Contextual Privacy Policy Modeling in IoT

Emmanuel Onu
Department of Computer Science
University of Calgary
Calgary, Canada
emmanuel.onu1@ucalgary.ca

Michael Mireku Kwakye
Department of Computer Science
University of Calgary
Calgary, Canada
michael.mirekukwakye@ucalgary.ca

Ken Barker
Department of Computer Science
University of Calgary
Calgary, Canada
kbarker@ucalgary.ca

***Abstract*—The Internet of Things (IoT) has been one of the biggest revelations of the last decade. These cyber-physical systems seamlessly integrate and improve the activities in our daily lives. Hence, creating a wide application for it in several domains, such as smart buildings and cities. However, the integration of IoT also comes with privacy challenges. The privacy challenges result from the ability of these devices to pervasively collect personal data about individuals through sensors in ways that could be unknown to them. A number of research efforts have evaluated privacy policy awareness and enforcement as key components for addressing these privacy challenges. This paper provides a framework for understanding contextualized privacy policy within the IoT domain. This will enable IoT privacy researchers to better understand IoT privacy policies and their modeling.**

***Index Terms*—Privacy Policy, Cyber-physical systems, IoT Privacy Taxonomy, Privacy Formalization, Cyberspace, Context, IoT**

I. Introduction

The Internet of Things (IoT) is a paradigm composed of a networked computing environment of uniquely identifiable physical objects which are referred to as "things" on the network [5], [27], [32]. These cyber-physical objects are equipped with several enabling technologies such as sensors, actuators, and communication infrastructures - which allow the objects to collect and exchange data over a network with little or no human intervention. Thus, the IoT exhibits typical characteristics of a ubiquitous computing technology in its ability to unobstructively work in the background [17].

An intrinsic part of many IoT devices is data collection. These IoT devices are usually embedded with sensors which allow the device to collect information about its surroundings - including information about people and environmental elements. Some of the typical sensors embedded in IoT devices include: temperature, proximity, pressure, water quality, chemical, gas, smoke, infrared, motion detection, accelerometer, gyroscope, and so on.

There are several benefits from the integration of the Internet of Things into our everyday lives - from increased convenience to improved health and safety efficiency [33]. These benefits are made possible through the deployment of several new services and applications that have been able to leverage on the interconnection of the physical and virtual realms [32]. For instance, an energy management system could adjust a room temperature based on the physiological status of an individual attained through a communicated measurement of the heart rate and body temperature of the individual [42].

A. Motivation

Despite the benefits realizable from the IoT, it also brings privacy and security challenges [33]. Ziegeldorf *et al.* [48] enumerates many of the privacy threats in the IoT. One of the most prevalent privacy issue with the IoT is unsolicited data collection and the lack of privacy choices [31], [33], [48]. These privacy issues are largely attributed to the lack of awareness of their existence and the data collection practices of the IoT devices. Hence, to address these privacy challenges, there is a need to create privacy awareness and ensure data use is in accordance with people's privacy preferences. Privacy awareness is created by making individuals in an IoT environment aware of the privacy policies of data collecting IoT devices in the environment. A privacy policy states information about the data collector, the purpose for the data collection, the type of data collected, and how the collected data will be used and shared.

Current research focuses on creating privacy awareness, managing users privacy expectations, and alleviating cognitive burden from users in an IoT environment [10], [25], [27], [31], [33]. Most of this work identifies key elements in an IoT privacy policy and investigates how to capture users' privacy preferences. This work leverages on either the IoT privacy policy or users' privacy preferences or both.

B. Assumptions

This paper proposes a privacy model as a framework for understanding IoT privacy policies. The assumptions made to frame this research are as follows:

1) The scope of the research study is based on the typical configuration of IoT devices in a smart building environment. This smart building may be composed of smart IoT devices, such as, smart thermostat, smart lights, smart security locks, smart TVs, smart CCTV camera, *etc.*, or any other IoT device that may be configured in a smart building. Additionally, the smart building may be instantiated as a smart home, smart office, or smart hospital.
2) With regards to the context privacy element, we consider *location* and *time* as the key factors in determining a privacy context for IoT devices.

978-1-7281-6609-4/20/$31.00 ©2020 IEEE
DOI 10.1109/DASC-PICom-CBDCom-CyberSciTech49142.2020.00030

3) We consider categories of IoT devices that primarily collect and process data from the users.

C. Contributions

The key contribution of this paper is to develop a framework for understanding and modeling privacy policy within the IoT environment. This framework could be used to classify the contributions of existing research work - which will show what aspect of IoT privacy policy are considered. Our technical contributions are summarized, as follows:

1) To develop a framework for modeling IoT privacy policy within the context of smart environments.
2) To define and classify a taxonomy representation of the key aspects of IoT privacy policy description and their relationships.
3) To establish a formal definition and expression of the semantics of privacy notations used in IoT privacy policies.
4) To adopt and use the formal privacy model and taxonomy to express or explain privacy policy description within the context of a smart building environment. Moreover, we use the formal privacy model to help develop IoT privacy policies from the viewpoint of IoT developers, and classification of existing privacy models.

The balance of this paper is organized as follows: Section 2 discusses related work on privacy policy and preference modeling. Section 3 discusses the different elements in an IoT privacy policy description and their related taxonomy specifications. In Section 4, we present a formal description of privacy policy. Section 5 shows the application of our privacy model and taxonomy. Finally, Section 6 concludes and discusses areas of future work.

II. Background and Related Work

Research studies conducted over the years have shown increasing privacy concern in people with regards to how their data is collected and used [30], [45]. This has led to further studies in several domains on how to address privacy concerns. This work, depending on the privacy problem, addresses privacy violations through various approaches, such as, avoiding identification of data subjects, reducing information leakage, and controlling data access. These are achieved using techniques, such as, anonymization [43], cryptography methods [41], and access control models [19].

