Ontology-based User Privacy Management in Smart Grid

Raisa Islam
New Mexico Tech
Socorro, USA
raisa.islam@student.nmt.edu

Tomas Cerny
Baylor University
Waco, USA
tomas_cerny@baylor.edu

Dongwan Shin
New Mexico Tech
Socorro, USA
dongwan.shin@nmt.edu

ABSTRACT

A smart grid system is one of the most complex cyber-physical systems consisting of power generation, distribution, consumption, customer domains, and millions of connected end devices. The widespread implementation of the smart grid raises concerns about the privacy of the data it collects. Since users' personal and non-personal data are going to be accessed by different entities involved in the smart grid and by third parties, the privacy concern can be a big obstacle in the adoption of the smart grid among people. Hence, there is an urgent need to provide the user with a privacy solution to support selective sharing of their usage data with different entities. In this paper, we propose an ontology-based user privacy management approach that will enable the user to release their data based on sensitivity and privacy factors, thus making an informed privacy decision on their usage data sharing. Green Button Initiative is a smart grid application that allows users to download and share their energy usage data with third parties. We present a proof-of-concept implementation extending the Green Button Initiative to test the feasibility of the proposed approach. Lastly, we discuss the results of a user study to investigate the effectiveness of our proposed approach.

ACM Reference Format:
Raisa Islam, Tomas Cerny, and Dongwan Shin. 2022. Ontology-based User Privacy Management in Smart Grid. In *Proceedings of ACM SAC Conference (SAC'22).* ACM, New York, NY, USA, Article 4, 9 pages. https://doi.org/10.1145/3477314.3508383

1 INTRODUCTION

Smart grid (SG) system is one of the most complex cyber-physical systems (CPS) consisting of power generation, distribution, consumption, multiple customer domains, and millions of connected end devices. Smart meters are the building blocks of the SG systems. The *Institute for Electric Innovation* (IEI) [1] estimated that in the U.S. about 107 million smart meters were installed at year-end of 2020 and the total number of installed smart meters would rise to 115 million by the end of 2021. Due to the widespread implementation, many attacks in SG systems have been reported in past years. Known attacks often focus on discovery, capture, and exfiltration of data [2]. It is also shown that by using energy consumption data, i.e., bills and electricity report, it is possible to extract complex usage patterns using off-the-shelf statistical methods [3, 4]. From these data, it is easy to find out the behavioral pattern, lifestyle and economical condition of consumers. Hence, privacy concerns about those data could be a significant roadblock towards SG system development and adoption.

Statistical and cryptographical solutions to allow consumers to set privacy policies have been proposed to thwart different types of privacy attacks against SG systems in the past [5, 6, 7, 8, 9]. However, most of them have limited support for consumers who want to share their data selectively with power companies or third parties, which could collect, store, and access sensitive consumer data to facilitate their services. In addition, since the data are exchanged between the consumer, power/service companies, and other third parties, it is extremely important to have a common data model for them to use for their SG system design and implementation. The data model will allow different SG components to understand each other and enforce a common privacy policy enacted for the user data for all the different parties. An ontology is described as a specification of concepts and relationships in a domain. The concepts in the ontology are formally defined as objects, classes, attributes, relations and restrictions. Due to the complex topology of SG, different components need to communicate with each other to conduct various tasks. The primary benefit of using ontology in a SG network is that it resolves the interoperability issue across heterogeneous components [10] by providing a common semantic.

In this paper, we propose an ontology-based user-controllable privacy management mechanism where end users will be able to manage their privacy attributes, control the exposure of those attributes and set up specific privacy policy based on the level of risk factor associated with each attribute. The contributions of our paper are mainly as follows:

- First, we designed an ontology for SG which can be used as a common data model to support communication between different entities i.e. consumer, service provider and third-parties.
- Second, we discussed a proof-of-concept implementation based on Green Button Initiative (GBI) which provides consumers a way to access their energy usage data [1]. We applied and extended the ontology to accommodate the GBI framework. Users are allowed to set up specific privacy preferences based on the level of risks associated with each attribute.
- Lastly, we designed and conducted a comparative user study. The goal of this user study is to investigate the effectiveness of our privacy management mechanism against the existing privacy management mechanisms.

The remainder of this paper is organized as follows: In Section 2, we present background and related work, and our approach to

[1] Information derived from www.greenbuttonalliance.org

SAC'22, March 25-29, 2022, Brno, Czech Republic

ACM ISBN 978-1-4503-5933-7/19/04…$15.00
https://doi.org/10.1145/3477314.3508383

developing an ontology model for use in SG is presented in Section 3. Section 4 discusses our proof-of-concept implementation based on GBI. Our user study and results are described in Section 5 and 6. Section 7 concludes this paper.

2 BACKGROUND AND RELATED WORK

2.1 Ontologies In SG

There have been prior approaches to designing an ontology for use in SG systems in literature. An integrated SG information model using semantic web techniques along with a case study was proposed in [11]. This model is extensible in a sense that it is capable of handling the large volume and rapid growth of data while also being able to manage the large discrepancy and variation among the different data types and sources. However, instead of creating a completely new ontology, it proposes a model to facilitate a dynamic data retrieval mechanism by analyzing and integrating the existing ontologies for the different components. Another ontology for SG was presented in [12] where the main focus is on distributed energy feeder with a consideration for its multilayered prosumer-centric components and network. The layered approach ensures that any addition or subtraction of components from the existing system will only require a change in that specific layer, not the entire model. Simmhan et al. proposed an integration approach between SG domain and external domain ontologies as a unified data model in [13]. An SG domain ontology is the one used by utility companies and systems, while external ontology is of the elements outside the SG, i.e., weather information, structural information, or social media information of the users. Their approach was focused more on the practical aspect of data handling in SG rather than establishing a conceptual data model.

