IEEE Access
The journal for rapid open access publishing

Received October 19, 2015, accepted November 10, 2015, date of publication December 1, 2015, date of current version December 29, 2015.

Digital Object Identifier 10.1109/ACCESS.2015.2504503

Smart Meters Big Data: Game Theoretic Model for Fair Data Sharing in Deregulated Smart Grids

ABDULSALAM YASSINE[1], (Member, IEEE), ALI ASGHAR NAZARI SHIREHJINI[2], AND SHERVIN SHIRMOHAMMADI[1], (Senior Member, IEEE)

[1]School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON K1N 6N5, Canada
[2]Sharif University of Technology, Tehran 11365-11155, Iran

Corresponding author: A. Yassine (abdulsalam.yassine@gmail.com)

ABSTRACT Aggregating fine-granular data measurements from smart meters presents an opportunity for utility companies to learn about consumers' power consumption patterns. Several research studies have shown that power consumption patterns can reveal a range of information about consumers, such as how many people are in the home, the types of appliances they use, their eating and sleeping routines, and even the TV programs they watch. As we move toward liberalized energy markets, many different parties are interested in gaining access to such data, which has enormous economical, societal, and environmental benefits. However, the main concern is that many such beneficial uses of smart meter big data would be severely curtailed if the data were excessively protected due to individuals' privacy. In this paper, we propose a game theoretic mechanism that balances between beneficial uses of data and individuals' privacy in deregulated smart grids. Our mechanism solves the problem of access control by fairly compensating consumers for their participation in the data market based on the concept of differential privacy. The results of our experiments show the importance of taking consumers' attitudes toward privacy as a crucial element in designing balanced markets for fair data sharing. Furthermore, the experiments provide a principled way to choose reasonable values for privacy levels that are more relevant to real-world scenarios.

INDEX TERMS Smart metering, smart grid, big data, privacy, game theory.

I. INTRODUCTION

Reading fine-granular data measurements from smart meters every 30 minutes transforms into 48 million readings for every one million consumers resulting in a massive volume of data [1]; i.e smart meters big data. Studying these big data not only provide utility companies with consumption data, but also allow them to derive a great deal of valuable details about how consumers use energy. To protect consumers' privacy, current work, such as [2]–[4] and [5], focus on means to protect household's power consumption information by various security and privacy mechanisms. As we move towards liberalized energy markets, many different parties are interested in gaining access to consumers' information (e.g. preferences and behavior), derived from power consumption data. This data has enormous economical, societal, and environmental benefit [6], [7]. However, the main concern is that many beneficial uses of smart meters big data would be severely curtailed if the data is excessively protected to individuals' privacy. Some of these benefits accrue directly to consumers, who are able to save on their energy consumption by understanding which devices and appliances consume the most energy, or which times of the day they can schedule their demand to reduce their energy bill [8]. Other benefits, such as accurately forecasting energy demands to optimize the use of green energy sources, are gained by the whole society [8]. Individuals recognize that many complex challenges with societal importance, such as environmental considerations, market-research or political decision-making [9], may benefit from accessing smart meters big data. Admittedly, however, they also recognize that their privacy may be jeopardized as several research studies show that personal information such as an individual's sleep-awake cycles, activities, preferences, TV programs, and multimedia content can be estimated with high accuracy from consumers power consumption patterns [10]–[16].

In this paper, we address the following issues: First, how to design mechanisms that balance between beneficial uses of data and individuals' privacy in deregulated smart grids. On one hand, concerns over privacy issues may limit the access to valuable information, which can dampen the

data economy, innovation and productivity [8]. On the other hand, it is particularly unacceptable to have the privacy door widely open while at the same time individuals are deprived of their opportunities to make decisions about their data, and whether they want to be involved in certain data markets [17]. After all, consumers' data has great commercial value in use as well as in exchange. The World Economic Forum (WEF) referred to consumers' data as the ''New Oil'' of the 21st Century [6]. Second, how to control access to data and guarantee fairness for all players in deregulated smart grid markets? Currently, several companies e.g. C3 Energy, Energy Savvy, K Wantera, WegoWise etc., have emerged and begun marketing their platforms for smart meters big data analysis. It is important that we design mechanisms that accurately reflect individuals' incentives to participate in smart meters big data markets [18].

The above challenging questions have served as motivation behind our work. We envision a market for smart meters big data, where data analysts are interested in obtaining access to a certain subset of the data that corresponds to a representative subset of individuals. The data analysts work with third party service providers who want to learn about consumers behavior and preferences to offer them some services. Smart meters big data is collected by the data aggregator (e.g. utility company) who acts as the market maker. The data aggregator engages in negotiation with the data analysts to strike a deal that allow them to access the consumers' data while considering the trade-off between privacy, the quality of data analysis results, and consumers willingness-to-participate in the data market.

Our approach of tackling the technical challenges is as follows: First, we use the concept of Activities in Daily Living (ADL) [19] to produce ontology of classes of activities from power consumption big data. This will allow the data aggregator and the consumer to identify which type of private information may potentially be derived from the activity. and which is the potential privacy risk depending on the sensitivity level of the private information. This modular ontology of classifying power consumption data will be used to define a privacy risk matrix, taking into account the value of data usage against the potential risks to privacy. We also propose a mechanism that aims at determining the privacy risk value which will be used to determine the user's payoff once the user decides to reveal the data to third party service providers. Second, we propose a game theoretic negotiation mechanism to investigate fairness among the consumers, the data aggregator and the third parties. The goal of each player involved in the game is to maximize the utility. The consumer wants to maximize his reward from allowing access to power consumption data, the data aggregator generally prefers to receive more money from third parties and provide less incentives to consumers, and finally third parties prefer to pay less for data and get higher quality and higher cardinality. To our knowledge, no other work has tried to assess mechanisms of sharing power consumption data among market players in deregulated smart grids, or how, in general, to engage energy consumers in such markets to commercially benefit from the amount and nature of extractable information from smart meters big data.

The rest of the paper is organized as follows: Next section presents the related work followed by the model description in section III. In section IV, we provide analysis of the negotiation game and the players utilities. Section V presents the experimental evaluation. Finally, in section VI we conclude the paper and provide our plan for future work.

II. RELATED WORKS

Smart meters big data has great societal importance that can be achieved by mechanisms that balance between consumers privacy and the benefit of sharing data. Similar mechanisms are currently being used for personal information markets in the online world (e.g. [20]–[24]). These markets have evolved such that users make choices in which they surrender a certain degree of privacy in exchange for benefits such as price discounts, money, improved qualities of service, customized offers, specials, etc. These benefits are perceived by the user to be worth the risk of information disclosure. The design of these mechanisms involves privacy, negotiation mechanisms, game theoretic and data market models. In this section, we review some of the existing studies, the technical approaches, and the research work on data sharing. Although some of these studies are not directly used in the context of smart meters big data, they are fairly applicable to the general concept of data sharing.

Research on smart meters big data is growing due to the valuable information that can be extracted from the collected datasets [25], [26]. For example, the work in [25] discussed the possibilities of data forensic on smart meters big data for detecting attacks implemented by compromised appliances at residential premises. Results of forensic analysis can then be shared with law enforcement agencies. The main issue is that privacy of individuals may be breached if there is no clear consent from the consumer. In this regard, many existing mechanisms can be leveraged to balance between the privacy trade-off and the usage of power consumption data. For example, third parties can purchase access to users' information [28], [29] by compensating them for the release of their data. The main idea of these approaches is to integrate transactional privacy in a privacy preserving systems and forming a market of personal information that can be managed by a trusted third party. The work in [27] argues that over protection of data limits the operational capabilities of the smart grid system. The study proposes methods to balance between privacy requirements and operational requirements in a smart grid system.

Providing incentives for consumers to share their data has been studied in [30]–[32]. In [30], the authors consider the case when a data analyst wishes to buy information from individuals in order to derive some statistical estimates. The objective of the data analyst is to minimize his costs while having an accurate estimate. The data owners experience some cost for their loss of privacy, and therefore must be

compensated for this loss. This is confirmed by [31] which also that consumers are concerned about their privacy, and experience some cost as a function of their privacy loss. Consumers are willing to participate in a survey if they receive an incentive (e.g. payment) at least equivalent to their privacy cost. The cost valuation of the privacy loss is assumed to be private. However, individuals are assumed to be rational, but in reality they may lie about their privacy cost valuation if it leads to best outcome. The work in [32] proposes mechanisms for modeling privacy in players' utility functions following the work presented by [33]. The proposed work in [32] measures privacy cost as mutual information between a player's type and the outcome of the mechanism. The difficulty is that it can only incentivize truthfulness by giving players an influence on the outcome, but such an influence also leads to privacy costs, which may incentivize lying.