Privacy is ideally personalized and contextualized [34], [35], [44] - hence, privacy protection should be designed to meet the expectations of users by capturing their privacy preferences and their situational context. However, challenges exist when we attempt to capture and model context, privacy policies, and user privacy preferences. Several approaches have been proposed over the years and some notable ones are discussed in the following.

A. Privacy Policy and Preference Languages

Privacy policy and preference languages allow for human-readable privacy policies to be expressed in precise and computer compatible formats [24]. There are several privacy policy languages, each of which have different syntax and were designed for specific purposes. For example, the Platform for Privacy Preferences (P3P) [9] is designed for expressing website privacy policies in machine-readable format by the World Wide Web Consortium (W3C). The W3C designed a language for expressing users' privacy preferences, called the P3P Preference Exchange Language (APPEL) [8]. IBM subsequently developed the Enterprise Privacy Authorization Language (EPAL) for the purpose of expressing the internal privacy policies of organizations [4]. A consortium of organizations developed the eXtensible Access Control Markup Language (XACML), which is a language for expressing privacy and security policies [3].

In general, many of these languages are designed to be lightweight and use the XML markup languages. Apart from XML, other approaches used for expressing policy languages include: JSON [36] and temporal logic [7].

B. Privacy Policy Modeling using Taxonomy, Visualizations, and Ontology

The study of better privacy-preserving polices and privacy-aware database models have garnered interest in the research community. Barker *et al.* [6] outline a privacy taxonomy for the collection, sharing, and disclosure of private data stored in repositories. The authors use their privacy taxonomy to define parameters for a privacy policy while protecting the data provider's personalized information. Ghazinour *et al.* [16] outline a model that focuses on policy visualizations to facilitate policy understanding for data owners, and data accessors, amongst others. Their study describes how the model is applied in various scenarios to demonstrate its utility.

Privacy ontologies form the structure and model for most context-based privacy policy formulations. Ontologies formalize diverse preferences, unique attributes, and distinct rules for managing, processing, and modelling privacy policies. Additionally, ontologies help in instantiating contextualized privacy policies for different individuals or scenarios (circumstances). Several research efforts have investigated ontologies, and these studies define and describe how ontologies are used in modelling real-world scenarios and application domains. The results from these studies allow insights from the ontologies and their predicates, especially in the framework of controlling privacy in context-aware systems [46]. Accordingly, the study of context-aware and rule-based ontologies suggest how to apply them to unique scenarios. Moreover, the outputs of context-aware and rule-based ontologies are relevant in defining the semantics and predicates of privacy ontology modelling [47].

C. Context

Dey and Abowd [1] define context as *"any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves."* They identified four key context types for characterizing the situation of a particular entity, which are: *location*, *identity*, *activity* and *time*. They do not address the problem of how to model or represent contextual information.

Dourish [13] presents two views of context which have been highly instrumental to context modeling in context-aware ubiquitous computing and recommender system. The first is the representational view, which makes four key assumptions of context:

1) context is a form of information,
2) context is delineable,
3) context is stable, and
4) context and activity are separable.

The second is the interactional view of context, which contradicts the key assumptions made with the representational view by making the following countering assumptions:

1) Contextuality is a relational property, which means that certain information may or may not be relevant to some activities.
2) The score of contextual features is defined dynamically.
3) Context is an occasioned property, which means that context is particular to each occasion of an activity or action.
4) There is a cyclic relationship between context and activity.

However, the representational view of context has been widely adopted because it makes the concept of context conceivable and applicable. Two major implications of this view of context are: 1) contextual information has to be collected along with any form of data collection, and 2) relevant contextual variables and their structure must be predefined.

III. IoT Privacy Policy Description and Taxonomy

This section discusses key elements of an IoT privacy policy as defined by IoT manufacturers [21], [39] or as used in Building Management System (BMS) [28], [29]. Moreover, in describing an IoT privacy policy, we define a taxonomy framework that classifies and defines the various aspects in the formulation of an IoT privacy policy, using a tree-like hierarchical form.

We consider smart facilities which involve, but are not limited to, smart building, smart home, and smart office. These smart facilities contains IoT devices, such as, smart thermostats, smart lights, smart TV, smart locks, and smart CCTV camera, amongst others. We consider these devices as the likely devices found in a smart facility, but our consideration is not limited to them [37].

A typical IoT device under consideration in this paper is a smart thermostat. A smart thermostat is usually part of an HVAC system which could considerably reduce heating and cooling cost by intelligently managing energy consumption. Essentially, a smart thermostat automatically adjust room temperature based on its occupancy through the help of several sensors and by learning over time the preferred room temperature of a user. The sensors are used to detect the presence, number, and location of people within a space. An example of a smart thermostat sensor is a passive infrared (PIR) motion sensor.

Based on previous research, IoT device manufacturers privacy policies, and typical setup configuration of a smart building, we propose a privacy taxonomy for a smart environment. The utility of this taxonomy is in modeling of privacy policies for IoT devices and users' preferences. The key privacy elements in our taxonomy include: *data collection frequency*, *purpose*, *data retention*, *data type collected*, *data granularity and inference*, *sensor activation*, *data recipients*, and *context*. These privacy taxonomy elements are sufficient and necessary for a privacy policy taxonomy description of an IoT setup for smart buildings. Fundamentally, privacy elements such as *purpose*, *type of data collected*, *data retention*, *data granularity*, and *data recipients* are necessary for modeling privacy policies. However, there are some privacy elements which are unique to IoT devices. These elements include: *data collection frequency*, *sensor activation*, and *context*. Data collection frequency refers to the rate of data collection by an IoT device. The type of data collected refers to the data collected by IoT sensor devices. Sensor activation refers to activities or elements which could trigger data collection by an IoT device. Context refers to situational factors during data collection by an IoT device such as location and time. Our privacy taxonomy elements are discussed in more details in the following sections.