To support interoperability, Santodomingo et al. proposed an ontology matching mechanism to facilitate interactions between common information model (CIM)-based and substation configuration language (SCL)-based systems in [14]. The proposed mechanism partially succeeded in designing the generic ontology matching system though aligning CIM and SCL based systems, which means either the matching system needs to achieve significantly more accuracy or a more unified standard data model. The consumer can play an active role in SG by becoming a prosumer. So there arises a necessity to manage production and consumption effectively as well as participating in demand response. [15] proposed a set of ontologies and presented a case study to enable interoperability between different types of simulators, i.e., electricity market, SG and residential energy management to develop a decision-support system.

2.2 Ontology based Privacy Management

There have been prior ontology-based privacy management for other domains such as the Web and the mobile computing in literature [10, 16, 17].

Built on top of Web Access Control (WAC), Privacy Preference Ontology (PPO) provides a platform-independent, fine-grained privacy management mechanism which enables the users to set up data access preferences [16]. The privacy preferences define which resources to protect, which access control privileges to grant and the attributes a requester must satisfy in order to access resources. It can also handle specific policies, such as allowing only one type of specific data. [17] proposed a ontology-based privacy preserving mechanism for big data which can measure the similarity degrees and cost of disclosing private data. Based on these values, users are warned when they choose to release data to services. Lastly, Loukil at el. proposed an approach that is close to the framework we propose in this paper [10]. They discussed the design and implementation of a new privacy ontology called LIoPY for IoT devices. In their framework, data owners are allowed to define the privacy permission settings before the data collection phase. However, the approach was for IoT devices, not SG systems encompassing a much wider diversity and complexity of devices.

2.3 Existing Privacy Management in SG

Several approaches have been proposed to tackle the privacy issue in SG. Access control mechanisms are one of the most popular privacy preserving mechanisms as they allow the user to control who can access their data; a privacy preserving mechanism against the service provider was proposed in [6] where consumers can grant access right over meter readings to a service provider at a time granularity; Speiser et al. proposed a semantic-web technology based privacy model that allows automated access control [7]. The proposed system is based on a traditional contractual model where users set up data policies and, based on the requester's intent, accept or reject the request; and a multi-agent user-centric privacy manager was proposed to protect sensitive information along with different levels of controls on metering data and personal identifiable information (PII) in [8]. Compared to these solutions, our approach is much more comprehensive in modeling data, policy specification, and has an implementation based on the existing GBI framework. On the other hand, other solutions provided in [6, 7, 8] work only with pre-defined policies which do not meet the dynamic requirement of the data requester.

A data sanitization based privacy mechanism is discussed in [18] which removes sensitive/identifiable data before sharing with any party. Authors adopted k-clustering based data anonymization and then applied secure multi-party computation (homomorphic encryption) technique to perform data aggregation. [19] enhanced user privacy by adding noise to power usage data and using non-interactive zero knowledge proofs. Methods proposed in [9, 20, 18, 19] are computationally expensive due to data processing or using encryption. Lastly, blockchain has become a popular technique for data sharing in recent years. [21] proposed a permissioned blockchain based data sharing method using group signature and secret channel authorization. Smart contracts are used to ensure the privacy aspect. In another paper [22], authors used the ring signatures and a one-time address scheme along with consortium blockchain to protect the privacy of both the data owner and data receiver.

3 ONTOLOGY AND PRIVACY RISK

3.1 The Ontology Design

As modern SG systems have become more complex due to the size and complexity of the network, entities, and data involved, it would be extremely difficult to design a standardized ontology which can accommodate all the different models and improvements over time.

