Protecting User Privacy in the Connected World

PI, Khalid Elgazzar
Ontario Tech University, Oshawa, Ontario L1G 0C5
Final Report, July 30, 2020

The project is supported by Office of the Privacy Commissioner of Canada, Grant #:

1. Abstract

This project develops technologies that place users at the central control of the sharing and dissemination of their private data to make the Internet of things an unobtrusive technology through the design and implementation of a generic privacy framework. The framework enables users to flexibly define abstract and high-level privacy goals that can be used to generate runtime policies that capture the user's constraints on private data sharing according to the operational context. The proposed framework includes a global resource registry for devices and services, personal privacy broker, and policy enforcement mechanisms. The personal privacy broker implements a privacy-aware discovery algorithm that uses both query and on-demand requests which can use high-level settings to only respond to authorized discovery requests using self-contained authorization tokens. Therefore, devices belonging to the user are only discoverable by authorized users/services. The broker also implements a novel policy negotiation protocol to enforce user-defined policies on IoT devices collecting data from the user domain and streaming it to the outside world. An Android application has been developed to provide a user-friendly multimodal interface where the user can define privacy policies and tie them to certain locations for access restrictions or approvals. The application can then connect to devices belonging to the user and monitor data streaming to identify any privacy violating data collection practices. The developed techniques could be used in both online interactions and local IoT environments. This project allows users to manage their changing privacy preferences and constraints according to who is receiving the data, for what purpose and in what context.

2. Introduction

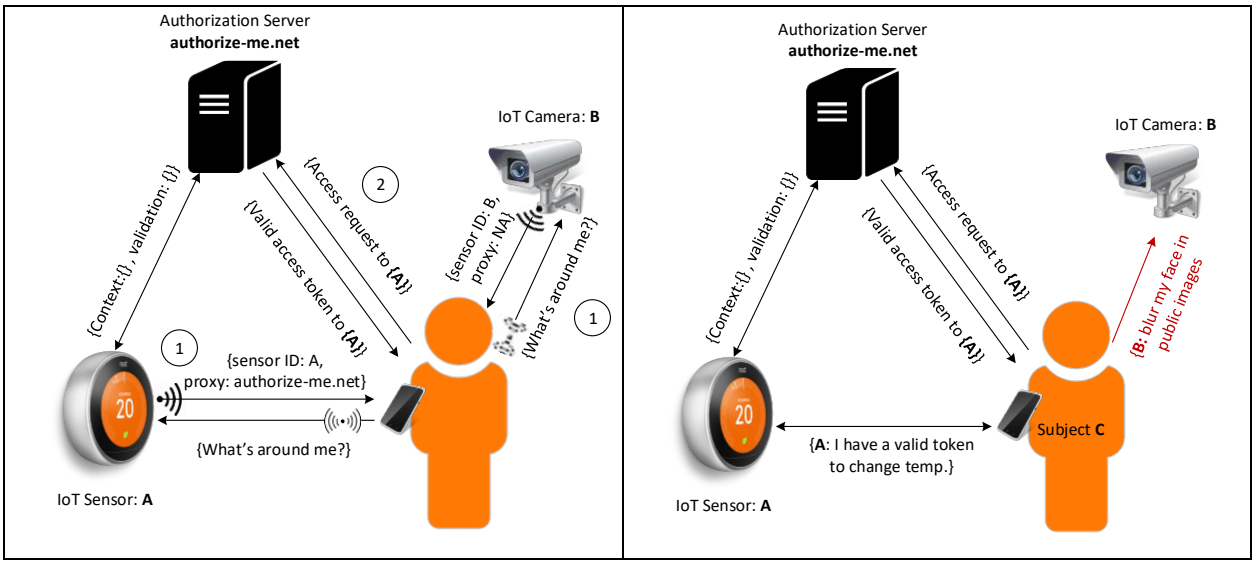
The Office of the Privacy Commissioner of Canada articulated in 2016 that IoT poses significant risks to our privacy, with significant reluctance from consumers to adopt its technologies. Similarly, in its 2015 privacy report, the Federal Trade Commission (FTC) identified that privacy is a core requirement in the Internet of Things (IoT) and is the major concern that makes users reluctant to embrace it. They also pointed out that notice and consent is one of the vital challenges associated with the widespread deployment of IoT. As of May 2018, the European General Data Protection Regulation (GDPR) mandates that data collection practices must be transparent and require a confirmed user consent for personal data collection. However, in reality people are not used to reading privacy notices when they browse the web from desktops, laptops, or portable devices, let alone understanding what these privacy notices imply and what it really mean from privacy perspectives to allow for a certain data collection.