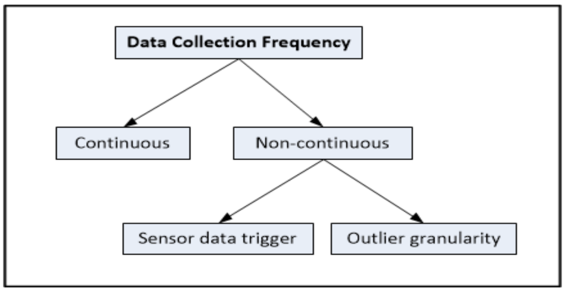

Fig. 1: Data Collection Frequency

A. Data Collection Frequency

The frequency of data collected refers to the rate or the amount of data collected within a specified period of time. IoT data is generally intermittent and highly unstructured. The frequency could vary depending on the prevailing conditions of location (for example, being at home or at a public place, such as, a library) for data collection. Generally, data collection frequency is categorized as either *continuous* or *non-continuous* [38], [40]. In terms of a non-continuous data collection frequency, suppose we have a smart home consisting of various IoT devices, such as, smart thermostat, smart light, or smart door. These devices will start collecting data from the data providers (humans subjects that enters the building) based on sensing human movement. Secondly, in the absence of human movement in the building no data is collected [40].

On the other hand, data collection frequency will be categorized as continuous based on the nature and type of data collected. This is attributed to the location where data is collected [38]. Suppose we have a smart CCTV camera in a public place, such as, a school campus, or a court house, where data is continuously collected. It is important to note that even within the same public place, the frequency of data collected within certain perimeters or rooms could vary from other rooms.

Figure 1 displays a hierarchy of the *data collection frequency* taxonomy for an IoT privacy policy description. In analyzing the frequency of data collected by IoT devices, a number of factors influence these data collection rates. Here, data collection frequency is categorized as, *time*, *location*, *sensor responses or triggers*, *etc.* It is assumed that these data collection tendencies are generally influenced by the presence and activities of the data providers at a particular location and at a particular time.

B. Purpose

Purpose is a fundamental component of any privacy policy because it captures the reason for the data collection [6], [20]. For example, the reason for data collection by a smart thermostat is to efficiently manage energy consumption while for a smart camera it could be for attendance management or security. A clearly defined purpose allows the data collector to identify and collect just enough data to fulfill the purpose and nothing more.

Purpose could capture information about how the data collection will benefit the user. Data collection is done to provide a particular service to a user. Furthermore, the deployment scenario of an IoT device could affect its purpose. For instance, a smart CCTV camera could be deployed in a building for security purposes.

Figure 2 illustrates a hierarchical taxonomy of *purpose* for IoT privacy policy description. The items on our purpose taxonomy are based on the study of previous research and typical IoT setup in a smart building. This implies that our purpose taxonomy can be extended by considering some unique IoT setup. This taxonomy is categorized mainly as *security*, *energy management*, *environmental factors tracking*, *user personalization*, *health*, and *entertainment*. In terms of *security*, the purpose for an IoT deployment is further categorized as *fire safety* or *monitoring*. Additionally, drafting an IoT policy for *energy management* will be for the reason of monitoring the amount of energy consumed by the *HVAC* or *lighting* installations. Some IoT systems are generally deployed for the sole purpose of managing or tracking the concentration of environmental factors on human lives within a smart building. Some of the factors that are usually considered are *temperature*, *carbon monoxide concentration*, *air quality*, and *humidity concentration*. Finally, there are times when an IoT device is deployed for the sole purpose of *user personalization* for either *room temperature* or *room lighting* preferences.

C. Data Retention

Data Retention refers to the maximum time duration or number of accesses after which the data collector is expected to remove the data from their database [6]. Data retention reflects one of the core tenets of many privacy regulations. For instance, privacy regulations such as the GDPR [11], OECD [15], and HIPAA [2], emphasize the need for data collectors to inform data providers of how long the data collected is kept.

Data retention impacts users' privacy preferences based on the duration of the retention and number of data access [18], [22], [23]. Previous research has shown reduced privacy concerns from people on data which they consider less sensitive but have a long retention period. Furthermore, increased privacy concerns has been seen on data which will be accessed frequently based on the specified data retention policy.

The need to retain data in an IoT privacy policy description could depend on definite factors and in, some instances, be related to other IoT privacy policy tenets. *Figure 3* illustrates the hierarchical data retention taxonomy for IoT privacy policy description. Data retention can be categorized as *stated period*, *indefinite*, and *number of access* based on typical IoT privacy policies of a smart building. In terms of *stated period*, there is a specified duration of time for IoT data collection. There are instances where the IoT data collected is kept for an *indefinite* period of time because of the peculiar need of the data, such as, a users' biometric fingerprint for personalized privacy policy. Finally, there are instances where the IoT data collected is specified with data retention on *number of access*. This factor can further be categorized as either *one time access* or *multiple accesses*.

D. Data Type Collected

Different IoT devices collect different types of information based on the sensors embedded in them. For example, a smart thermostat usually has, at least, a temperature and motion sensor for collecting temperature and room occupancy information, respectively. Details about the type or category of data collected by an IoT device is key in a privacy policy because it makes users aware of what information about them is collected. A number of surveys [27], [33] on users IoT privacy preferences have highlighted the importance of creating awareness on the type of data collected by an IoT device.