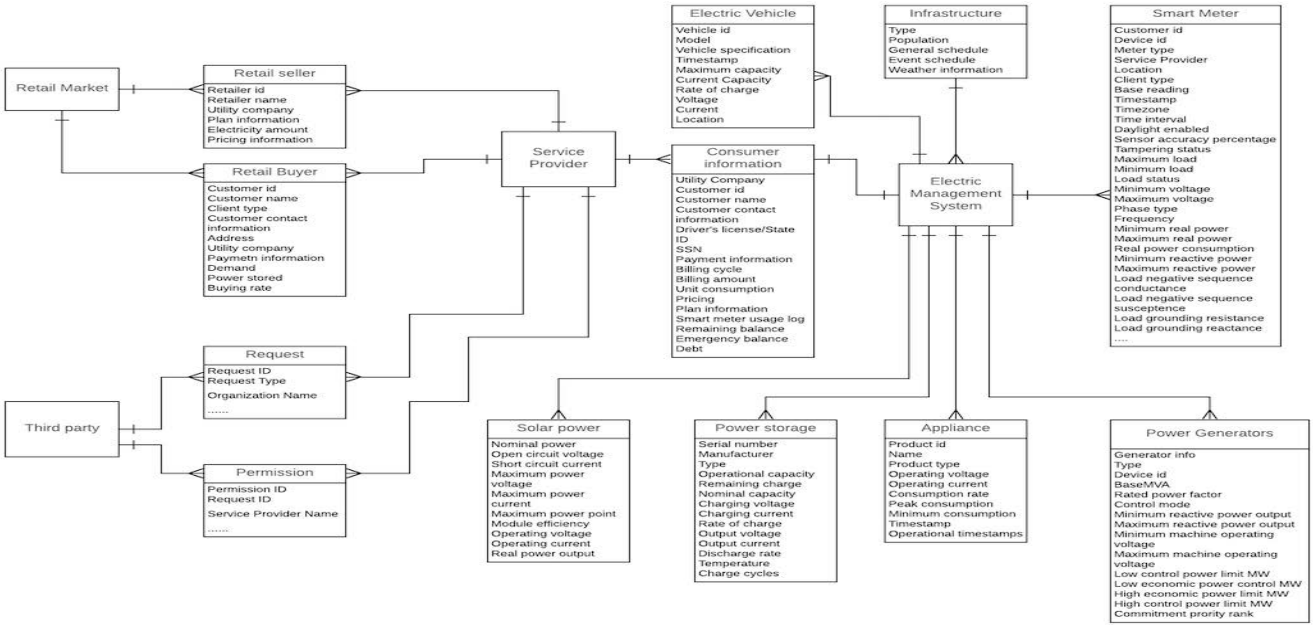


Figure 1: Smart Grid Base Ontology

The more reasonable and scalable approach is to adopt a dynamic modeling approach which has a generic base knowledge model and can be adapted to accommodate any change in the structure and granularity to match a new environment. Hence, we propose a base SG ontology, as shown in Figure 1, where entities, attributes, and their relationships are depicted in the entity-relationship (ER) diagram. The ontology proposed here works as a generic model which can be used as a base model for any new SG environment. A brief description of the entities and attributes is presented below:

(1) Consumer information: Consumer information consists of all the information the service provider has about a specific user, i.e., personal and financial information, energy usage data, and so on. If any unauthorized third party gets access to the information, they can use it for malicious purposes which can cause tremendous personal and/or financial damage to the user. Thus, consumer information has the most sensitive attributes and requires a high level of privacy control.
(2) Smart Meters: Smart meter sends the meter reading and other technical information within a specific time interval. The energy consumption rate of a household can be determined by analyzing the readings in the time intervals. Higher consumption rate in a time period indicates that the house is occupied by people and lower consumption indicates that the household is likely to be empty. Analyzing the meter reading over a large period of time can also used to find out behavioral pattern, i.e., when the consumer is at home and not.
(3) Appliances: Information of appliances, i.e., type and number, can give a general idea of the financial condition of a household. Other than that, each appliance has a pattern of energy usage (load signature), by analyzing these data it is possible to figure out household occupancy.
(4) Electric Vehicle: Electric vehicles could contain several privacy related information, such as real-time location information, and thus a location log could reveal details of all the places that the driver has visited. Additionally, the charging state of the vehicle might reveal user movement information.
(5) Electrical Management System (EMS): EMS is a computer-based application which controls and maintains connected devices at home, and communicates with SG systems to manage the electricity flow. EMS collects the connected components' reading, both consuming and producing electricity, and relays to SG systems to ensure a balance between demand and supply.
(6) Request: This represents third parties' access to the customer data. It contains data attributes about third party name, organization information, types of data required, time frame, service provider name, etc.
(7) Permission: Permissions are granted to third parties by the customer to access usage data. It contains identifiers for permission and requests, permission status, timestamp, duration of authorization, etc. The permission contains identifiable information for the consumer, so it is considered sensitive information.
(8) Third Party: The entity or entities other than the service provider who uses the customer data.
(9) Power Storage: Power storage information i.e. battery capacity, charging/discharging rate can be combined with appliances information to get a more accurate idea of a household.
(10) Power Generators: Power generators can give an idea if a property is vacant or occupied.