In the context of data sharing and control, [34] presented a MultiParty Access Control (MPAC) model to identify and resolve privacy conflicts for collaborative data sharing. Parties involved in accessing consumers' data aim to maximize their own benefit. The model is formulated as a multiparty control game and show the existence of unique Nash Equilibrium (NE). Similarly, [35] studied the problem of sharing data in digital repositories. The authors addressed the issues of participation of consumers in data analytical projects using a game-theoretic model in which individuals take control over participation in data analytics. The main idea is that individuals can contribute data at a self-chosen level of precision and they can decide whether they want to contribute or not. The study also investigated the options of the analyst to set requirements for data precision, so that individuals are still willing to contribute to the project, and the quality of the estimation improves. Another work [38] has also considered data markets where data analysts can access unbiased samples of private data by appropriately compensating the individuals to whom the data corresponds according to their privacy attitudes. Similar to [38] and [36] studied the problem of a data analyst who may purchase an unbiased estimate of some statistic from multiple data providers. The analyst is provided with a menu of options with estimate variance levels, where each level has a cost associated with it. These estimates are reported by the data owner. The study solves an optimization problem by combining the purchased estimators into an aggregate estimator that has a variance of at most equal to some fixed desired level. In such a setting, the analysts are willing to compensate the data providers in order to truthfully report their costs to the mechanism.

Our work falls under the same auspice of the above mentioned studies. These studies assume that for a certain mechanism (negotiation, mechanism design, game theoretic), data subjects individually decide on the degree of data accuracy given a trade-off between their privacy and the revelation of their data. Our negotiation game is similar to the work presented in [37]. As in [37], we believe that it is not feasible for an individual consumer to negotiate directly with data analysts. The data aggregator is the entity that collects information and has the means to influence the level of precision of the released data, i.e., adding noise according to the agreed upon privacy level. By so doing, the consumer has prior knowledge about the privacy level and can opt in or opt out according to the estimation of the privacy loss cost.

III. MODEL DESCRIPTION

In this section, we provide an overview of the deregulated smart grid market model. We describe the role of the involved players in this market. We also discuss our proposed model of data classification using Activities in Daily Living (ADL).

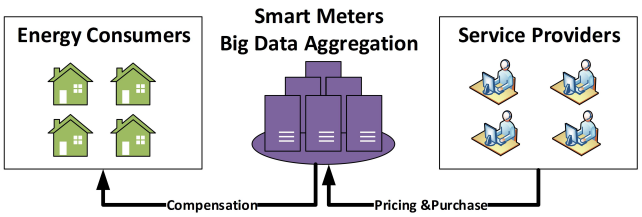

FIGURE 1. Market players in deregulated smart grids.

A. MARKET MODEL

In deregulated smart grids, data aggregators harvest terabytes of fine granular energy consumption measurements from smart meters installed at households. Third party service providers are interested to obtain a certain subset of the data. We envision that a data aggregator, acting on behalf of energy consumers, involves in the negotiation with third party service providers (represented by data analysts) as shown in figure 1. There are many services that can be offered from third parties by having access to power consumption profiles of households [12]–[14], [16], [25]. For example, a company that offers targeted energy-saving products may want to sell its services to consumers with high-energy consumption. The offered service could be a tool that enables consumers to analyze their energy usage and create a personalized energy-saving plan. By getting access to energy consumption patterns of consumers, such companies can build compelling campaigns that quickly and easily deliver personalized services to customers. Furthermore, advertisement companies want to know what type of equipment the house is using, so they can target their advertisement to maximize consumers' attention to their products. From energy consumption data they can identify the type of equipment in the house [14] and its usage frequency, as well as its brand [16]. This information helps advertisement companies to move from large, broad-based, inefficient marketing programs to smaller, more agile, and higher-yield campaigns. Other examples of offered services may include dynamic billing, insurance deals, health care premiums etc. Such services generate lucrative amounts of money, which would be ''unfair'' for the rightful owner of the data not to receive any.

The aggregator collects energy consumption data and works on behalf of consumers. The aggregator uses the

ADL concept to identify which type of private information may potentially be derived from the activity and the potential privacy risk depending on the sensitivity level of the private information. This step is critical because it helps the consumer to identify the privacy risk level based on data subsets, which are included in the analysts query. The aggregator receives offers from data analysts, and based on their needs and the expected cardinality of the dataset announces a privacy level and some rewards to individual consumers. Then, the consumer decides to opt-in or opt-out of the deal. Once an agreement is concluded the aggregator protects the data of the consumers at the agreed level of privacy and provides the data analyst access to the data. In this model, we assume that the consumers are aware of the fact that their private information and preferences can be derived from power consumption data. We also assume that consumers are informed of the data market and they are willing to share their information for a certain benefit that is proportional to loss of privacy. Their cost valuation of their privacy loss can be as diverse as their privacy concerns. There are many reasons that justify their concern, for example, some individuals may perceive the release and use of their data as an intrusion on their personal sphere, or as a violation of their right. In addition, they may fear that the data is used for social and economical discrimination.

Next, we illustrate the ADL concept, which is used to produce ontology of classes of activities from power consumption big data.

B. DATA CLASSIFICATION MODEL

The concept of Activities in Daily Living (ADL) is used to classify activities and tasks mainly inside a house. The concept of ADL is widely used to measure the ability of people to maintain their regular life in the field of medical care for high risk patients and elderly. Among the traditional assessment of everyday functions, ADL includes basic ADL (BADL), instrumental ADL (IADL), extended ADL (EADL), and advanced ADL (AADL). The BADL includes basic activities for independent living functions, such as bathing, using the toilet, and eating. The IADL includes activities with instruments, such as using a telephone, food preparation, cleaning the home, and doing the laundry. EADL and AADL are references for measuring the activities of independent living. To capture the essence of modern activities, [19] added complex ADL (CADL) and quality ADL (QADL) which are used as a reference to match the operation of appliances inside the house. In our model, we use ADL to classify smart meters data and categorize power consumption patterns, which represent the operation of appliances inside the house. The operation of appliances are in fact directly related to users' tasks that can be used to infer preference and behavior.

The ability of measuring the power consumption of individual appliances throughout the course of each day can provide a detailed portrait of an individual ADL at home. In our study, the data aggregator uses the ADL concept as a classification reference to identify which type of private information

FIGURE 2. An example of ADL classification.

may potentially be derived from the activity and the potential privacy risk associated with the combination of private data. Figure 2 shows an example of ADL classification. This modular ontology of classifying power consumption data will be used to define a privacy risk matrix, taking into account the value of different uses of data against the potential risks to individual privacy. For example, a data analyst for a health insurance company wants to learn about unusual behaviors that might indicate illness of an individual before negotiating a health care premium. The insurance company might want to know if the individual cooks at home, which is an indication of healthy eating and if this individual exercise regularly or watches TV for extended hours. Combining these information together gives the analyst a possibility matrix about the health risk of this particular individual. By analyzing the power consumption pattern of cooking appliances (e.g. stove, oven, microwave), entertainment appliances (e.g. TV), and exercising equipment (e.g. treadmill, stationary bike), the analyst will be able to draw a good picture about the potential health risk of an individual. If the individual does not cook at home and watches TV for extended period of times, meaning higher probability of eating unhealthy and not exercising regularly, then this individual is at risk of developing illness. When aggregating data from different subsets of ADL classification, the data analyst will have better insight about the individual behavior. For the consumer, it means higher risk of exposure to privacy loss.

The data aggregator classifies the data using ADL as follows: Consider we have M different categories $(C_1, C_2, \ldots, C_M)$, such as BADL, IADL, EADL, AADL etc. In every category C_j, data is further divided into subsets $S_{j,k}$ for $k = 1, 2, \ldots$ based on the data ontology that applies to the dataset. For example, in figure 2, IADL category is divided into two subsets, Laundry and Cleaning. The aggregator

takes into consideration the following: first, data in different categories may have different privacy risk level. Therefore, the composition of the data may have different implications on the level of privacy. Second, the substitution of different private data in the same subset is constant; i.e., assuming that one of the data has been revealed; revealing the rest of the data in the same subset will only increase the privacy risk by a small margin. For instance, in the Laundry category revealing consumption data about the washing machine and dryer at the same time has little implication on the consumers' privacy. The data analyst can improve his knowledge of consumer's Laundry behavior by a small fraction. Third, data in different subsets are not substitutable, revealing any one of them will increase the privacy risk.