The type of data collected by an IoT device is identifiable in different forms. *Figure 4* illustrates the hierarchical data type collection taxonomy for IoT privacy policy description. Generally, the data types for IoT transaction processing is broadly categorized as, *location*, *biometric*, and *environment*. In terms of the *location*, the specific coordinate details in a smart building will indicate or trigger the collection of data, such as, *human motion*, *lighting*, *video*, *etc.* Data in the form of *biometrics* also indicate user-specific information in identifying individuals so the IoT devices to deliver such user-specific services. Biometric data is usually in the form of *facial image*, *voice*, *retina*, *fingerprint*, or *gait*. *Environmental data* is processed by some IoT devices to monitor environmental

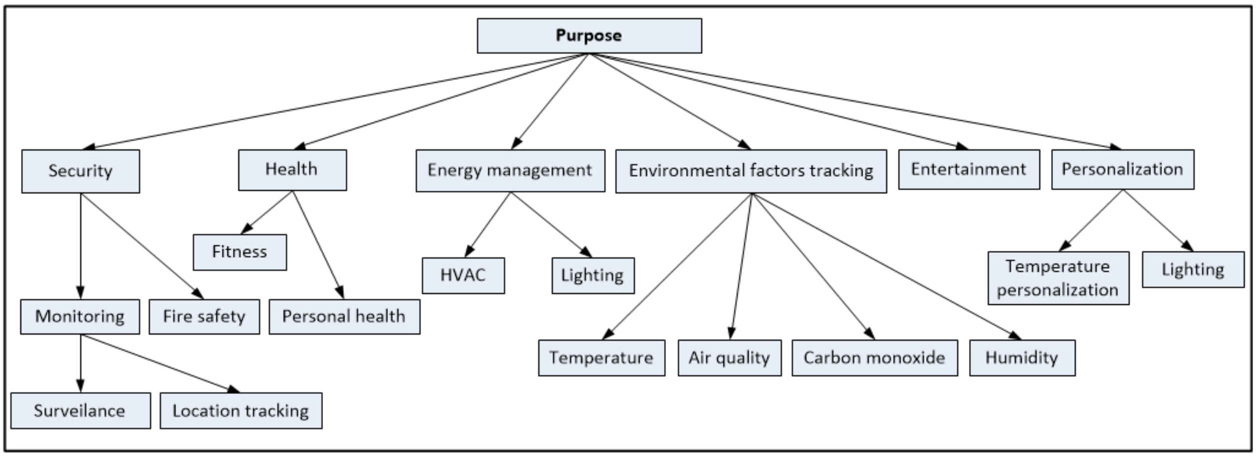

Fig. 2: Purpose

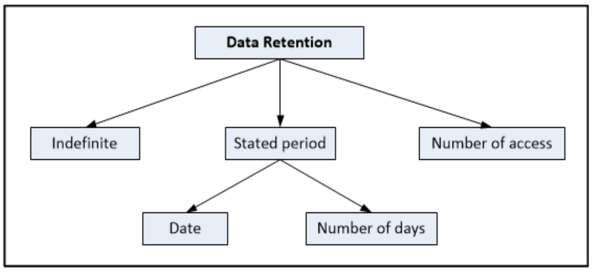

Fig. 3: Data Retention

factors, such as, *temperature*, *carbon monoxide*, *air quality*, and *humidity*, amongst others.

E. Data Granularity and Inference

Data granularity refers to the level of data precision that is stored in a repository. It could indicate what type of information can be inferred about a person. Dirkzwager *et al.* [12] study the data collection practices of the smart home giant, *Nest*. Their study reveals that Internet Protocol (IP) address and location of users are collected by the devices and these are used to uniquely identify a user. Hence, the definition of the granularity of collected data in a privacy policy allows users' to know what information about them can be inferred.

In *Figure 5*, we illustrate the hierarchical data granularity and inference taxonomy for IoT privacy policy descriptions. We categorized data granularity as *exact* and *partial*. An *exact* level of granularity means the data collected by an IoT device is not perturbed during the data transmission, processing and storage. Data at this level of granularity could be used to uniquely identify an individual and vice versa. For example, the collection of biometric data, such as fingerprint, can be used to uniquely identify an individual. However, environmental data such as temperature, cannot be used in itself to identify an individual.

The *partial* level of data granularity means the data is perturbed after collection by an IoT device to reduce its preciseness and make it less identifiable. Two key approaches for perturbing data are: *generalization* and *anonymization*.

F. Sensor Activation

Sensor activation defines factors which trigger data collection by IoT devices. For example, with a smart thermostat, the presence of someone in the room triggers the collection of temperature information by the thermostat to regulate the room temperature [37]. Some IoT devices, like a *smart camera*, which performs image recognition could pervasively collect data about their surrounding without the need for activation.

Figure 6 illustrates the hierarchical sensor activation taxonomy for IoT privacy policy description. The idea of sensor activation in IoT privacy policies are generally categorized as *human motion*, *temperature*, *carbon monoxide*, *air quality*, and *humidity concentration*. These kinds of related data constitute the primary data elements that activate sensors.

G. Data Recipients

A data recipient is an entity or individual who processes (or uses) the data provided to, either deliver service to the intended users (data providers) or fine-tune IoT device performance. While information about the data collector is key in a privacy policy, information about any third party data processor is equally important in the privacy policy. The data collected by IoT devices are sometimes shared with third-parties to aid service delivery. Though the European General Data Protection Regulation (GDPR) [11] makes the data collector accountable for any data misuse by third party data processors, providing information about data recipients creates transparency as to how users data will be managed and used.

In *Figure 7*, we illustrate the data recipients taxonomy for IoT privacy policy descriptions. The data recipients taxonomy is generally categorized into *device manufacturer*, *service provider*, and *third-party*. These categories indicate that for