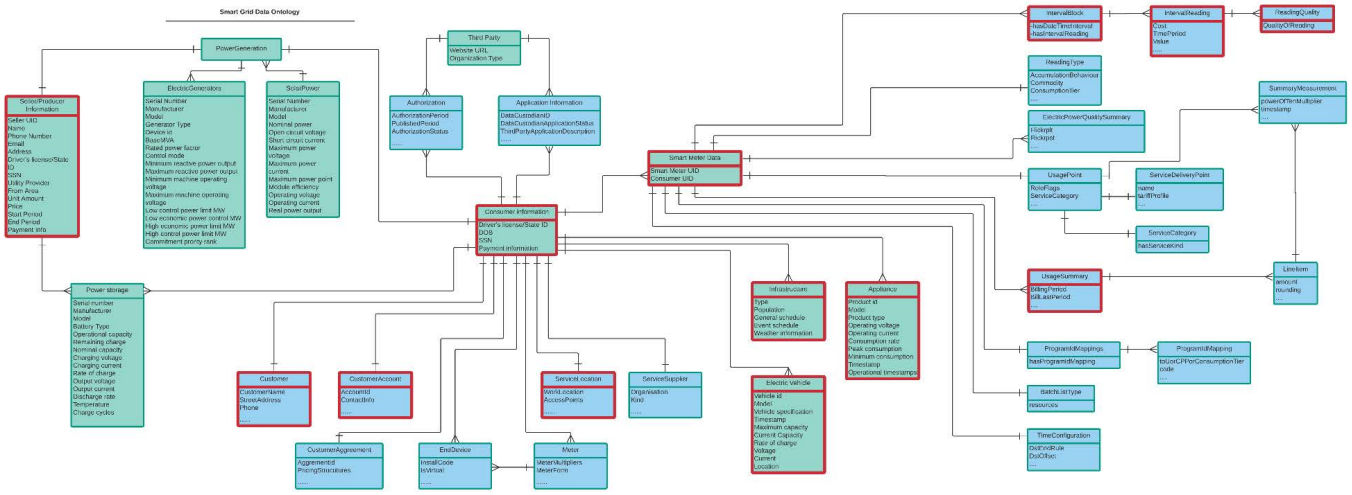


Figure 2: Extended ontology with GBI integration

3.2 Privacy Ratings on Attributes

Each attribute in the ontology is assigned a privacy rating based on the sensitivity of the attribute. The privacy rating is a reflection of how much impact the disclosure of that specific attribute can have on the user's personal life. The numeric range for a privacy attribute rating can be assigned based on the need of an application where the lowest and highest rating indicate least and most sensitive attribute, respectively. By assigning ratings on each attribute, the user can control the degree of exposure he wishes to have. The privacy ratings for each attribute can be divided into primarily three categories based on the different effects it has on a user's personal life.

- Financial factor: Sensitivity of information directly related to a person's financial aspect is represented by financial factors.
- Identifiability factor: Sensitivity of information directly related to the identifiability of a person is represented by identifiability factors.
- Personality factor: Sensitivity of information related to a person's embarrassment in real life is represented by personality factors.

3.3 Privacy Risk Calculation

Privacy risk can be calculated by using attributes and their privacy sensitivities. Following notations are used to define our privacy risk calculator function:

Notation 1: w is the tuple of weight constants for all three privacy factors in the format (w_p, w_f, w_i) which specifies personality, financial and identifiable weight respectively. The value of each element in this set ranges from 0.0 to 1.0. The weight values are used to give users more flexibility in calculating their privacy risk values for attributes.

Notation 2: $\mathcal{A}$ represents the set of all attributes described in Figure 1.

Notation 3: $\mathcal{F}$ is a multi-set where the k^{th} element corresponds to the privacy factor values of $\mathcal{A}_k$ in the set $\mathcal{A}$. An example format is $\{(p_1, f_1, i_1), (p_2, f_2, i_2)..(p_n, f_n, i_n)\}$ where p,f and i denotes personal, financial and identifiable factor, respectively.

Notation 4: r is the output of risk calculator function which represents the cumulative risk factor. r ranges over 0.0 to 1.0, where 0.0 indicates no risk and 1.0 indicates the highest risk.

We calculate risk value for k^{th} attribute in $\mathcal{A}$ as

$$r_k = (w_p.p_k + w_f.f_k + w_i.i_k)/(p^{max} + f^{max} + i^{max})$$

where $p^{max}, f^{max}, i^{max}$ are the maximum possible value of p, f, i respectively. Then the combined risk value is,

$$r = \sum_{k=1}^{n} r_k/n$$

If r value of the requested attributes exceeds the customer's predetermined maximum tolerance limit, the customer will be notified and given options [23]. The options include:

(1) Accept the request: If the user feels that the request is reasonable and do not violate the threshold, the user can accept the request based on the required information.
(2) Reject the request: If the user feels that the required information violates the privacy threshold, the user can reject the request for information.

4 A PROOF OF CONCEPT IMPLEMENTATION

4.1 Extending GBI Schema with the Ontology

We demonstrated the feasibility of our ontology-based approach by extending the base ontology to accommodate the data schema used in GBI. After the extension, the updated ontology is compatible with GBI standards. Specifically, in order to make our ontology compatible with GBI, the following changes have been made. Firstly, we added new entities that are not in the base ontology but in the GBI to the extended ontology. Secondly, for the entities which have

similar semantics between our base ontology and GBI, we used the entities in the base ontology with the superset of attributes from both entities. Finally, we removed the duplicate entities. Figure 2 shows the extended ontology with the GBI integration. Entities in the green background are entities from the base ontology, entities in the blue background are those from GBI, and entities with red borders indicate ones containing highly sensitive privacy information.

4.2 GBI Implementation Framework

Green Button Alliance (GBA) has created a codebase for developers to implement GBI into their own respective web applications for their customers. The codebase has components and functions as follows:

4.2.1 Components. GBI implementation can be categorized into three major components.

(1) OpenESPI-Common-Java : Provides a Common library for Third Party and Data Custodian Applications. Additionally, contains the schema and structure of data and persistence.
(2) OpenESPI-DataCustodian-Java : Provides a user interface for Data Custodians (DC) to download their energy usage data in just one click. Also, it provides a mechanism to approve data authorization requests from third-parties applications.
(3) OpenESPI-ThirdParty-Java : Provides a user interface for Third-party(TP) to request data authentication from Data Custodian and download authenticated energy usage data.

4.2.2 Customer Functions. Green Button sandbox provides two customer usable functions.

(1) Green Button Download My Data® (DMD): The starting point for every customer when they log onto their energy providers' website with Green Button implemented. The DMD provides access to XML formatted energy usage data with a simple click.
(2) Green Button Connect My Data® (CMD): This provides a tool for the applications to request authorization from consumers using OAuth 2.0. In this process, applications request to approve access to consumers' accounts, and consumers provide authorization using a webpage. After the authentication process is done, the application can retrieve energy data without any further involvement of the consumer.