The ultimate source of user discomfort is the lack of control over personal raw data that is directly streamed from the user domain to the outside world. In most cases users are unaware of ongoing

data collection whether it be from their physical devices (sensors) or as they interact with online services. Even if they do, users are often lacking the correct under- standing of what it means to share/release their personal data at a certain granularity. For example, a utility company may collect consumption measurements (e.g., water usage) with fine granularity to study consumption patterns and possibly improve customer experience. However, data analytics techniques could provide significant insights on the behavior of the occupants of the house including daily schedules, usage patterns, anomalies, and health status. The most alarming issue is that most things are quarriable and can be accessed by remote entities. The rapid evolution of the Internet of Things will continue to aggravate privacy concerns, and penetration to people's life will grow deep through identification, data linkage, and profiling, along with new challenges that will continue to unfold. To protect user privacy in open Internet and pervasive environments, systems must instill mechanisms to respect people's privacy by design from the ground up. In this project, we developed a technology that flexibly enables users to control the flow of their private data to the outside world.

3. System Architecture Overview

In dynamic IoT environments, users encounter IoT things that collect personal data in everyday life scenarios. These devices collect a wide range of data that might be classified as sensitive data by itself or lead to the revelation of personal practices related to personal interest, religious beliefs, sexual orientation, etc. Figure 1 illustrates a high-level view of the proposed architecture. Users discover resources collecting data in their surroundings, query data collection procedures, negotiate privacy policies and configure their own privacy preferences. For the sake of simplicity, we only provide visionary interactions, not a complete system architecture in its final design blueprint. In scenarios in which users may have insufficient privileges to configure the way IoT things collect their personal data, users must be warned of possible privacy threats that may violate their current preferences by surrounding things (whether directly or indirectly). Therefore, users can make informed decisions when applicable (e.g., choose to leave a domain when enforcing privacy policies is infeasible). The system architecture primarily encompasses the following major components, (1) resource registry, resource discovery, and personal privacy broker.



(a)users discover/request access to nearby things (b)users set/change privacy configurable parameters

3.1 Resource Registry

IoT embeds heterogeneous sensors (aka. resources) in our surroundings that continuously observe and collect data of different types, analyze it and make decisions. Examples of such resources and services include, companies deploying HVAC systems, presence detectors, audio/video equipment, and location tracking in office buildings. Public services include, but not limited to, airborne or fixed traffic monitoring services, computer vision-based event detection and classification systems, public safety and remote health monitoring systems.

Nowadays, we also deploy private systems such as smart home services, door locks, surveillance cameras, thermostats, and voice-enabled home assistants. All these services deploy connected devices to capture public and personal data that could be sensitive by nature or revealing by data analytics. In all these services, resource owners must register their services and declare any data collection practices associated with these services. We anticipate that at some point of time, either regulation will have to enforce service providers to register their resources and follow certain defined frameworks, or providers will be incentivized by the overall ecosystem to integrate with other services, otherwise will be left alone. The resource registry facilitates resource registration and is based on an open and distributed architecture in which any number of actors may collaborate in the deployment and management of these resources. We refer to resources as any service, whether it is going to be an online ser- vice or a sensor that is deployed (or can be deployed) and observe a property or a phenomenon and collect user data. The resources registry maintains a comprehensive list of available resources whether permanently deployed or can by deployed on-demand according to needs. In large scale deployments (e.g., smart city systems), the large number of heterogeneous resources and diversity of required functionality pose major scalability issues on the under- lying networking infrastructure. Thus, the management of such resources must be carefully designed to facilitate efficient resource management and monitoring, yet ensure high performance and quality of service delivery. The registry offers a registration service to register resources using a unified resource profile template in JSON encoded format for interoperability and portability purposes. This allows for dynamic resource maintainability, where resources can be easily added or removed. The resource profile registration template includes the capabilities, properties of a registered resource, data collection practices, and access per- missions. Properties describe the various states (e.g., active, suspended, ready, disabled, etc.) and metadata (e.g., UUID, name, location, spatiotemporal attributes) of a resource. In general terms, capabilities describe actions/functional operations that can be performed

3.2 Resource & Service Discovery

An important aspect of our design is how users can discover active resources and data collection practices in their proximity that may pose risks on their privacy. Our design objective from this perspective is two-fold: (1) allow users to discover resources already registered in the global/geographically-bound resource registry. Resource registries are discovered through centralized directories of registries covering different geographic areas (pretty much similar to domain name resolution); (2) enable users to discover resources in their proximity range based on a P2P discovery procedure. The first objective is to facilitate the establishment of a well-defined data collection notice and consent framework based on predefined and declared data collection procedures. This empowers service provider-user negotiations and settles on a common and well-

defined agreement between users and collecting entities. The second objective is primarily a best-effort approach to protect user privacy based on ad-hoc resource discovery. This enables users to discover devices collecting data in their domain, or domains they just walk in or pass through, query or challenge data collection procedures, and advertise privacy preferences to these collecting devices. We assume that this approach will essentially serve as a warning service to users rather than an enforcing mechanism until the practice is well-established and adopted in the entire echosystem by both vendors, system developers and users.