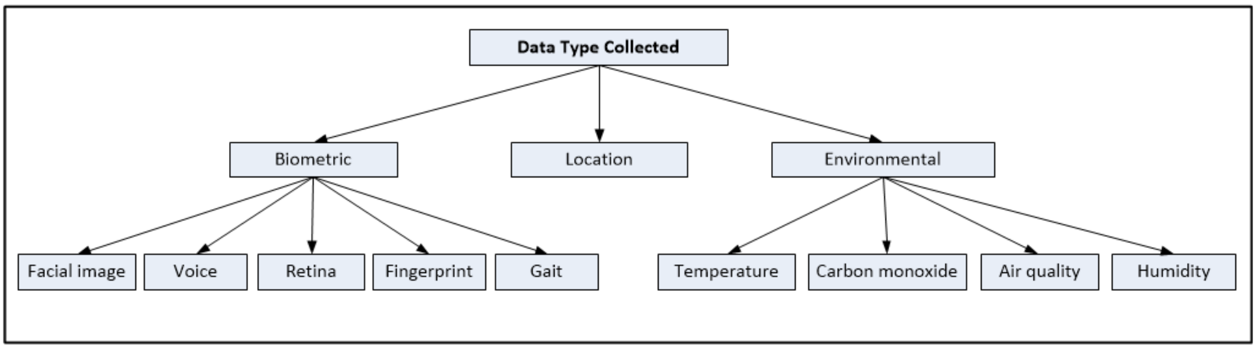

Fig. 4: Data Type Collected

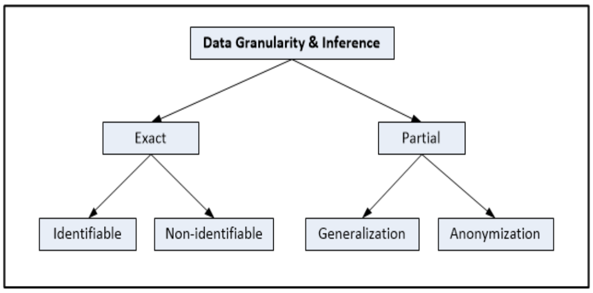

Fig. 5: Data Granularity and Inference

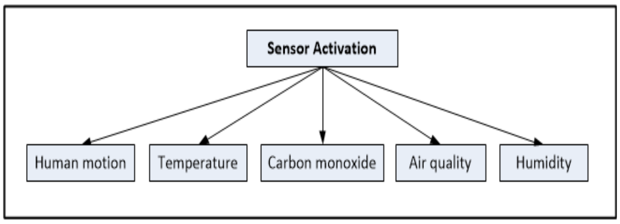

Fig. 6: Sensor Activation

the data collected from the users (data providers), there are instances where the data will be used solely by the *device manufacturer* to fine tune the performance of the IoT device. In other instances, the data collected will be used by the *service provider* to deliver efficient services to meet user needs. Additionally, the *device manufacturer* and *service provider* could be the same entity in which case it is the same vendor who manufactures the IoT device and offers service utility on the device platform for use by the data providers (users). Data collected and delivered to *third-party* recipients are data that these providers require to deliver additional services.

H. Context

Conventionally, contextual information is not captured in a privacy policy, but with the IoT it is important because it uncovers information about the deployment environment of the IoT device. For instance, context could describe the meta information of a smart building such as who is responsible for data collection, location of sensors, *etc.* [36]. The most important contextual information for the IoT is *location*, because previous research has shown that the privacy expectations of people varies with location context [27], [33]. While some people might be comfortable with the data collection practices of a smart thermostat in an office environment, some might be uncomfortable with such collection in a private space like their homes as it might reveal sensitive information.

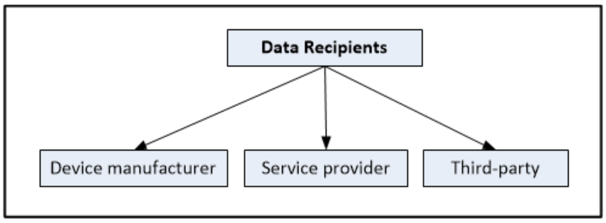

Fig. 7: Data Recipients

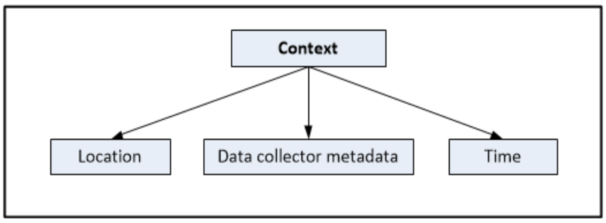

Fig. 8: Context

The *context* for IoT data collection, storage, access, and sharing varies from one IoT environment to another, even within the same application domain. We illustrate the context taxonomy for IoT privacy policy description in *Figure 8*. Context is mainly categorized as *location*, *data collector metadata*, and *time*. Most forms of transactional data processing for IoT systems are related to *location* semantics. The specific location co-ordinates of the user (data provider) is considered during data collection and processing. Additionally, the context for IoT data collection and processing factors in *data collector metadata* such as data collector name and designation. The information about the data collector and how data is to be used gives a clearer context for IoT data processing. The *time* context provides information about when the data collection took place.

IV. Privacy Formalization

A. Privacy Policy Formalization

In general, an IoT privacy policy contains one or more privacy statements which defines how collected data will be used or managed. This means that a privacy policy is a set of privacy statements and a privacy statement represent the privacy practices for the IoT device. However, not all privacy statements in a policy are important in the service delivery process of the IoT device. Hence, some privacy policies provide choice to users by allowing them to opt-in and opt-out from those privacy statements.

Definition 1. (Privacy Policy): Privacy Policy, $\mathscr{P}$, is a set of privacy statements, that is $\mathscr{P} = \{S_1, S_2, \ldots, S_n\}$ where S represent a privacy statement and n is the number of statements in the policy.

Definition 2. (Privacy Statement): A privacy statement, S, represents how a piece of data collected from an individual will be managed and used. Formally, a privacy statement is a tuple which contains a data element D, privacy practices Pa, context C, and compulsory status W, that is $S_i =< D_i, Pa_i, C_i, W_i >$ where i is a data items identifier. S_i represent a privacy statement in a privacy policy. D_i represent the type of data collected. Pa_i is a tuple which contains privacy practices for the collected data, which include elements such as purpose (p), retention (r), data recipients (q), and granularity (g). That is, $Pa_i =< p_i, r_i, q_i, g_i >$. C_i represent context for the data collection such as the location where the collection will take place. Finally, W_i indicates if a user can opt-out from the data collection.