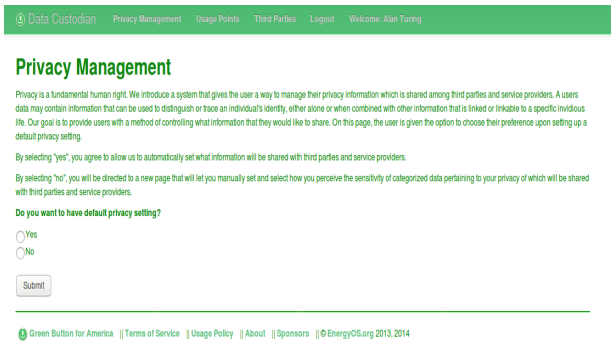
4.3 Privacy Management Implementation

Based on and extended from our ontology-based approach, a user privacy management system was designed and implemented within the Green Button Sandbox framework. While implementing the user-controllable privacy feature, we had to consider the data propagated in various components within the Open-ESPI framework. After carefully analyzing the extended ontology, we considered the following attributes to have a privacy risk.

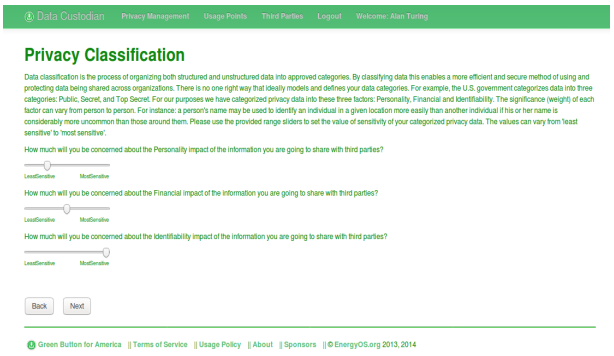
- IntervalBlockValue: Time period (the start and duration) of consumer's energy usage information.
- LocalTimeParameters: Enables the ability to set a local time parameter (timezone) to consumer's energy usage information.
- MeterReading: A utility meter that automatically collects consumption, diagnostic, and status data and transfers that to a central database.
- Commodity: The type of energy resource being read. (E.g., air, natural gas, propane etc.).
- ReadingTypeTimeAttribute: The time period of when consumer's energy usage information was being recorded.
- MeasuringPeriod: Time granularity at which the utility company collects consumer's data.
- ServiceCategory: The specific type of service a meter will read. (E.g., electricity, gas, water, heat etc.)
- OutageRegion: The specific type of service that is currently experiencing outages in the region.

4.3.1 Front-end Component. The front-end implements the user interface (UI) that allows consumers to adjust and set their privacy data values. The user is given the option to modify the sensitivity values that will then display how much risk their settings constitute for them. To implement the Privacy Management UI, we added the following views: home page, classification page, attribute page, and summary page. We have used a three-color indicator to show risk values, red, yellow, and green; red indicates high risk, yellow moderate, and green lowest risk.

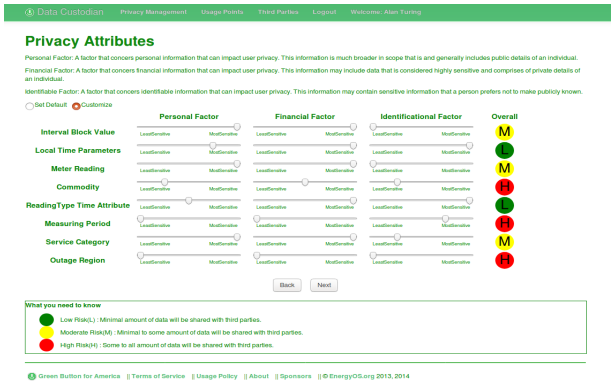
(1) Home page: The homepage displays the starting point of the privacy management setting. The user can choose to have a default privacy setting or opt-out for manual settings, as displayed in Figure 3a. If the user chooses not to have default privacy settings, they will be redirected to the next page, allowing them to customize their privacy settings.
(2) Classification page: This webpage [Figure 3b] is designed to allow the users to set weight values for the three privacy factors. A description of how each factor affects the users' private data is displayed at the top. Using the range slider, users can select how sensitive they are about each factor type. We used a Likert scale for sensitivity values, which ranges from: least sensitive, less sensitive, likely sensitive, more sensitive, to most sensitive. These selected weight values will be used in the back-end component to display information regarding risk values for the user on the next two web pages. 'Back' and 'Next' buttons allow the users to navigate back and forth. Upon clicking 'Next', the user will be directed to the Attribute page.
(3) Attribute page: This allows the user to set sensitivity values for each attribute under every factor category, as shown in Figure 3c. For easier understanding, each factor is described at the top of the page. Again, the user can choose default values or customize each attribute. The default sensitivity value is 'Most Sensitive'. Based on the selections, the 'Overall' column will dynamically display letters and colors. At the bottom of the page, a short description of the color values is displayed.
(4) Summary page: A summary report of the user's overall privacy, shown in Figure 3d, will be displayed with a risk-evaluation-meter that changes dynamically based on the