Let $\mathcal{Q}_j$ be the cardinality of C_j (i.e. $\mathcal{Q}_j = |C_j|$). We perform normalization over the whole data set. The normalized data size $\overline{\mathcal{Q}}_j$ reflects the ratio of each category in a consumer's privacy.

$$\overline{\mathcal{Q}}_j = \frac{\mathcal{Q}_j}{\sum_{k=1}^{M} \mathcal{Q}_k} \tag{1}$$

where M is the number of the data categories.

Every consumer i of type t has a privacy risk value $PR_{i,t}$ (privacy risk and consumer types are discussed in subsection IV.B) for private data in each category, the aggregator calculates the weighted privacy risk, as in (2), and then normalizes it, as in (3). The value in (3) reflects the risk of revealing all data in category j.

$$\Omega_j = \overline{\mathcal{Q}}_j . PR_{i,t} \tag{2}$$

$$\overline{\Omega}_j = \frac{\Omega_j}{\sum_{g=1}^{j} \Omega_g} \tag{3}$$

The privacy risk revealing α_j subsets from category j is calculated as in (4).

$$\Omega_j = \frac{\alpha_j}{\mathcal{Q}_j} . \overline{\Omega}_j \tag{4}$$

The privacy risk value in (4) will be used to determine the compensation value that the consumer should receive given the privacy risks that are involved. Intuitively, we want to associate high benefit/compensation with Ω_j that allow for greater revelation of consumers private data. In this manner, the compensation paid to the consumer is justified, at least in part, from the damages that might occur to his privacy.

IV. NEGOTIATION GAME AND PLAYERS UTILITY

In this section, we introduce the details of the negotiation game and the players' utility. Specifically, we discuss the negotiation offers and the responses from the data aggregator to balance the negotiation framework.

A. NEGOTIATION GAME

The negotiation model in our study is adopted from [37]. The negotiation between the data aggregator and the data analyst is modeled as a game with perfect and complete information. This means, the players have full knowledge of the strategic moves and payoff functions of each other. More specifically, both players know the consumer behavior model, but not necessarily the privacy/reward trade-off functions of each consumer because an individual consumer is not directly modeled as a player in the negotiation. The outcome of the negotiation game is affected by the chosen privacy level. The data aggregator employs anonymization techniques such as [39]–[41], which provide data privacy at the cost of losing some information. All these techniques have one common privacy parameter denoted by ρ, which indicates the level of privacy protection. In general ρ affects the valuation cost of the data loss in the sense that the higher ρ, the lower the cost of privacy loss.

The game starts with an offer from the data analyst to the data aggregator. In this framework, the data analyst wishes to propose a privacy level that corresponds to some data precision requirements, so that individuals are still willing to contribute to the dataset, and the cost for accessing the dataset is minimized. The analyst proposes a privacy parameter (precision of the anonymization technique) for which he is willing to purchase access to the data. In the offer, the required value for privacy parameter ρ and the price γ (per each record) must be specified. We denote an offer by $offer = \langle \rho, \gamma \rangle$. Once the data aggregator receives the offer he can either reject or accept it. In case of a rejection, the game terminates with payoff zero to both the analyst and the data aggregator. If the data aggregator decides to accept then he needs to announce an incentive in exchange for allowing the data analyst to access consumers' data. Here, we assume that r represents monetary value of the incentive and its domain is in $\mathbb{R}^+$. The outcomes of the game are either an accepted offer $\langle Offer, Accept \rangle$ or a rejected offer $\langle Offer, Reject \rangle$. The number of consumers who will opt-in is determined by each consumer utility function (discussed in VI.B). Consequently, preferences of the data analyst and the data aggregator are determined based the number of records in the dataset and values ρ, γ, and r.

The aggregator is trusted by both the data analyst and data owners. He collects data from the owners and sells it in the form of queries or data access. When a data analyst decides to purchase a query, the aggregator collects payment, computes the answer to the query, adds noise (anonymization) as appropriate, returns the result to the analyst, and finally distributes individual payments to the data owners. The aggregator retains a fraction of the price as profit at least to cover basic cost of data maintenance. The interaction between the parties is captured in figure 3. In order for the framework to be balanced [42], each consumer must be fairly rewarded whenever the answer to some query results in some privacy loss, and the data analyst is charged to cover all these payments. This definition involves three quantities: the payment that the data analyst needs to pay the market maker, a measure of the privacy loss of data item, and a payment by which the market maker compensates the data owner for this privacy loss. In the following subsection, we provide the details of these quantities when we discuss the utility of each player.

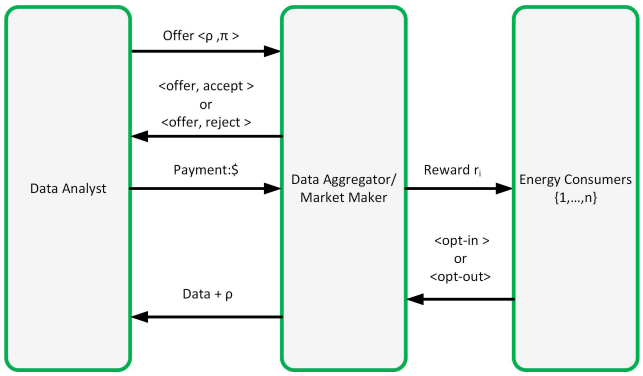

FIGURE 3. Negotiation among data analyst and data aggregator.

B. ENERGY CONSUMER UTILITY

As discussed above, individuals' decision to provide or deny access to their power consumption data depends on the offered level of privacy and the offered reward. We suppose that each consumer $i \in \mathcal{N}$, such that $\mathcal{N} = \{1, \ldots, n\}$, has a privacy risk valuation depending on his type $t \in \mathcal{T}$, where $\mathcal{T}$ is the set of consumer types. The type of the consumer is determined by the consumer's preference to protect his private information. In real life, individuals are categorized into three types [20] marginally not concerned individuals who are willing to accept lower offers to allow others to access their information, pragmatic individuals who expect offers that are worth the risk of privacy exposure, and privacy fundamentalists who are strictly conscious about privacy exposure; but willing to allow access to their personal information with high assurance provisions and payoff. The privacy preferences are modeled by $\lambda_{i,t} \in [\underline{\lambda}, \overline{\lambda}]$. For instance, $\underline{\lambda} = 0$ models a consumer who is willing to surrender his privacy in exchange for a deal. On the other hand, $\overline{\lambda} = 1$ models a consumer who is fundamentalist and in favor of maximizing his privacy. Values of $\lambda_{i,t}$ between $\underline{\lambda}$ and $\overline{\lambda}$ represent various degrees of pragmatism with respect to privacy preferences.

The privacy risk $PR_{i,t}(\alpha)$ of an individual consumer i of type t over a data item $\alpha \in \mathcal{A}_i$, where $\mathcal{A}_i$ is the set of private data, is given as follows:

$$PR_{i,t}(\alpha) = PC(\alpha).SL(\alpha) \quad \forall \alpha \in \mathcal{A}_i, \forall i \in \mathcal{N}, \forall t \in \mathcal{T} \tag{5}$$

where $PC(\alpha)$ denotes the general privacy concern of a consumer i ($0 \leq PC(\alpha) \leq 1$), and $SL(\alpha)$ denotes the sensitivity level of private data ($0 \leq SL(\alpha) \leq 1$). Equation (5) explicitly represents the privacy attitude of the consumer for a specific piece of data. The privacy risk $PR_{i,t}(\alpha)$ in this regard does not take into consideration the combination of multiple data items as discussed in Subsection (III.B). The data aggregator derives the over all privacy risk value Ω_j from $PR_{i,t}(\alpha)$. Each consumer is described by a privacy risk value as well as a nondecreasing cost functions $v_{i,t}(\alpha) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ that measures the consumers dis-utility for having his private data revealed. Therefore, there is a mapping l between the privacy risk and the cost of privacy loss such that $l = [\forall \alpha \in \mathcal{A}_i : \Omega_j(\alpha) \rightarrow v_{i,t}(\alpha)]$. The cost valuation of privacy loss $v_{i,t}(\alpha)$ is private to the consumer and represents the ''pure cost'' of revealing his private data without any anonymization. The data aggregator who is working on behalf of the consumers considers only those who are interested in both privacy and reward. Consumers weigh their decision based on the following model:

$$\mathcal{N}' \subseteq \mathcal{N} = af_1(\rho, v_{i,t}) + bf_2(r) \tag{6}$$

where $\mathcal{N}'$ is the number of individuals who accept the offer, f_1 and f_2 are functions of ρ, $v_{i,t}$ and r, and a and b are the intercept and marginal effects of $f_1(\rho, v_{i,t})$ and $f_2(r)$ on individual's decision to participate in the market. The functions $f_1(\rho, v_{i,t})$ and $f_2(r)$ are two non-decreasing functions. In equation (6), we assume that consumers cannot increase their privacy cost valuation by lying about their true privacy concerns. Hence, individual rationality is implied. Furthermore, we assume that privacy costs have known bounded support, that is, $v_{i,t} \in [v_{min}, v_{max}]$.