V. Application of Privacy Model

This section discusses the utility of our privacy taxonomy and its application to various domains. The application is explained through three major examples. First, the application of our methodology in the classification of existing research work. Second, the application to demonstrate the usability of our privacy model in a smart building environment. Third, the application of a real privacy policy of an IoT device manufacturer (or vendor).

A. Example 1: (Classification of Existing Research Works)

There are several research efforts that leverage privacy policy in their research to create IoT privacy protection schemes. This work considers certain dimensions of IoT privacy policy or preference based on the nature of their research. We use two of these contributions to demonstrate the robustness of our privacy taxonomy.

Pappachan *et al.* [36] propose a framework for protecting users' privacy preferences in a smart building environment. The framework involves an IoT assistant that captures and transfers the privacy preferences of a user to a privacy-aware smart building for enforcement during any data collection or third-party data sharing process. Based on the privacy policy description in their work, they considered purpose, retention, type of data collected, and context. Using our taxonomy, the sample privacy policy in their work for a smart building at the University of California has a purpose of *security*, a data retention of *stated period*, a type of data collected of *location* and context of *location* and *data collector metadata*.

Lee *et al.* [26], carried out a survey using 200 participants on Amazon Mechanical Turk to understand users' privacy preferences in an IoT environment. The privacy preference data is analyzed using *k*-mode clustering algorithm to discover factors that affect users' preferences in an IoT environment. The IoT scenarios for the survey are constructed using five contextual parameters: where, what, who, reason, and persistence. The privacy preferences of users is captured using five reaction parameters: notification, permission, comfort, risk, and appropriateness. The privacy policy description in their work considers collection frequency, purpose, type of data collected and context. The contextual parameters in their work corresponds to certain dimensions on our taxonomy: *where* is equivalent to context, *what* is equivalent to type of data collected, *who* is equivalent to data recipient, *reason* is equivalent to purpose, and *persistence* is equivalent to collection frequency.

Finally, Naeini *et al.* [33] study users' privacy expectations and factors that affect there preferences in an IoT environment. The study involved 1,007 participants on Amazon Mechanical Turk and 380 IoT data collection scenarios. The resulting data were analyzed using machine learning algorithms to discover useful insights on privacy preferences. Each participant is presented with 14 IoT scenarios which are varied across eight factors: type of data collected, location of data collection, user benefit, data collecting device, purpose, retention, data sharing, and inference. Based on these eight factors in their work, in relation to our taxonomy, considered *purpose*, *retention*, *type of data collected*, *granularity*, *data recipients*, and *context*.

Table I illustrates our proposed formal IoT privacy policy descriptions and taxonomy in this paper, to classify the research studies discussed.

B. Example 2: (Ecobee)

Ecobee is a Canadian-based company that manufacture IoT devices, such as, smart thermostats. The privacy statement below from the website [14] of Ecobee describes the privacy policy regarding usage of the IoT device by data providers (users).

> *Your Ecobee Device will collect environmental data such as temperature and humidity as well as operational data such as temperature set points (i.e., "runtime" data) from your HVAC equipment. Some Device models may include additional types of data such motion sensing (i.e., "occupancy sensing"). Depending on your Device model, your Device may*

TABLE I: Classification of Existing Research Works

Research	Collection Freq.	Purpose	Retention	Data Type	Granularity	Sensor Activation	Data Recipients	Context
Pappachan *et al.* [36]		✓	✓	✓				✓
Lee *et al.* [26]	✓	✓		✓			✓	✓
Naeini *et al.* [33]		✓	✓	✓	✓		✓	✓

> *also collect data from remote sensors in addition to the Device itself. This environmental and operational HVAC data and is used to optimize the heating and cooling algorithms on your Device to minimize energy usage when you are home or away.*

TABLE II: Ecobee Privacy Statement Taxonomy Representation

Taxonomy Dimensions	Instance Values
Data Collection Frequency	Continuous
Purpose	Energy Management
Data Retention	Indefinite
Data Typed Collected	Environmental
Data Granularity	Exact
Sensor Activation	Human Motion
Data Recipients	Service Provider
Context	Not Applicable

Table II shows the taxonomy representation of the stated privacy statement. The data collection frequency is *continuous* since the device constantly collects environmental and operational HVAC data. The purpose for the data collection is *energy management* considering the device is used to minimize energy consumption. The data retention is *indefinite* since the retention period was not explicitly specified. The type of data collected is *environmental* as it entitles temperature and humidity data. Though the transmission of data is encrypted, there is no indication from the privacy statement that the data is perturbed. Hence, the data granularity is *exact*. The sensor activation is *human motion* since data collection by the smart thermostat could be triggered by room occupancy. In this scenario, it can be seen that Ecobee is both the device manufacturer and data collector, and that is why the data recipient is *service provider*. The context is *not applicable* because device manufacturers might be unaware of the device deployment and usage.

The formal IoT privacy policy for *Example 2* is expressed in the following privacy statement:

$$s_1 =< Environmental,\ Pa,\ Not\ Applicable,\ Mandatory >$$
$$Pa =< Energy\ Management,\ Indefinite,\ Ecobee,\ Exact >$$

In the above privacy statement, S, *environmental* represent the type of data collected by the smart thermostat, Pa represents the privacy practices, *not applicable* means context is not under consideration, and *mandatory* means that a user cannot opt-out from the privacy statement.

In the privacy practices, Pa, expressed above, *energy management* is the instance for the purpose of data collection, *indefinite* is the instance for data retention period, *Ecobee* is the entity representing the instance for data recipient, and *exact* instantiate a representation for the data granularity; which means the data is not perturbed.

VI. Conclusion

We propose a framework for understanding and modeling IoT privacy policies within the context of smart environments, such as, smart buildings. This research identifies key components from previous research, IoT device manufacturer privacy policies, and typical IoT environment setup - that can be used to efficiently understand IoT privacy policies. Apart from conventional privacy policy elements, we identified *data collection frequency*, *sensor activation*, and *context*, as important elements for modeling IoT privacy policy. Additionally, we present a formal representation for privacy policy, which is also based on our taxonomy. Finally, we demonstrate the utility of our taxonomy and privacy model by evaluating it on two application scenarios. The first application scenario is the classification of the contributions from previous research and the second example is based on the modeling of a real privacy policy from an IoT device manufacturer (Ecobee).