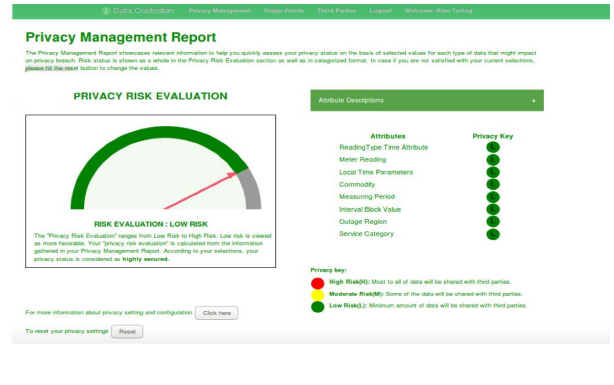
(a) Home Page



(b) Classification Page



(c) Attributes Page



(d) Summary Page

Figure 3: User Interface for Privacy Management within GBI Sandbox

selections. A short description of their risk status is provided below the risk-evaluation-meter. On the right-hand side, a drop-down menu consists of every attribute and their descriptions. Below the drop-down menu is a list that displays the attributes and privacy risk status. In the bottom right of the page, privacy keys are explained. The user can reset all their settings using the 'Reset' button.

4.3.2 Database Component. All the user-specified privacy attribute values are stored in a table named 'PrivacyManagement' with the unique customer ID and XML formatted data. Columns are named as 'id' and 'xmlData'. This table is created by modifying the existing 'datacustodianmysql.sql' file, and to pre-populate some of the user's ratings, we added a new file 'prepopulatesql_usersPrivacy_dc.sql'.

4.3.3 Back-end Component. The back-end implementation is the link between front end and database component. We have implemented classes and objects to fetch data from the database and render them to the UI. We have added the following files: PrivacyFactorController.java, PrivacyManagement.java, RiskCalc.java, PrivacyRatings.java, AttrRating.java, and JAXBParser.java. The controller class, implemented in PrivacyFactorController.java, navigates the flow of data that is transferred to and from the database and the front-end. There are two model classes implemented in PrivacyManagement.java and RiskCalc.java. The former uses methods to retrieve and insert user input into the database. The latter uses methods to calculate risk values to be sent back to be used by the front-end. AttrRating.java consists of an object class that represents single attributes. PrivacyRatings.java class uses attribute objects and calculates the weighted values of the privacy factors and risk levels. Object to XML and XML to object conversion is handled in JAXBParser.java.

5 USER STUDY

We conducted a user study[2] to test if the participants feel more secure and in-control about their privacy after using our proposed ontology based privacy management mechanism in GBI. Additionally, we wanted to investigate the effectiveness by comparing it with the existing privacy mechanisms.

Before starting the user study, we developed a set of hypotheses as follows:

[2]This user study was approved by our university's IRB and conducted for one week in Amazon's Mechanical Turk(MTurk) and Google site.

(1) (**General awareness about data privacy in the smart grid system**) - Consumers are likely to be concerned about usage data when they share them with others.
(2) (**Usefulness of more information to increase the awareness about data privacy**) - Given more information on data privacy users will be more concerned about their data privacy risks when sharing their data in smart grid systems.
(3) (**Ineffectiveness of existing privacy management mechanisms**) - Our hypothesis is that the existing privacy management mechanisms are not effective and consumers do not feel safe to share their data using the existing methods.
(4) (**Effectiveness of ontology-based privacy management**) - Ontology-based privacy management mechanism is effective in reducing data privacy risks and thus consumers feel more secure to share their data using this mechanism.
(5) (**Efficacy of privacy management mechanism**) - Ontology-based privacy management mechanism performs better in reducing data privacy risk than the existing mechanisms.

The first hypothesis assumes that the average user is aware of the privacy risk when they share their personal data with third-party and utility companies. In other words, users know what kind of information is being collected in the smart grid system and what is the privacy risk if this data is obtained by an untrustworthy authority. Given this information, our second hypothesis tests if users are more concerned about their privacy than before. Our third hypothesis tests the assumption that the traditional privacy management mechanisms are not that effective. We have chosen privacy notice based and social media based privacy mechanisms among the existing privacy management mechanisms to test our hypothesis. The last two hypotheses are related to our proposed approach. We hypothesize that our proposed ontology based privacy mechanism is effective and it is better than the existing privacy management mechanisms.

5.1 Experiment Design

We defined four separate user groups, each of which is exposed to a different privacy mechanism scenario:

- Group 1: Exposed to data sharing without any privacy mechanism. For the rest of this paper we will refer to this group as the *Control Group*.
- Group 2: Exposed to data sharing with privacy notice based mechanism implemented. This group of users were asked to agree to a simple privacy notice (the terms of service agreement) informing that their data is collected and being shared. We will refer to this group as the *Notice Group*.
- Group 3: Exposed to data sharing with social media based privacy mechanism implemented. In this approach, users were asked to choose if they want to share their data after informing what data was collected. For the rest of this paper we will refer to this group as the *Social Group*.
- Group 4: Exposed to data sharing with privacy ontology based mechanism implemented. In this mechanism, users were given an in depth look at what data was collected and how to manage it and preserve their data. As the participants can select the attributes they want to share, we will refer to this group as the *Selection Group*.

An online survey was conducted to obtain the user feedback from all the four user groups. Four different versions of GBI implementation was made available in Amazon Web Services(AWS) and each participant was randomly assigned to a group. The participants were asked to share the data with third parties, acting as if the data was their own. A well-defined set of questions was given to the users to record demographic information and user feedback about the assigned privacy mechanism. To avoid framing effect, the questions were designed in such a way that users were not aware about the results we were looking for.

