby); iOS users, in contrast, will get notifications from Chrome browsers installed on their devices.

(2) Users click on URLs of the IoT services of interest and then check various privacy-related information on these web pages being hosted by ISS. Each web page first visualizes all possible combinations of sensor data as well as personal information which can be inferred from each combination of data (i.e., privacy properties). Next, it provides user interfaces for both receiving an individual user’s evaluation of utility-privacy tradeoffs and showing all other users’ evaluations with statistics. The last part of the web page is a list of all available actions for the user (e.g., disallowing some data collection).

(3) Users are also able to browse other IoT services running at their current location by clicking a “Browse other services” button. When this happens, users will be redirected to a different page enumerating nearby IoT services with short descriptions. In this page, IoT services are sorted in descending order by the users’ ratings about the utility-privacy tradeoffs.

(4) If users find any other interesting services, they can check the details by clicking a “Learn more” button.

C. Privacy Properties of IoT Services

To make users more aware of privacy risks related to IoT services they are using, we defined an information architecture capable of expressing privacy properties of the services. As discussed earlier, users have a limited understanding of the implications about the collection of various kinds of sensor data. The primary goal of the proposed system is reinforcing users’ privacy awareness by letting them understand each IoT service’s privacy properties, namely relationships between raw sensor data (e.g., Wi-Fi MAC address) and higher-level descriptions of personal information (e.g., user identity) which can be inferred from the sensor data.

1) Sensor Data and Personal Information

To begin with, we need to build a taxonomy of sensor data collectible in IoT environments. Since we plan to integrate the proposed system with the TIPPERS IoT framework, currently running on a six-story building (DBH) at UCI, we first defined the types of sensor data according to available sensor devices installed in DBH (see Table I). As explained, IoT service providers need to register their services with detailed data collection policies. In order to systematically express the policies for collecting sensor data, we also defined the following attributes: “service_id” (identifier of service collecting and processing data), “mandatory” (indication that data is mandatory or optional for using the service), “source” (sensor device collecting data), “storage” (location where data is stored), “retention” (time duration for which data is stored), “protection” (security mechanism for protecting data), and “sharing” (the existence of third-parties allowed to access data). As all this information is stored in a MySQL database, service providers (or system administrators) can update their data collection practices via simple Web user interfaces.

Next, we developed a taxonomy of (inferable) personal information in the context of IoT. Since we were unable to find a classification scheme for personal information in IoT, we created a broad-brush classification scheme using the P3P specification V1.1 [11] as a baseline. To do this, we considered the CATEGORIES element that describes 16 different types of personal information available in Web environments. This element was originally designed to help Internet users define generalized preferences and rules (i.e., P3P privacy policy) for the exchange of their personal data through the Web. We augmented this element to consider personal data available in physical environments, as opposed to Web environments. For instance, IoT devices are argued to accurately recognize users’ emotional states by analyzing video footage captured by a security camera [12]. We added the following types of personal information: “Physical Activity” (e.g., biking or cooking), “Physical State” (e.g., sitting position), “Emotion” (e.g., Ekman’s six basic emotions), “Personality” (e.g., The Big Five personality traits), “Cognitive Activity” (e.g., intention), and “Social Relationship” (e.g., workplace dynamics). We also excluded some personal information which is not directly related to IoT (e.g., user-generated Web content). As a result, we wound up with 16 types of personal information as described in Table I. Note that each type of sensor data or inferable personal information may have subcategories (e.g., “Image>Face Photo”) to better express its meaning.