The outcome of this research could be used to model privacy policy of IoT devices in a smart building environment, which is a key domain in current IoT privacy research. Our formalization is important for the conversion of a privacy policy into a machine-readable format. This will generally allow us to be able to compare two privacy policies.

As future work, we plan to broaden the scope of the taxonomy framework to handle other forms of IoT devices. We envisage the expansion of the taxonomy elements to handle IoT devices configured into smart environments or platforms, such as, smart cities (which may be composed of smart traffic lights, smart roads, smart CCTV cameras), smart energy management, smart waste and pollution management, smart infrastructure and transportation systems, and smart utility grids, amongst others. Additionally, we plan to evaluate users' preferences based on our taxonomy.

References

[1] Gregory D. Abowd, Anind K. Dey, Peter J. Brown, Nigel Davies, Mark Smith, and Pete Steggles. Towards a Better Understanding of Context and Context-Awareness. In *International symposium on handheld and ubiquitous computing*, pages 304–307. Springer, 1999.

[2] Accountability Act. Health Insurance Portability and Accountability Act of 1996. *Public law*, 104:191, 1996.

[3] Anne Anderson, Anthony Nadalin, B. Parducci, D. Engovatov, H. Lockhart, M. Kudo, P. Humenn, S. Godik, S. Anderson, S. Crocker, et al. Extensible Access Control Markup Language (XACML) Version 1.0. *OASIS*, 2003.

[4] Paul Ashley, Satoshi Hada, Günter Karjoth, Calvin Powers, and Matthias Schunter. Enterprise Privacy Authorization Language (EPAL). *IBM Research*, 30:31, 2003.

[5] Luigi Atzori, Antonio Iera, and Giacomo Morabito. The Internet of Things: A Survey. *Computer Networks*, 54(15):2787 – 2805, 2010.

[6] Ken Barker, Mina Askari, Mishtu Banerjee, Kambiz Ghazinour, Brenan Mackas, Maryam Majedi, Sampson Pun, and Adepele Williams. A Data Privacy Taxonomy. In *British National Conference on Databases*, pages 42–54. Springer, 2009.

[7] Adam Barth, Anupam Datta, John C. Mitchell, and Helen Nissenbaum. Privacy and Contextual Integrity: Framework and Applications. In *2006 IEEE Symposium on Security and Privacy (S&P'06)*, pages 15–pp. IEEE, 2006.

[8] Lorrie Cranor, Marc Langheinrich, and Massimo Marchiori. A P3P Preference Exchange Language 1.0 (APPEL1. 0), W3C Working Draft 15 April 2002. *World Wide Web Consortium (W3C), URL: http://www. w3. org/TR/P3P-preferences*, 2002.

[9] Lorrie Cranor, Marc Langheinrich, Massimo Marchiori, Martin Presler-Marshall, and Joseph Reagle. The Platform for Privacy Preferences 1.0 (P3P1. 0) specification. *W3C recommendation*, 16, 2002.

[10] Anupam Das, Martin Degeling, Daniel Smullen, and Norman Sadeh. Personal Privacy Assistants for the Internet of Things.

[11] Paul De Hert and Vagelis Papakonstantinou. The Proposed Data Protection Regulation Replacing Directive 95/46/EC: A Sound System for the Protection of Individuals. *Computer Law & Security Review*, 28(2):130–142, 2012.

[12] Aimee Dirkzwager, J. Cornelisse, T. Brok, and L. Corcoran. Where Does Your Data Go? Mapping the Data Flow of Nest. *Masters of Media*, 2017.

[13] Paul Dourish. What We Talk About When We Talk About Context. *Personal and Ubiquitous Computing*, 8(1):19–30, 2004.

[14] Ecobee. Privacy Policy Terms of Use, 2019.

[15] Organisation for Economic Co-operation and Development. *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*. Organisation for Economic Co-operation and Development, 2002.

[16] Kambiz Ghazinour, Maryam Majedi, and Ken Barker. A Model for Privacy Policy Visualization. In *2009 33rd Annual IEEE International Computer Software and Applications Conference*, volume 2, pages 335–340. IEEE, 2009.

[17] Serge Gutwirth, Yves Poullet, Paul De Hert, and Ronald Leenes. *Computers, Privacy and Data Protection: An Element of Choice*. Number 2006. 2011.

[18] Markus Hinkelmann and Andreas Jakoby. Preserving privacy versus data retention. In *Theory and Applications of Models of Computation, 6th Annual Conference, TAMC 2009, Changsha, China, May 18-22, 2009. Proceedings*, pages 251–260, 2009.

[19] Vincent C. Hu, D. Richard Kuhn, and David F. Ferraiolo. Attribute-Based Access Control. *Computer*, 48(2):85–88, 2015.

[20] Mohammad Jafari, Philip W. L. Fong, Reihaneh Safavi-Naini, Ken Barker, and Nicholas Paul Sheppard. Towards Defining Semantic Foundations for Purpose-Based Privacy Policies. In *First ACM Conference on Data and Application Security and Privacy, CODASPY 2011, San Antonio, TX, USA, February 21-23, 2011, Proceedings*, pages 213–224, 2011.

[21] Bilal Javed, Mian Waseem Iqbal, and Haider Abbas. Internet of things (IoT) Design Considerations for Developers and Manufacturers. In *2017 IEEE International Conference on Communications Workshops, ICC Workshops 2017, Paris, France, May 21-25, 2017*, pages 834–839, 2017.

[22] Panayiotis Kotzanikolaou. Data retention and privacy in electronic communications. *IEEE Security & Privacy*, 6(5):46–52, 2008.