In our study, participants were recruited by MTurk job post and distribution of a Google site web link. A total of 186 people participated in our survey but only 84 of them were verified as valid responses based on task completion and database entry. Table 1 shows the total number of valid participants among all four groups along with the number of participants who chose to share their data with third parties and utility companies.

	Shared data	Not share data	Total
Notice group	15	2	17
Social group	17	7	24
Selection group	17	2	19
Control group	18	6	24
Total	67	17	84

Table 1: Comparison of Data Sharing and Not Sharing

The last step requested the participants to submit their feedback through a questionnaire on the overall understanding of privacy concern and opinion on the different privacy management mechanisms. Each participant was exposed to only one scenario in the study and they were asked to rate the privacy mechanism they used. Additionally, they were asked to quantify their concern about data privacy in the smart grid system using a five-point Likert scale where one indicates that users are least concerned about their data and five indicates that they are extremely concerned.

6 USER STUDY RESULTS

The results from the survey demographic information shows that 75% of the users are undergrad students or with an education higher than that. About 95% of the users claimed that they pay their own utility bills and 83% said that they use the utility company's websites to pay their bills. From our survey, we found that 81% users will prefer to have a privacy mechanism over no privacy mechanism to address their privacy concerns.

6.1 Hypothesis 1: General awareness about data privacy in the smart grid system

This hypothesis was tested by asking the question if they were concerned about their data privacy in smart grids before this user study. They were asked the rate of their concern level in the Likert scale of 5 where 1 indicates that the user is unconcerned and 5 indicates they are extremely concerned.

To have a clear view, we have drawn a normal distribution graph from user responses (Figure 4). From the graph, it is visible that the

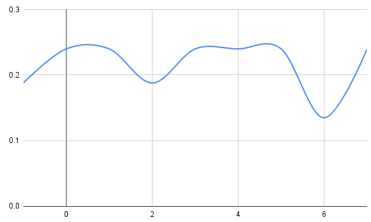


Figure 4: Normal Distribution of Participants Privacy Concern without Additional Knowledge

distribution curve does not follow any normal pattern, it is quite random. From this result, we can conclude that users are generally concerned about their privacy even though they don't have any detailed knowledge about what might happen if they share those data.

6.2 Hypothesis 2: Usefulness of more information to increase the awareness about data privacy

In our second hypothesis, we assumed that if the users are given more information about what data is being collected and how this information can affect their privacy then they will be more concerned. To test this hypothesis, participants were asked to use different versions of privacy mechanisms where they can have a general idea about what data is being collected. Then we asked them if they would be more concerned about their data privacy when more information on the privacy is given. 60.7% of the participants think that more information on data privacy increased their awareness.

Figure 5: Rating Comparison between Prior and Post Additional Knowledge Given

Additionally, we asked the participants to rate their concern level after they performed the user study. Figure 5 shows the histograms of privacy concern before and after given some additional knowledge about data privacy. We performed a paired two sample for means t-test on prior and post knowledge rating data. For a hypothesized mean difference of 0, the p-value is 0.037, 92% confidence interval(CI) confirms our hypothesis as the difference is statistically significant.

6.3 Hypothesis 3: Ineffectiveness of existing privacy management mechanisms

The two most common privacy management methods are privacy notice based mechanism and social media based privacy management mechanism so we selected these two to test our third hypothesis. We hypothesize that these two existing methods are not effective. We compared both groups with control group data where no privacy mechanism was implemented. To compare the groups, we used two-sample t-test assuming unequal variances. For a hypothesized mean difference of 0, the p-value for control and notice group is 0.027 and the p-value for control and social group is 0.38. From the p-values, we can confirm that in 92% CI privacy notice based mechanism is more efficient than no privacy mechanism whereas there is no significant difference between the social media based mechanism and no privacy mechanism.

6.4 Hypothesis 4: Effectiveness of ontology-based privacy management

In our ontology based privacy management mechanism, users are given an in depth look at what data was collected and how to manage it and preserve their data. They are able to select the level of privacy for any particular attribute. We used two-sample t-test assuming unequal variances to find out the effectiveness of the ontology based mechanism. For a hypothesized mean difference of 0, the p-value for control and selection group is 0.018. 92% CI confirms our hypothesis as the difference between these two groups is statistically significant.

6.5 Hypothesis 5: Efficacy of privacy management mechanism

In this hypothesis we assumed that the ontology based privacy mechanism is more effective than privacy notice and social media based privacy management mechanisms in reducing data privacy risks in utility company customer data. We tested this by asking the users to rate the privacy management. Among the users 1 user from each group chose not to rate the privacy mechanism they used. Figure 6 graphically represents the ratings given by the participants.

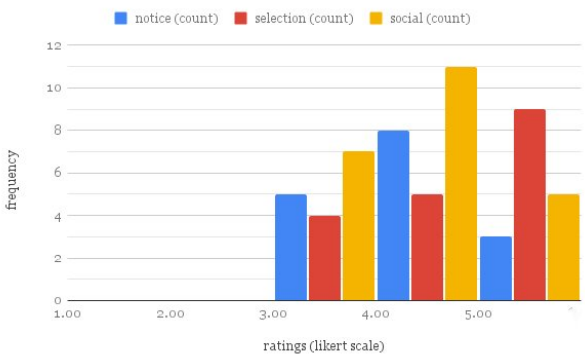


Figure 6: Privacy Mechanisms Rating Comparison

We used two-sample t-test assuming unequal variances to find out the efficacy of the ontology based mechanism. For a hypothesized mean difference of 0, the p-value for notice and selection group is 0.069. In similar setup, the p-value for social and selection group is 0.075. 92% CI confirms our hypothesis as the difference between these groups is statistically significant.