<table>
<thead>
<tr>
<th>Sensor Data</th>
<th>(Inferable) Personal Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Image/Video</td>
<td>1. Identity</td>
</tr>
<tr>
<td>3. Wi-Fi</td>
<td>3. Financial</td>
</tr>
</tbody>
</table>

\(a\) Attributes: source, storage, retention, protection, sharing

\(b\) Newly defined for IoT

2) Inference of Personal Information

Providing users with knowledge of the possible inferences of sensitive personal information is important for increasing their privacy awareness in IoT environments. The most straightforward way would be for IoT service providers to generally describe the inferred personal information from the collected sensor data. This declaration may, however, reveal confidential business strategies [9]. Also, each service provider may not know other possible inferences that provide no utility to the service it offers. For instance, a service provider who is utilizing facial recognition software for the purpose of user authentication (“Identity”) might be unaware of the fact that a similar technique can be used to infer users’ sexual orientation (“Preference>Sexuality”) [13] from previously collected image data. For these reasons, we believe that the construction of a knowledge base about the inference of personal information is necessary for reinforcing awareness of the privacy risks in IoT. As a starting point, we therefore defined if-then rules for specifying the privacy properties of the IoT services, composed of the combination of available sensor data (antecedent) and possible inferences of personal information (consequent). In doing this, we referred to literature on mobile sensing and data mining related to personal information of the user [14-16]. We will also...
We generated 36 different rules for five hypothetical IoT services registered to the proposed system. As an example, AutoMusicOn is a service that aims to automatically play music upon the user’s entrance into a specific room. To check whether a registered user enters the room, AutoMusicOn is collecting Wi-Fi MAC addresses of mobile devices and/or face photos of people inside the room. In this case, the Wi-Fi MAC address will need to be collected because it can imply both the current location of nearby devices (see “Device>Location” in Fig. 3) and the identity of their owners (“Identity”), thereby possibly inferring user location. Optionally, the service may also collect face photos in order to verify a user’s identity via facial recognition and finally confirm the presence of this user in the room (“Presence”). With the collected facial images, however, the service provider (or third-party) might infer the sexual orientation of the user (“Preference>Sexuality”) in the future. We manage these rules in a MySQL database under the system for commonly applying them to all the registered services. Therefore, each IoT service’s web page summarizes its privacy properties as depicted in Fig. 3. As can be seen, (mandatory/optional) sensor data and inferable personal information are distinguished through color coding. In addition, all items are visualized as clickable buttons; users are then able to click the button to view the popup window giving additional explanations about the item.

![Privacy Properties Popup Messages](image)

**Fig. 3. User interfaces showing privacy properties**

### D. User-driven Assessment and Control of Privacy Risks

Aside from providing the privacy properties, ISS lets users evaluate and share their opinions about nearby IoT services. We adopted a conventional five-star rating system to allow users to collaboratively assess the subjective balance between the benefits and privacy risks of using a specific IoT service. After checking privacy properties of the service, users can leave their ratings (ranging from 1 to 5 stars), just as they do in general e-commerce systems like Amazon (see “Your Rating” in Fig. 2).

Since the system distinguishes users with server-generated identifiers (i.e., random number assigned to $\_SESSION variable in PHP), users are able to change previously submitted ratings whenever they want. For instance, a user who gave 1 star to AutoMusicOn solely based on its stated privacy properties might update his/her rating after actually using the service. Except for the session identifier, the system does not collect or use any personal data (e.g., phone number) to recognize the user at this moment. To present multiple users’ evaluations about the service at a glance, ISS also calculates its average rating and displays it with the distributions of star ratings (see “Crowd Rating” in Fig. 2).

Lastly, each service web page provides users with the following control options: (1) subscribe service, allow all data collection, (2) subscribe service, disallow some data collection, (3) unsubscribe service, and (4) browse other services. Users can make choices by clicking buttons in the web page (see “Your Action” in Fig. 2). Regarding option (2), ISS is designed to send a message requesting the “opt-out” of specific data collection (i.e., user’s privacy preferences) in using the service. It can be simply done through JSON-based REST APIs and a policy enforcement engine, both of which are being provided by the TIPPERS framework [5]. Using the abovementioned session identifiers, the system keeps track of all actions performed by the users, thereby allowing them to change their preferences if necessary.

### IV. DISCUSSION AND FUTURE WORK

In this section, we explore future opportunities for research that will potentially help further improve the effectiveness of the proposed system, in making people aware of the privacy implications regarding the IoT service and available options for them to avoid potential privacy breaches.