[23] Panayiotis Kotzanikolaou and Christos Douligeris. Privacy threats of data retention in internet communications. In *Proceedings of the IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2007, 3-7 September 2007, Athens, Greece*, pages 1–4, 2007.

[24] Ponnurangam Kumaraguru, L. Cranor, Jorge Lobo, and Seraphin Calo. A Survey of Privacy Policy Languages. In *Workshop on Usable IT Security Management (USM 07): Proceedings of the 3rd Symposium on Usable Privacy and Security, ACM*, 2007.

[25] Marc Langheinrich. A Privacy Awareness System for Ubiquitous Computing Environments. In *International Conference on Ubiquitous Computing*, pages 237–245. Springer, 2002.

[26] Hosub Lee and Alfred Kobsa. Understanding user privacy in internet of things environments. In *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*, pages 407–412. IEEE, 2016.

[27] Hosub Lee and Alfred Kobsa. Privacy Preference Modeling and Prediction in a Simulated Campuswide IoT Environment. *2017 IEEE International Conference on Pervasive Computing and Communications, PerCom 2017*, pages 276–285, 2017.

[28] Georgios Lilis, Gilbert Conus, Nastaran Asadi, and Maher Kayal. Towards the Next Generation of Intelligent Building: An Assessment Study of Current Automation and Future IoT based Systems with a Proposal for Transitional Design. *Sustainable Cities and Society*, 28:473–481, 2017.

[29] Huichen Lin and Neil Bergmann. IoT Privacy and Security Challenges for Smart Home Environments. *Information*, 7(3):44, 2016.

[30] Yi-Ning Liu, Yan-Ping Wang, Xiao-Fen Wang, Zhe Xia, and Jingfang Xu. Privacy-preserving Raw Data Collection without a Trusted Authority for IoT. *Computer Networks*, 148:340–348, 2019.

[31] Mateusz Mikusz, Steven Houben, Nigel Davies, Klaus Moessner, and Marc Langheinrich. Raising Awareness of IoT Sensor Deployments.

[32] Daniele Miorandi, Sabrina Sicari, Francesco De Pellegrini, and Imrich Chlamtac. Internet of Things: Vision, Applications and Research Challenges. *Ad Hoc Networks*, 10(7):1497–1516, 2012.

[33] Pardis Emami Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer, Lorrie Cranor, and Norman Sadeh. Privacy Expectations and Preferences in an IoT World. *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, (Soups):399–412, 2017.

[34] Helen Nissenbaum. Privacy as Contextual Integrity. *Wash. L. Rev.*, 79:119, 2004.

[35] Daniel E. O'Leary, S Bonorris, W Klosgen, Yew-Tuan Khaw, Hing-Yan Lee, and W Ziarko. Some privacy issues in knowledge discovery: the oecd personal privacy guidelines. *IEEE Expert*, 10(2):48–59, 1995.

[36] Primal Pappachan, Martin Degeling, Roberto Yus, Anupam Das, Sruti Bhagavatula, William Melicher, Pardis Emami Naeini, Shikun Zhang, Lujo Bauer, Alfred Kobsa, et al. Towards Privacy-Aware Smart Buildings: Capturing, Communicating, and Enforcing Privacy Policies and Preferences. In *2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW)*, pages 193–198. IEEE, 2017.

[37] Andreas P. Plageras, Kostas E. Psannis, Christos Stergiou, Haoxiang Wang, and Brij B. Gupta. Efficient IoT-based sensor BIG Data collection-processing and analysis in smart buildings. *Future Generation Comp. Syst.*, 82:349–357, 2018.

[38] Taj Rahman, Xuanxia Yao, and Gang Tao. Consistent Data Collection and Assortment in the Progression of Continuous Objects in IoT. *IEEE Access*, 6:51875–51885, 2018.

[39] Richard L Rutledge, Aaron K Massey, and Annie I Antón. Privacy Impacts of IoT Devices: a SmartTV Case Study. In *2016 IEEE 24th International Requirements Engineering Conference Workshops (REW)*, pages 261–270. IEEE, 2016.

[40] Vasilios A. Siris, Nikos Fotiou, Alexandros Mertzianis, and George C. Polyzos. Smart application-aware IoT data collection. *J. Reliable Intelligent Environments*, 5(1):17–28, 2019.

[41] Nicolas Sklavos and Ioannis D. Zaharakis. Cryptography and Security in Internet of Things (IoTs): Models, Schemes, and Implementations. In *8th IFIP International Conference on New Technologies, Mobility and Security, NTMS 2016, Larnaca, Cyprus, November 21-23, 2016*, pages 1–2, 2016.

[42] John A. Stankovic. Research directions for the internet of things. *IEEE Internet of Things Journal*, 1(1):3–9, 2014.

[43] Sweeney, L. k-Anonymity: A Model for Protecting Privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst*, 10(05):557–570, 2002.

[44] Xiaokui Xiao and Yufei Tao. Personalized privacy preservation. In *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*, pages 229–240. ACM, 2006.

[45] Yuchen Yang, Longfei Wu, Guisheng Yin, Lijie Li, and Hongbin Zhao. A Survey on Security and Privacy Issues in Internet-of-Things. *IEEE Internet of Things Journal*, 4(5):1250–1258, 2017.

[46] Ni Zhang and Chris Todd. Developing a Privacy Ontology for Privacy Control in Context-Aware Systems. *Dept. of Electronic & Electrical Eng., Univ. College London*, 2006.

[47] Ni Jenny Zhang and Chris Todd. A Privacy Agent in Context-Aware Ubiquitous Computing Environments. In *IFIP International Conference on Communications and Multimedia Security*, pages 196–205. Springer, 2006.

[48] Jan Henrik Ziegeldorf, Oscar Garcia Morchon, and Klaus Wehrle. Privacy in the internet of things: Threats and challenges. *Security and Communication Networks*, 7(12):2728–2742, 2014.