3.3 Personal Privacy Broker

In open IoT environments, it's often difficult for users to realize that data collection is underway (e.g., camera or audio recording). Privacy preserving brokers provide users with the necessary tools to enforce their own privacy preferences using local trusted computing infrastructure. Brokers are deployed on the user's smartphone and monitor all data streaming out from the user's domain and ensure user privacy is respected. Our design strategy is that users define access policies governing access to their private data with a simple and user-friendly mobile interface. The broker discovers IoT devices in close proximity, monitors data collection practices, and authorizes access requests according to these policies. It also provides necessary data manipulation (e.g., creating multi-tier data fidelity from raw sensor data, data anonymization, and denaturing) to conform with privacy policies when responding to access requests. For example, we can generate a varying level of abstractions from location raw data, including country, city, neighborhood, postal code/zip code, and exact longitude and latitude. Thus, the location can be sharded at different levels of granularity according to who is accessing the data and for what purpose.

Brokers are continuously running services that can be queried by requesting entities. They can also announce privacy policies to surrounding environments, negotiate access terms and conditions, and reconfigure privacy parameters. Additionally, since users may typically have flexible privacy constraints in different scenarios according to the situation, privacy brokers allow users to maintain multiple configurable profiles to choose/tune privacy policies according to situations (think of this feature as multiple privacy avatars). The privacy broker leverages machine learning techniques (e.g., reinforcement learning) to evolve user policies and constraints according to user behavior and decisions made over time. These inferred policies are presented to users for approval and feedback.

Our design supports two types of scenarios: (1) data collection services are registered and declared by their providers, hence, data collection practices are well framed and provide user configurable parameters; (2) users just discovered devices that are collecting personal data in their proximity (e.g., cameras in conference rooms, and microphones in meeting rooms), hence, begin to negotiate privacy preferences or make proper decisions. In both scenarios, users can request access to collecting devices and reconfigure privacy settings based on predefined rules/options or via a negotiation procedure. Access will be granted through a valid access token from the resource-designated authorization server associated with the data collecting entity. Access tokens are self-contained JSON web tokens RFC [7519] that deter- mine the scope of access and expiration time based on the user's submitted credentials. The token is signed using standard RSA signing algorithms and can be validated using the authorization server's public RSA key.

4. Milestones and Deliverables

This research has developed technologies that place users at the central control of sharing and dissemination of their private data and make the Internet of things an unobtrusive technology through the design and implementation of a generic privacy framework. The user is able to flexibly define abstract and high-level privacy goals that can be used to generate runtime policies that capture the user's constraints on private data sharing according to the operational context. The framework includes a global resource registry, personal privacy broker, and enforcement mechanisms. These components allow users to manage their changing privacy preferences and constraints according to who is receiving the data, for what purpose and in what context.

The project has developed the following Key capabilities:

1. **Resource Registry:** A distributed resource registry to provide necessary data storage and handling for resource registration (resources include devices and services).
2. **Privacy-aware Discovery Protocol:** This protocol supports ubiquitous data access and enables users to discover nearby resources on-demand in a P2P fashion, even if resources are not registered using direct signal scanning.
3. **Policy Announcement and Negotiation Protocol:** This protocol facilitates privacy negotiation and enforce user-defined policies on surrounding data collection practices.
4. **Privacy Broker:** This acts as a personal privacy guard by watching out and analyzing all data streaming out of the user's domain, whether from their own connected devices or devices surrounding them.
5. **Proof-of-concept Prototype**: An Android application puts these concepts altogether in action as a proof-of-concept prototype to demonstrate the utility and usability of the proposed framework.

5. Conclusion

This project developed novel technologies that enable users to intelligently define, manage and enforce their privacy policies over online interactions as well as pervasive/IoT environments. For online interactions, a mobile application analyzes data streaming from user devices and informs users about ongoing data collections. In smart and pervasive environments, the developed framework enables users to: (1) declare the presence of their resources and their procedure of data collection and for what purpose; (2) discover resources (devices and applications) around them that may collect data. The developed technology is able to alert users on data collections practices that violate user-defined privacy policies. Machine learning techniques are utilized to learn user privacy preferences from interactions and proactively apply these preferences without distracting the user attention.