#### A. Scalable Knowledge Base for Privacy Properties

We consulted existing literature to define a set of rules specifying possible inferences that can be drawn from sensor data available in IoT, and then used these rules to present the privacy properties of the IoT services. However, we recognize that in practice, claimed inferences are complex, highly context-dependent, and open to misinterpretation. This approach will necessarily have scalability issues as long as humans are a component of the system. One possibility, albeit not ideal, is to build a probabilistic information retrieval system that automatically extracts relationships between entities (e.g., sensor data and personal information) in text data (e.g., machine learning literature). To build such a system, however, we still need labeled training data (i.e., known relationships between sensor data and personal information), which does not exist to our knowledge. One approach is to utilize systems such as Snorkel, designed for programatically generating labeled training datasets from raw data without much human intervention [17].

#### B. Understanding Users’ Perception of Privacy

Through our previous location-based survey study, we showed that people’s privacy preferences (or decisions) are significantly affected by the awareness of inferable personal information in a simulated IoT environment [18]. Regarding the collection of image data, for instance, users are more worried when they realize the implications of image-based human age estimation. In contrast, they are very open to providing information about their devices (e.g., phone identifier) if they perceive that this information is not related to their sensitive personal information. However, these findings are based on people’s stated privacy preferences towards hypothetical service scenarios, not actual behavior of using working IoT systems. It is therefore necessary to analyze users’ privacy behavior captured in real world situations, both for validating our previous findings and for extracting additional insights about the privacy awareness in IoT. In this vein, we plan a collaboration...
with the TIPPERS research team in order to incorporate their internally (or externally) developed service apps into the proposed system. Thereafter, we will deploy the integrated system to DBH and conduct field experiments with real users (i.e., building inhabitants) to collect their privacy-related behavioral data generated while using the system.

C. Privacy Decision Support

Even though the proposed system presents users with information that will help them understand some privacy implications of using various IoT services, users still need to configure their privacy settings by themselves. However, some users may have difficulties in doing so due to limits in their available time, motivation, and cognitive decision-making abilities [19, 20]. Therefore, the system may need to assist users with making better privacy choices, perhaps by predicting future decisions based on their historical decision-making behavior and recommending privacy settings accordingly (i.e., privacy decision support). We are currently considering using machine learning (ML) methodologies to realize this functionality. By using the abovementioned privacy behavioral data (i.e., users’ interaction logs of using the system and their submitted privacy settings) as training data, we can train ML model(s) predicting privacy settings of the users. We then embed the trained ML model(s) into the proposed system (see “ML Engine” in Fig. 1) for providing users with machine-generated privacy recommendations.

V. Conclusion

In this paper, we designed and implemented a novel web-based system called IoT Service Store (ISS). Our goal is to allow users to comprehend the privacy implications of nearby IoT services through gaining a better understanding of the data collected and possible inferences that may be drawn from this data. ISS also allows users the ability to control the collection of their data. For concreteness, we adopted the Physical Web as an underlying communication channel between the users and IoT services, thereby realizing an easy discovery of IoT services operating near the user. In order to efficiently express and convey the detailed privacy properties of each available IoT service to the user, we developed an information architecture for describing the relationships between collected sensor data and inferable personal information. We designed and implemented user interfaces, as a presentation layer of ISS, not only visualizing the privacy properties of an IoT service, but also enabling users to collaboratively assess its potential privacy risks and configure privacy settings according to their privacy expectations. Future work will mainly focus on the following topics: the automated extension of a knowledge base used for presenting IoT services’ privacy properties, collection and analysis of people’s perceptions of privacy in operational IoT environments, and a ML-based privacy decision support system that alleviates users’ cognitive burden of configuring privacy preferences for diverse IoT services.

ACKNOWLEDGMENT

This work was done while Hosub Lee was a summer intern at Intel Labs, Intel Corporation, Santa Clara, CA. The authors would like to thank the reviewers for their valuable comments on earlier versions of this paper.