7 CONCLUSION

In this paper, we presented a user privacy management framework for SG systems. Specifically, we discussed a data modeling approach to designing a scalable ontology, along with a user controllable scheme to allow consumers to manage their privacy attributes and selectively share their attributes based on their sensitivity and privacy risks. Additionally, we presented a proof-of-concept implementation to test the feasibility of the proposed approach using the GBI framework. Lastly, we conducted a user study to test the efficacy of the proposed approach.

ACKNOWLEDGEMENT

This work was partially supported by the grant from the National Science Foundation (NSF-OIA-1757207). We also thank Md Umar Sharif, Manjusha Raavi, Jiwan Chong, and Branndon Montano for their invaluable contributions to this work.

REFERENCES

[1] A. Cooper and M. Shuster. Electric company smart meter deployments: foundation for a smart grid (2021 update), 2021.
[2] C. Wueest. Targeted attacks against the energy sector, 2014.
[3] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin. Private memoirs of a smart meter. In *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, BuildSys '10. Association for Computing Machinery, 2010.
[4] Y. Liu, W. Guo, C. Fan, L. Chang, and C. Cheng. A practical privacy-preserving data aggregation (3pda) scheme for smart grid. *IEEE Transactions on Industrial Informatics*, 15, 2019.
[5] J. Marks, B. Montano, J. Chong, M. Raavi, R. Islam, T. Cerný, and D. Shin. Differential privacy applied to smart meters: a mapping study. In C. Hung, J. Hong, A. Bechini, and E. Song, editors, *SAC '21: The 36th ACM/SIGAPP Symposium on Applied Computing, Virtual Event, Republic of Korea, March 22-26, 2021*. ACM, 2021.
[6] H.-Y. Lin, S.-T. Shen, and B.-S. P. Lin. A privacy preserving smart metering system supporting multiple time granularities. In *2012 IEEE Sixth International Conference on Software Security and Reliability Companion*, 2012.
[7] S. Speiser, A. Wagner, O. Raabe, and A. Harth. Web technologies and privacy policies for the smart grid. *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*, 2013.
[8] H. Simo Fhom, N. Kuntze, C. Rudolph, M. Cupelli, J. Liu, and A. Monti. A user-centric privacy manager for future energy systems. In *2010 International Conference on Power System Technology*, 2010.
[9] L. Zhang, J. Ren, Y. Mu, and B. Wang. Privacy-preserving multi-authority attribute-based data sharing framework for smart grid. *IEEE Access*, 8, 2020.
[10] F. Loukil, C. Ghedira-Guegan, K. Boukadi, and A. N. Benharkat. Liopy: a legal compliant ontology to preserve privacy for the internet of things. In *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, volume 02, 2018.
[11] Q. Zhou, S. Natarajan, Y. Simmhan, and V. Prasanna. Semantic information modeling for emerging applications in smart grid. In *2012 Ninth International Conference on Information Technology - New Generations*, 2012.
[12] S. Gillani, F. Laforest, and G. Picard. A generic ontology for prosumer-oriented smart grid. In *EDBT/ICDT Workshops*, 2014.
[13] Y. Simmhan, Q. Zhou, and V. Prasanna. Semantic information integration for smart grid applications. In Springer, 2011.
[14] R. Santodomingo, S. Rohjans, M. Uslar, J. Rodríguez-Mondéjar, and M. Sanz-Bobi. Ontology matching system for future energy smart grids. *Engineering Applications of Artificial Intelligence*, 32, 2014.
[15] G. Santos, F. Silva, B. Teixeira, Z. Vale, and T. Pinto. Power systems simulation using ontologies to enable the interoperability of multi-agent systems. In *2018 Power Systems Computation Conference (PSCC)*, 2018.
[16] O. Sacco and R. Passant. A privacy preference ontology (ppo) for linked data. In *in Linked Data on the Web Workshop at the World Wide Web Conference*, 2011.
[17] C. Ke, F. Xiao, Z. Huang, Y. Meng, and Y. Cao. Ontology-based privacy data chain disclosure discovery method for big data. *IEEE Transactions on Services Computing*, 2019.
[18] L. Yang, H. Xue, and F. Li. Privacy-preserving data sharing in smart grid systems. In *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2014.
[19] D. Mashima and A. Roy. Privacy preserving disclosure of authenticated energy usage data. In *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2014.
[20] Y. Ye, L. Zhang, W. You, and Y. Mu. Secure decentralized access control policy for data sharing in smart grid. In *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2021.
[21] K. Gai, Y. Wu, L. Zhu, L. Xu, and Y. Zhang. Permissioned blockchain and edge computing empowered privacy-preserving smart grid networks. *IEEE Internet of Things Journal*, 6(5), 2019.
[22] Z. Guan, X. Zhou, P. Liu, L. Wu, and W. Yang. A blockchain based dual side privacy preserving multi party computation scheme for edge enabled smart grid. *IEEE Internet of Things Journal*, 2021.
[23] D. Shin, R. Lopes, W. Claycomb, and G.-J. Ahn. A framework for enabling user-controlled persona in online social networks. In *2009 33rd Annual IEEE International Computer Software and Applications Conference*, volume 1, 2009.