C. DATA AGGREGATOR UTILITY

In our model, the data aggregator develops and maintains the ontology of classes of activities which is produced from the ADL. This will allow the data aggregator and the consumer to identify which type of information may potentially be derived from the activity and the potential privacy risk. Once the data is categorized, the data aggregator receives offers from the data analyst, and based on their needs and the expected cardinality of the collected dataset provides reward and price for both the consumers and the data analyst respectively. The data aggregator, while being truthful, aims at maximizing his benefits. The data aggregator receives a payment γ for each data record γ. For $\mathcal{N}$ consumers' records, the aggregator's revenue, $\mathcal{R}$, is as follows:

$$\mathcal{R} = \gamma.\mathcal{N} \tag{7}$$

The data aggregator incurs costs related to data categorization, maintenance, storage, energy etc. we denote these costs by β. In addition, the data aggregator must compensate consumers with payment r. The total cost to the data aggregator can be defined as:

$$\mathcal{TC} = \beta + r.\mathcal{N} \tag{8}$$

The payoff to the data aggregator is therefore defined as [37]:

$$\mathbf{U}_{ag} = \mathcal{R} - \mathcal{TC} = (\gamma - r).\mathcal{N} - \beta \tag{9}$$

D. DATA ANALYST UTILITY

A data analyst is an entity interested in accessing power consumption information for some data analysis purposes. The data analyst is equipped with a mechanism that offers a menu specifying a discrete, finite range of possible levels of anonymization $0 < \rho_1 < \rho_2 < \ldots < \rho_m < \infty$ which are associated with prices $(\gamma_i)_{i=1}^m$. The offer made by the data analyst specifies the amount of payment that the analyst is willing to pay for a certain privacy parameter. Adding noise reduces

the price. The data analyst prefers a dataset with higher quality and a wider range of data records. The added noise to the data set after anonymization will have impact on the quality of the data as well as the number of consumers opting in. If the consumers are privacy fundamentalists, this will lead to higher anonymization, which in tern means higher cost of accessing the data. The data analyst starts by submitting a request which specifies the query of accessing the data. In this request the data analyst includes the price γ for each data record and the noise level that he is willing to accept for the offered price. Obviously, the data analyst specifies the price that maximize his benefit. He prefers to access a wider range of data items and pay the minimum cost per record. However, the wider the range of the dataset, the higher the privacy risk for the consumer. Let ϑ denote the absolute worth of the data record before any added noise. If the number of data records acquired from the data aggregator is denoted by $\mathcal{N}$, then the analyst's revenue is defined as $\vartheta.\mathcal{N}$. However, after adding noise the benefit of data decreases. Let η; $(0 \leq \eta \leq 1)$, be the parameter that captures the level variance of precision. The revenue $\mathcal{I}_{da}$ of the data analyst is then given as follows [37]:

$$\mathcal{I}_{da} = \vartheta.\mathcal{N}.\eta \tag{10}$$

The main factor that affects the precision function is the level of noise agreed between the data aggregator and the analyst. In this regard, for any number of data records N, η is a decreasing function of ρ. If the analyst pays price γ per record, his cost is $\gamma.N$, and therefore his utility $\mathbf{U}_{da}$, is as follows [37]:

$$\mathbf{U}_{da} = \mathcal{N}(\vartheta.\eta - \gamma) \tag{11}$$

There are several methods to introduce noise to a dataset to minimize privacy loss. One of these methods is known as ϵ-differential privacy [41]. In the next section, we provide the equilibrium strategies for the players when ϵ-differential privacy is used as a means of introducing noise to data records.

V. EQUILIBRIUM STRATEGIES WITH ϵ-DIFFERENTIAL PRIVACY

In this section we provide details about the equilibrium strategies of the negotiation game. Specifically, we discuss the stages needed to reach the game's equilibria. We start by giving a brief summary about the ϵ-differential privacy and the notion of privacy loss.

A. ϵ-DIFFERENTIAL PRIVACY AND PRIVACY LOSS

The main objective of differential privacy is to provide a privacy-preserving model that can be satisfied by a given mechanism of data analysis. The basic idea of differential privacy is captured in the following known definition [41]: "*A randomized function k gives ϵ-differential privacy if for all data set D1 and D2 differing on at most one element, and all $S \leq Range(k)$*"

$$Pr[k(D1) \in S] \leq exp(\epsilon).Pr[k(D2) \in S] \tag{12}$$

where ϵ is a parameter used to define the strength of the privacy. In this sense, ϵ is set by the mechanism depending on the probability of an event happening that affects the leakage of the information. Therefore, any mechanism satisfying this definition addresses concerns that any participant might have about the leakage of his personal information. In our setup, for each data analyst query, the data aggregator defines a random function $\mathcal{H}_q$, such that, for any database instance α, the random variable $\mathcal{H}_q(\alpha)$ has expectation $q(\alpha)$ and a noise value equal to ρ. By answering the query through this mechanism, the data aggregator leaks some information about each data item α, and its owner expects to be compensated appropriately. The mechanism uses the Laplacian distribution as a source of random noise. Laplacian noise is commonly used to obtain differential privacy. A Laplacian distribution with mean 0 and parameter $e > 0$ is denoted by $Lap(e)$. The probability density function of $Lap(e)$ is

$$f(x) = \frac{1}{2e} exp(\frac{-|x|}{e}) \tag{13}$$

The aggregator uses the Laplacian mechanism denoted by $\mathcal{L}$ to add a noise variable drawn from the probability density function $Lap(e)$. The mechanism returns $\mathcal{L}_{q(\alpha)} = q(\alpha) + \rho$, where ρ is noise with distribution $Lap(e)$ and $e = \sqrt{\upsilon/2}$ and υ is the variance of Laplacian distribution. Given such mechanism the privacy loss of each individual is bounded by:

$$\epsilon(\mathcal{L}_{q(\alpha)}) \leq \frac{s_q}{\sqrt{\upsilon/2}} \tag{14}$$

where s_q is the sensitivity of a query q defined in [41].

From equation (14), we can construct the decision for each consumer as a combination of privacy protection level and incentive. This model is explained in equation (6). By substituting the values in equation (6) we get:

$$\mathcal{N} = a.\epsilon(\mathcal{L}_{q(\alpha)}).v_{i,t} + b.r \tag{15}$$

Following the work provided in [35], [36], and [38], we considered in equation (15) a linear relation between the privacy loss ϵ and the expected cost $v_{i,t}$. Furthermore, for the sake of simplicity, we considered the function for the incentive $f_2(r) = r$. Thus, for consumer i of type t to agree to the use of his data, his expected payment should be at least $\epsilon.v_{i,t}$. Here, we assume a linear relation between cost of privacy loss and the ϵ, however, any other functions are possible without loss of generality.

B. DATA AGGREGATOR ACTIONS

The main goal of the aggregator is to take actions that balance the pricing framework. This means that each consumer must be fairly rewarded whenever the answer to some query results in some privacy loss, and the data analyst is charged to cover these rewards. In this sense, aggregator actions are geared towards finding an equilibrium where everybody is satisfied with the outcome of the negotiation game. The data aggregator can estimate the expected number of data records

in the dataset for each anonymization value and the reward based on the combination of equation (9) and (15).

$$\mathbf{U}_{ag} = (\gamma - r)(a.\epsilon(\mathcal{L}_{q(\alpha)}).v_{i,t} + b.r) - \beta \tag{16}$$

The goal of the data aggregator is to find the reward value $\hat{r}$ that maximizes his utility $\mathbf{U}_{ag}$ as follows:

$$\begin{aligned}\hat{r} &= \arg\max_{r}(\mathbf{U}_{ag}) \\ &= \arg\max_{r}[(\gamma - r)(a.\epsilon(\mathcal{L}_{q(\alpha)}).v_{i,t} + b.r) - \beta]\end{aligned} \tag{17}$$

subject to the constraint that $\hat{r} \geq 0$

The data aggregator will have the following actions:

- $\langle Offer, Accept\rangle$ if $max(\mathbf{U}_{ag}) \geq 0$
- $\langle Offer, Reject\rangle$ if $max(\mathbf{U}_{ag}) < 0$
- $\langle Offer, Accept/Reject\rangle$ if $max(\mathbf{U}_{ag}) = 0$

The best response $\mathbf{BR}_{ag}$ strategy for the data aggregator is then as follows:

$$\mathbf{BR}_{ag} = \begin{cases} Reject & if\ (\gamma - \hat{r})(a.\epsilon(\mathcal{L}_{q(\alpha)}).v_{i,t} + b.\hat{r}) - \beta \leq 0 \\ Accept & if\ (\gamma - \hat{r})(a.\epsilon(\mathcal{L}_{q(\alpha)}).v_{i,t} + b.\hat{r}) - \beta \geq 0 \end{cases} \tag{18}$$

Every time the aggregator receives an offer from the data analyst, he must calculate the value of $\hat{r}$ that maximizes his payoff. This means finding a combination of $\hat{r}$ and γ such that $\hat{r} > \gamma$.

C. DATA ANALYST ACTIONS

As mentioned earlier, the data analyst would like to maximize the number of consumers who are willing to participate in the dataset. Therefore, he must find the most profitable action that achieves this goal. In other words, the data analyst must find the best combination of $\epsilon(\mathcal{L}_{q(\alpha)})$ and γ that maximizes the analyst's utility. The main factor that affects the best action of the data analyst is the optimal reward $\hat{r}$ that the consumers are receiving from the data aggregator based on the offer $\langle \epsilon(\mathcal{L}_{q(\alpha)}), \gamma\rangle$. In this case, if $\hat{r}$ is the optimal solution for the combination of $\epsilon(\mathcal{L}_{q(\alpha)})$ and γ, then $\hat{r}$ can be defined as a function of $\epsilon(\mathcal{L}_{q(\alpha)})$ and γ (i.e., $\hat{r} = z(\epsilon(\mathcal{L}_{q(\alpha)}), \gamma)$). Furthermore, if one value of r maximizes $\mathbf{U}_{ag}$, then this value is part of the equilibrium that maximizes $\mathbf{U}_{da}$ and the number of records $\mathcal{N}$ will be determined based on the following:

$$\mathcal{N} = \begin{cases} a.\epsilon(\mathcal{L}_{q(\alpha)}).v_{i,t} + b.r & if\ max(\mathbf{U}_{ag}) \geq 0 \\ 0 & otherwise \end{cases} \tag{19}$$

Equation (19) states that the data analyst will be granted access to the dataset only when the data aggregator accepts the offer, the noise level is $\epsilon(\mathcal{L}_{q(\alpha)})$, and the reward is $\hat{r} = z(\epsilon(\mathcal{L}_{q(\alpha)}), \gamma)$. In case the data aggregator rejects the offer $\langle \epsilon(\mathcal{L}_{q(\alpha)}), \gamma\rangle$, the analyst will not be given access to the dataset. Substituting the function definition of $\hat{r} = z(\epsilon(\mathcal{L}_{q(\alpha)}), \gamma)$ into equation (19), $\mathcal{N}$ becomes a function of $\epsilon(\mathcal{L}_{q(\alpha)})$ and γ as well. Recall that the precision value η defined in subsection IV.D depends on the anonymization technique used by the aggregator and hence it is a function of $\epsilon(\mathcal{L}_{q(\alpha)})$ and γ and denoted as $\eta'(\epsilon(\mathcal{L}_{q(\alpha)}), \gamma)$. In equation (11), after substituting $\mathcal{N}$ and η with $\mathcal{N}'(\epsilon(\mathcal{L}_{q(\alpha)}), \gamma)$ and $\eta'(\epsilon(\mathcal{L}_{q(\alpha)}), \gamma)$, then $\mathbf{U}_{da}$ becomes a function of two variables $\epsilon(\mathcal{L}_{q(\alpha)})$ and γ. The best response for the data analyst $\epsilon(\mathcal{L}_{q(\alpha)})$ and γ that maximize his payoff:

$$max[\mathcal{N}'(\rho, \gamma)(\vartheta.\eta'(\rho, \gamma) - \gamma)] \tag{20}$$

In equation (20), the upper bound for γ is ϑ. Parameter $\epsilon(\mathcal{L}_{q(\alpha)})$ is not necessarily bounded from above. According to [41], the value of ϵ can take any value depending on the situation. When there is a situation of higher possibility of privacy breach, ϵ must be specified to reflect such probability. There are many economical methods to choose ϵ as have been surveyed in [43]. If $\mathbf{U}_{da}$ has an absolute maximum subject to the bounds defined on $\epsilon(\mathcal{L}_{q(\alpha)})$ and γ, the game has equilibria of the forms $((\hat{\epsilon}(\mathcal{L}_{q(\alpha)}), \hat{\gamma}), reject)$ or $((\hat{\epsilon}(\mathcal{L}_{q(\alpha)}), \hat{\gamma}), \hat{r})$. This means the outcome of the game will be as follows: If the data aggregator determines that the offer will lead to potential loss, then the offer will be rejected and the negotiation ends unsuccessfully. When there is at least one combination of $\epsilon(\mathcal{L}_{q(\alpha)})$ and γ for which the data aggregator can make profit, then the negotiation game ends successfully. The data aggregator must balance between the cost of anonymization, the reward amount, the number of consumers who are willing to participate in the data set and the payment received from the data analyst. When this balance is achieved, we are then assured of a game equilibrium.

D. GAME EQUILIBRIUM

Having explained the strategic actions of the players in the negotiation game. We are now ready to discuss the equilibrium of the game. The first step is to determine the optimum reward $\hat{r}$ from equation (16). If the data aggregator accepts the offer $\langle \rho, \gamma\rangle$ with incentive r, his payoff will be:

$$\mathbf{U}_{ag} = (\gamma - r)(a.\epsilon(\mathcal{L}_{q(\alpha)}).v_{i,t} + b.r) - \beta \tag{21}$$

Calculating the derivative of $\mathbf{U}_{ag}$ with respect to r and setting it to zero reveals the maximizing r:

$$\frac{d\mathbf{U}_{ag}}{dr} = \frac{d[(\gamma - r)(a.\epsilon(\mathcal{L}_{q(\alpha)}).v_{i,t} + b.r) - \beta]}{dr} = 0 \tag{22}$$

Solving equation (22) for $\hat{r}$ we get:

$$\hat{r} = \frac{\gamma.b - a.\epsilon(\mathcal{L}_{q(\alpha)}).v_{i,t}}{2b} \tag{23}$$

The second derivative $\frac{d\mathbf{U}_{ag}}{dr} = -2b < 0$, this means the $\hat{r}$ is the local maximum. The restriction here is $r \geq 0$. If $\hat{r} < 0$, the maximizing r will be zero. The lower bound on r leads us to consider two separate cases:

1) If $\gamma.b > a.\epsilon(\mathcal{L}_{q(\alpha)}).v_{i,t}$ then the amount of reward that maximizes $\mathbf{U}_{ag}$ is $\hat{r} = \frac{\gamma.b - a.\epsilon(\mathcal{L}_{q(\alpha)}).v_{i,t}}{2b}$. Substituting for $\hat{r}$ in equation (22) gives us the maximum payoff to the data aggregator:

$$\hat{\mathbf{U}}_{ag} = \frac{b}{4}(\gamma + \frac{a.\epsilon(\mathcal{L}_{q(\alpha)}).v_{i,t}}{b}) - \beta \tag{24}$$

IEEE Access
The journal for rapid open access publishing

The data aggregator will accept the offer if $\hat{\mathbf{U}}_{ag} \geq 0$. This means when the following condition is satisfied:

$$\gamma + \frac{a.\epsilon(\mathcal{L}_{q(\alpha)}).v_{i,t}}{b} \geq \sqrt{\frac{4\beta}{b}} \tag{25}$$

2) If $\gamma.b < a.\epsilon(\mathcal{L}_{q(\alpha)}).v_{i,t}$ then the optimum reward value would be $\hat{r} = 0$ which gives the data aggregator a maximum payoff:

$$\hat{\mathbf{U}}_{ag} = \gamma(a.\epsilon(\mathcal{L}_{q(\alpha)}).v_{i,t}) - \beta \tag{26}$$

The data aggregator will accept the offer if $\hat{\mathbf{U}}_{ag} \geq 0$. This means when the following condition is satisfied:

$$\gamma(a.\epsilon(\mathcal{L}_{q(\alpha)}).v_{i,t}) \geq \beta \tag{27}$$

Substituting the values in conditions 1) and 2) into equation (19), then we can define the number of consumers as follows:

$$\mathcal{N} = \begin{cases} a.\epsilon(\mathcal{L}_{q(\alpha)}).v_{i,t} + b.\gamma & \text{if } \gamma.b \geq a.\epsilon(\mathcal{L}_{q(\alpha)}).v_{i,t} \\ & \text{and } \gamma + \dfrac{a.\epsilon(\mathcal{L}_{q(\alpha)}).v_{i,t}}{b} \geq \sqrt{\dfrac{4\beta}{b}} \\ a.\epsilon(\mathcal{L}_{q(\alpha)}).v_{i,t} & \text{if } \gamma.b < a.\epsilon(\mathcal{L}_{q(\alpha)}).v_{i,t} \\ & \text{and } \gamma(a.\epsilon(\mathcal{L}_{q(\alpha)}).v_{i,t}) \geq \beta \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

Next, we provide the experimental evaluation of the mechanism.

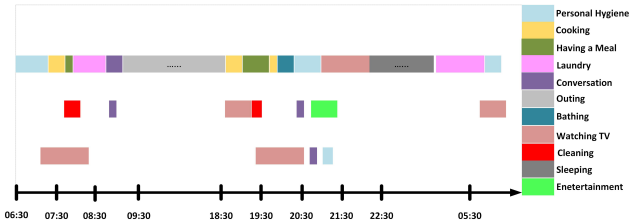

FIGURE 4. ADL example derived from power consumption data.

VI. EXPERIMENTS AND RESULTS

In this section, we present numerical experiments to test the validity of the proposed mechanism. For the experiments, we use a dataset of household power consumption provided by "UMass Trace Repository" [44], which provides daily power consumption records of appliances inside a house. We have used the data to extract the relevant information and categorize the data using the ADL model for several days worth of power consumption. A sample view of power consumption data categorized according to ADL is shown in figure 4. The data is for one house only over many months of power consumption. We built our experiment for 1000 users. For each data category we have assigned a privacy risk value $PR_{i,t}$ to be uniformly distributed between 0 and 1 to reflect the privacy attitude of different types of users. For the privacy cost valuation $v_{i,t}$ we assumed that the data record cost varies between 1 and 5 cents. Each consumer value the offer and opt in or opt out of the dataset after valuating the privacy level and the reward value. In every setting, the experiment examines the effect of the privacy level variation on the overall benefit of both the data analyst and the data aggregator. This is measured by the number of consumers that opts in. The choice of ϵ, which reflects the privacy level, is based on the suggestion provided in [41]. We simulated the value of ϵ to take values 0.1, 0.2, 0.3, 0.4, and 0.5 these values reflect different probabilities of privacy breach events. We also assumed that the payment corresponding to these values are 13, 12, 11, 10, 9, and 8 cents respectively.

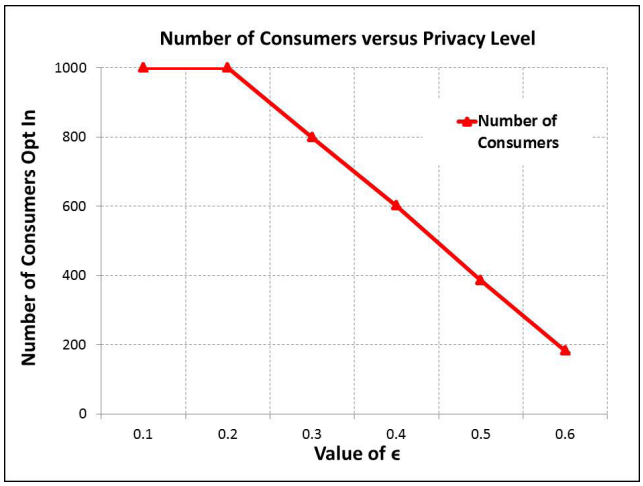

FIGURE 5. Number of consumers versus the privacy level.

In every run of the experiment we aim at testing the behavior of consumers, i.e. opting out or opting in, measured by the number of records, then the profitability of the aggregator given the number of records. We assumed that each consumer has one record and the outcome of the game would be the inclusion or exclusion of this record in the dataset. We first start by fixing the values of parameters a and b in equation (6). Specifically, we give more weight to the privacy cost over the reward for each value of ϵ. We assumed that the cost incurred by the aggregator is fixed and equal to 10 cents per record, which is in line with real scenarios. In figure 5, we show the number of consumers records versus the chosen value of ϵ. As the value of ϵ increases, user become concerned about their privacy and as a result fewer number of consumers would opt in if the value of reward does not change at least for the pragmatic consumers who try to weigh between the reward and the privacy level. Smaller values of ϵ attracts more data users to opt in without severely affecting the precision of the queries. The data user can make more profit in this case as can be seen in figure 6. Furthermore, based on the settings chosen, after a certain point the cost becomes too high for condition of equation (27) to be satisfied. In this case, the data aggregator is receiving a lower payment that limits him to announce non-zero rewards. In such case, many consumers prefer to opt out. At this stage, the data aggregator would not be able to find a combination of ϵ and γ that is acceptable

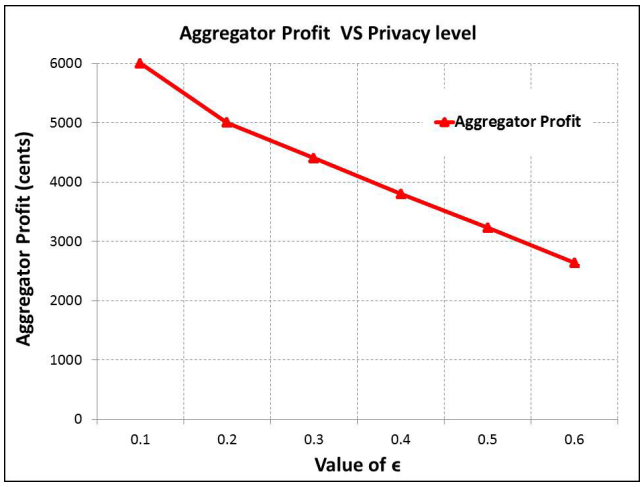

FIGURE 6. Aggregator Profit versus the privacy level.

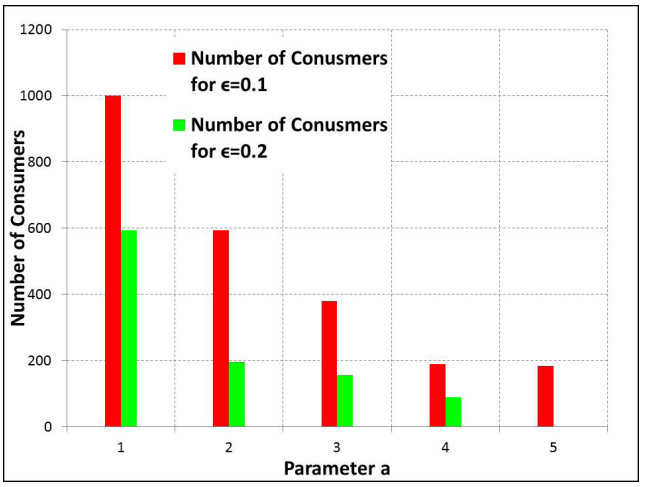

FIGURE 8. Privacy Cost effect on the number of consumers' decision.

to maintain $\mathbf{U}_{ag} \geq 0$. As ϵ increases, the test shows that individual's marginal expected harm increases endlessly. But, this could be unreasonable and there should be a maximum cost for participating. The cost curve could be refined for very small and very large values of ϵ.

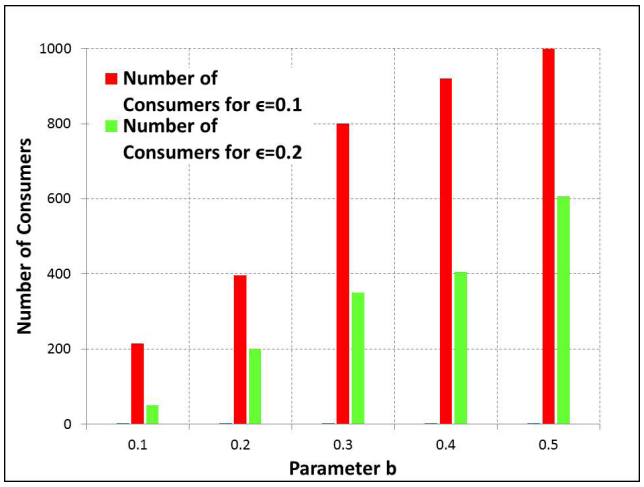

FIGURE 7. Reward effect on the number of consumers' decision.

In figures 7 and 8, we present an experiment where the value of ϵ is fixed for 0.1 and 0.2, but we let the parameters a and b to vary so that they reflect the privacy attitude of a wider spectrum of consumer types. In the first experiment, we fixed the value of a as before, but we let the value of parameter b to take the values $0.1, 0.2, 0.3, 0.4$, and 0.5. The aim of this experiment is to study the effects of consumers' privacy attitude on stable values of ϵ. According to figure 7, as the number of privacy unconcerned group who value the reward more than the privacy increase, the data user can receive larger volume of data without asking for sanitized dataset. This is clear in figure 7 for both values of ϵ. In figure 8, we fixed the value of b and we let a take values 1, 2, 3, 4, and 5. By increasing the value of a we model a privacy aware population. As can be seen in figure 8, when privacy has more significant impact on consumers' decisions, the data analyst must choose appropriate anonymization levels to convince more consumers to participate in the dataset. In both experiments, we show how the attitude of consumers reflected here by the choice of parameters a and b influence the stability values of ϵ. If b is less than a certain level then it mainly influences the price of information and not the level of noise added to the data records. However, when consumers become more interested to receive a larger reward, the data analyst can maximize his utility by just increasing the price and asking for less privacy. These experiments show how consumers attitude towards privacy is crucial in designing balanced markets for fair sharing of data. The experiments also provide a principled way to choose reasonable values for privacy parameter ρ, here as ϵ, and γ based on parameters with more immediate connections to the real world.

VII. CONCLUSION AND FUTURE WORK

In this paper, we presented a mechanism for sharing power consumption data in deregulated smart grids. We presented the concept of Activities in Daily Living as a means of data categorization and to help the data aggregator and the consumers to identify privacy risk values. We provided explanations of privacy cost based on privacy risk, and we modeled the privacy tradeoffs using a game theoretic model and presented the negotiation aspects of the game. Also, we explained the general approach of solving the game and its equilibrium points. We used the concept of differential privacy as an anonymity mean to minimize the leakage of information and as a negotiation parameter to specify the privacy level and the associated payment. The results of or experiments show the importance of consumers' attitude towards privacy. The experiments also provide a principled way to choose reasonable levels of anonymization to have a balanced framework. However, like any model, ours relies on some simplifying assumptions; for instance, we assume that

participants increase their cost linearly with the privacy risk value or with the anonymization level. These assumptions may require a more detailed modeling and user based studies to reflect reality.

Our plan for the future is as follows: We first would like to consider different setup of the negotiation game where the data analyst sends the query first without specifying the level of privacy and price, but instead presented with a menu of optimal prices and anonymity level. This will allow the data aggregator to ensure that prices are set such that, whatever disclosure is obtained by the analyst, all contributing individuals are properly compensated. Second, we would like to study the case where intelligent mechanism installed in smart meters provide personalized anonymity levels for automatic involvement in the market. This will reduce the cost burden on the aggregator and provides consumers with better sensibility of involvement in the decision making of their own assets.

REFERENCES

[1] T. Yu, N. Chawla, and S. Simoff, Eds., *Computational Intelligent Data Analysis for Sustainable Development*. London, U.K.: Chapman & Hall, 2013, ch. 7.

[2] J. Baek, Q. H. Vu, J. K. Liu, X. Huang, and Y. Xiang, "A secure cloud computing based framework for big data information management of smart grid," *IEEE Trans. Cloud Comput.*, vol. 3, no. 2, pp. 233–244, Apr./Jun. 2015.

[3] C. Efthymiou and G. Kalogridis, "Smart grid privacy via anonymization of smart metering data," in *Proc. 1st IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Oct. 2010, pp. 238–243.

[4] S. H. M. Zargar and M. H. Yaghmaee, "An efficient privacy preserving scheme of high frequency reports for secure smart grid communications," in *Proc. 3rd Int. Conf. Comput. Knowl. Eng. (ICCKE)*, Oct./Nov. 2013, pp. 368–373.

[5] H.-Y. Lin, S.-T. Shen, and B. P. Lin, "A privacy preserving smart metering system supporting multiple time granularities," in *Proc. IEEE 6th Int. Conf. Softw. Secur. Rel. Companion (SERE-C)*, Jun. 2012, pp. 119–126.

[6] *Personal Data: The Emergence of a New Asset Class*, World Economic Forum, Cologny, Switzerland, 2011.

[7] *Energy Retailers' Perspective on the Deployment of Smart Grids in Europe*, document EU WG3, Demand and Metering from ETP SmartGrids, 2010.

[8] O. Tene and J. Polonetsky, "Big data for all: Privacy and user control in the age of analytics," *Northwestern J. Technol. Intellectual Property*, vol. 11, no. 1, p. 27, Apr. 2013.

[9] H. R. Varian, "Beyond big data," *Bus. Econ.*, vol. 49, no. 1, pp. 27–31, 2014.

[10] L. AlAbdulkarim and S. Lukszo, "Impact of privacy concerns on consumers' acceptance of smart metering in The Netherlands," in *Proc. IEEE Int. Conf. Netw., Sens. Control (ICNSC)*, Apr. 2011, pp. 287–292.

[11] A. Barenghi, G. M. Bertoni, L. Breveglieri, M. G. Fugini, and G. Pelosi, "Smart metering in power grids: Application scenarios and security," in *Proc. IEEE PES Innov. Smart Grid Technol. Asia (ISGT)*, Nov. 2011, pp. 1–8.

[12] G. Kalogridis, C. Efthymiou, S. Z. Denic, T. A. Lewis, and R. Cepeda, "Privacy for smart meters: Towards undetectable appliance load signatures," in *Proc. IEEE SmartGridComm*, Oct. 2010, pp. 232–237.

[13] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin, "Private memoirs of a smart meter," in *Proc. ACM BuildSys*, Zürich, Switzerland, Nov. 2010, pp. 61–66.

[14] U. Greveler, B. Justus, and D. Loehr, "Multimedia content identification through smart meter power usage profiles," in *Proc. Comput., Privacy Data Protection*, 2012, pp. 1–10.

[15] Y. Yan *et al.*, "A survey on cyber security for smart grid communications," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 4, pp. 998–1010, Oct. 2012.

[16] M. A. Lisovich, D. K. Mulligan, and S. B. Wicker, "Inferring personal information from demand-response systems," *IEEE Security Privacy*, vol. 8, no. 1, pp. 11–20, Jan./Feb. 2010.

[17] J. Strüker and F. Kerschbaum, "From a barrier to a bridge: Data-privacy in deregulated smart grids," in *Proc. 33rd Int. Conf. Inf. Syst.*, Orlando, FL, USA, 2012, pp. 1–15.

[18] J. Strüker, H. Weppner, and G. Bieser, "Intermediaries for the Internet of energy—Exchanging smart meter data as a business model," in *Proc. ECIS*, 2011, pp. 1–13, paper 103.

[19] H. S. Cho, T. Yamazaki, and M. Hahn, "AERO: Extraction of user's activities from electric power consumption data," *IEEE Trans. Consum. Electron.*, vol. 56, no. 3, pp. 2011–2018, Aug. 2010.

[20] A. Yassine, A. A. N. Shirehjini, S. Shirmohammadi, and T. T. Tran, "Knowledge-empowered agent information system for privacy payoff in eCommerce," *Knowl. Inf. Syst.*, vol. 32, no. 2, pp. 445–473, Aug. 2012.

[21] A. Yassine and S. Shirmohammadi, "A business privacy model for virtual communities," *Int. J. Web Based Commun.*, vol. 5, no. 2, pp. 313–335, Mar. 2009.

[22] V. Gkatzelis, C. Aperjis, and B. A. Huberman, "Pricing private data," *Electron. Markets*, vol. 25, no. 2, pp. 109–123, Jun. 2015.

[23] S. Spiekermann, R. Böhme, A. Acquisti, and K.-L. Hui "Personal data markets" *Electron Markets*, vol. 25, no. 2, pp. 91–93, 2015. DOI: 10.1007/s12525-015-0190-1

[24] S. Spiekermann, R. Böhme, A. Acquisti, and K.-L. Hui, "Personal data markets," *Electron. Markets*, vol. 25, no. 2, pp. 91–93, Jun. 2015.

[25] M. Erol-Kantarci and H. T. Mouftah, "Smart grid forensic science: Applications, challenges, and open issues," *IEEE Commun. Mag.*, vol. 51, no. 1, pp. 68–74, Jan. 2013.

[26] G. Taban and A. A. Cárdenas, "Data aggregation as a method of protecting privacy in smart grid networks," *IEEE Smart Grid*, Mar. 2012. [Online]. Available: http://smartgrid.ieee.org/newsletters/march-2012/data-aggregation-as-a-method-of-protecting-privacy-in-smart-grid-networks

[27] F. Knirsch, D. Engel, M. Frincu, and V. Prasanna, "Model-based assessment for balancing privacy requirements and operational capabilities in the smart grid," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, Feb. 2015, pp. 1–5.

[28] C. Riederer, V. Erramilli, A. Chaintreau, B. Krishnamurthy, and P. Rodriguez, "For sale: Your data by: You," in *Proc. 10th ACM Workshop Hot Topics Netw.*, 2011, pp. 13:1–13:6.

[29] I. Bilogrevic, J. Freudiger, E. De Cristofaro, and E. Uzun, "What's the gist? Privacy-preserving aggregation of user profiles," in *Computer Security ESORICS* (Lecture Notes in Computer Science), vol. 8713, M. Kutyłowski and J. Vaidya, Eds. Wroclaw, Poland: Springer, 2014, pp. 128–145.

[30] A. Ghosh and A. Roth, "Selling privacy at auction," in *Proc. 12th ACM Conf. Electron. Commerce*, 2011, pp. 199–208.

[31] K. Ligett and A. Roth, "Take it or leave it: Running a survey when privacy comes at a cost," in *Internet and Network Economics* (Lecture Notes in Computer Science), vol. 7695, P. W. Goldberg, Ed. Berlin, Germany: Springer, 2012, pp. 378–391.

[32] Y. Chen, S. Chong, I. A. Kash, T. Moran, and S. Vadhan, "Truthful mechanisms for agents that value privacy," in *Proc. 14th ACM Conf. Electron. Commerce (EC)*, 2013, pp. 215–232.

[33] D. Xiao, "Is privacy compatible with truthfulness?" in *Proc. ITCS*, 2013, pp. 67–86.

[34] H. Hu, G.-J. Ahn, Z. Zhao, and D. Yang, "Game theoretic analysis of multiparty access control in online social networks," in *Proc. SACMAT*, London, U.K., Jun. 2014, pp. 93–102.

[35] M. Chessa, J. Grossklags, and P. Loiseau, "A game-theoretic study on non-monetary incentives in data analytics projects with privacy implications," in *Proc. IEEE 28th Comput. Secur. Found. Symp. (CSF)*, Verona, Italy, Jul. 2015, pp. 90–104.

[36] R. Cummings, K. Ligett, A. Roth, Z. S. Wu, and J. Ziani, "Accuracy for sale: Aggregating data with a variance constraint," in *Proc. Conf. Innov. Theoretical Comput. Sci. (ITCS)*, 2015, pp. 317–324.

[37] R. K. Adl, M. Askari, K. Barker, and R. Safavi-Naini, "Privacy consensus in anonymization systems via game theory," in *Proc. 26th Data Appl. Secur. Privacy (DBSec)*, 2012, pp. 74–89.

[38] C. Aperjis and B. A. Huberman, "A market for unbiased private data: Paying individuals according to their privacy attitudes," vol. 17, nos. 5–7, May 2012. [Online]. Available: http://journals.uic.edu/ojs/index.php/fm/article/view/4013/3209

[39] L. Sweeney, "K-anonymity: A model for protecting privacy," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 10, no. 5, pp. 557–570, Oct. 2002.

[40] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "L-diversity: Privacy beyond k-anonymity," *ACM Trans. Knowl. Discovery Data*, vol. 1, no. 1, Mar. 2007, Art. ID 3.

[41] C. Dwork, "Differential privacy," in *Automata, Languages and Programming* (Lecture Notes in Computer Science), vol. 4052, M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, Eds. Heidelberg, Germany: Springer, 2006, pp. 1–12.
[42] C. Li, D. Y. Li, G. Miklau, and D. Suciu, "A theory of pricing private data," in *Proc. ICDT*, 2013, pp. 33–44.
[43] J. Hsu *et al.* "Differential privacy: An economic method for choosing epsilon." [Online]. Available: http://arxiv.org/pdf/1402.3329v1.pdf, accessed Oct. 16, 2015.
[44] *SmartData Set for Sustainability*. [Online]. Available: http://traces.cs.umass.edu/index.php/Smart/Smart, accessed Dec. 13.

ABDULSALAM YASSINE received the B.Sc. degree in electrical engineering from Beirut Arab University, Lebanon, in 1993, and the M.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Ottawa, Canada, in 2004 and 2010, respectively. From 2001 to 2013, he was a member of the Technical Staff with the Wireless Communication Division, Nortel Networks, and later at Alcatel-Lucent, Ottawa, Canada. He is currently a Post-Doctoral Fellow with the DISCOVER Laboratory, School of Electrical Engineering and Computer Science, University of Ottawa. His current research interests are mostly focused on systems and networks, multimedia, artificial intelligence, smart environments, and smart grids.

ALI ASGHAR NAZARI SHIREHJINI received the Ph.D. degree in computer science from the Technische Universität Darmstadt, Darmstadt, Germany, in 2008. From 2008 to 2011, he was one of the four vision 2010 post-doctoral fellows with the University of Ottawa, Ottawa, ON, Canada. From 2001 to 2008, he was with the Fraunhofer Institute for Computer Graphics and GMD-IPSI, Darmstadt. He is currently an Assistant Professor with the Sharif University of Technology, Tehran, Iran. His research interests include ambient intelligence, human factors, intelligent agent and multiagent systems, pervasive and mobile games, game-based rehabilitation, massively multiplayer online gaming, and electronic commerce.

SHERVIN SHIRMOHAMMADI (SM'04) received the Ph.D. degree in electrical engineering from the University of Ottawa, Canada. He is currently a Professor with the School of Electrical Engineering and Computer Science, University of Ottawa. He is currently the Director of the Distributed and Collaborative Virtual Environment Research Laboratory, and an Affiliate Member with the Multimedia Communications Research Laboratory, conducting research on multimedia systems and networking, specifically video systems, gaming systems, and multimedia-assisted biomedical engineering. The results of his research have led to over 280 publications, over 75 researchers trained at the post-doctoral, Ph.D., master's, and research engineer levels, over 20 patents and technology transfer to the private sector, and received a number of awards and prizes. He is a University of Ottawa Gold Medalist, a Licensed Professional Engineer in Ontario, and a Lifetime Professional Member of the ACM. He is the Associate Editor-in-Chief of the *IEEE Instrumentation and Measurement Magazine*, a Senior Associate Editor of *ACM Transactions on Multimedia Computing, Communications, and Applications*, and an Associate Editor of the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT. He was an Associate Editor of the *Journal of Multimedia Tools and Applications* (Springer) from 2004 to 2012.