

The Costs of Privacy in Local Energy Markets

Erik Buchmann, Stephan Kessler, Patrick Jochem and Klemens Böhm
Karlsruhe Institute of Technology (KIT)
76131 Karlsruhe, Germany

Abstract—

Many renewable sources for electricity generation are distributed and volatile by nature, and become inefficient and difficult to coordinate with traditional power transmission paths. As a part of the transition from fossil fuel to renewable sources, local energy markets allow an efficient allocation and distribution of energy from local sources to nearby households. When using a discrete time double auction model, bids in such markets reflect the supply and demand of energy. However, since the energy demand of a household contains personal information, such markets are not in line with privacy legislation.

In this paper, we investigate the influence of anonymization methods on local energy markets. In particular, we anonymize the bids of the order book, and we compare the CO₂ emissions and the expenses of market participants of this allocation with a non-anonymous one. We have modeled the flows of personal data for a local energy auction platform, and we have developed a model for the supply and demand of electricity of a small town in the near future. Our experiments show that with elementary anonymization methods, the impact of anonymization on the costs and on the CO₂ emissions is small.

I. INTRODUCTION

Currently, many countries are working on the shift of the electricity supply from fossil to renewable sources [1, 2, 3, 4, 5]. This requires significant changes in the distribution infrastructure [5]. While classical power plants are most efficient with centralized power grid infrastructures, most renewable sources are distributed by nature, and they do not necessarily perform well with long power transmission paths [3]. An important option to distribute renewable energy efficiently is by means of local energy markets [3, 4]. These markets match the regional energy demand with the renewable supply. They bring together local electricity producers, e.g., the operators of small combined heat and power units, biogas plants and photovoltaic installations, and electricity consumers in the vicinity of the producers, e.g., households or enterprises.

Research prototypes [6, 7, 8] have already confirmed that local energy markets allocate renewable energy sources efficiently. However, such markets face a number of privacy challenges: On the one hand, energy consumption data carries sensitive information about consumers. In particular, bids for the usual 15-minute intervals of energy allow the inference of personal details from the buyers, e.g., regarding wealth, habits, daily routine, the presence of specific devices or the number of people of a household. Thus, use, storage and transfer of this kind of information is restricted by privacy directive 95/46/EC [9] and national privacy acts. On the other hand, energy markets must be transparent: National energy laws like the German Renewable Energy Act [10] include the obligation to make the production and distribution of energy transparent. This includes verifiable orders, hence cryptographically sealed

bids do not solve this specific problem [11, 12]. Furthermore, market transparency has a positive impact on the bid-ask spread, the market efficiency and liquidity, and the economic welfare [13, 14]. In addition, grid operators need information to balance an over- and undersupply of energy [15]. See Section II for further arguments.

In this paper, we investigate the impact of simple privacy enhancing technologies on the performance (in economic terms) of local energy markets. Mainly to keep our results intuitive, we use the well-known notion of *k-Anonymity* as an exemplary anonymity measure [16]. To complement our study we also look at the effects of perturbation with *random noise*. We quantify the expenses of the market participants and the CO₂ efficiency of anonymizing the order books of energy auctions. To this end, we have developed a model for the volume and costs of the supply and demand of electricity of a small German town in the near future. Furthermore, we have developed a local energy auction, based on the discrete time double auction model [6], which allows the anonymization of the order book. With a series of experiments, we compare the costs and the CO₂ emissions of market allocations with and without applying anonymization and perturbation methods.

In more detail, we make the following contributions:

- We model the personal data flows for a local electricity marketplace based on a double-sided auction.
- We provide a model for the supply and demand of electricity in a low-voltage circuit in the near future, based on data that corresponds to a small German town.
- We investigate how existing anonymization and perturbation methods can be applied to the order book in order to enhance privacy of the market participants.
- We analyze the impact of privacy methods on costs and CO₂ emissions by means of an extensive study with real-world data.

We show that anonymization is applicable in a local energy market, and that the effects are small for most households. However, unusual power consumption profiles may lead to high costs with the anonymization methods used here.

Paper structure: Section II reviews related work. Section III describes the design of our market, followed by the simulation setup in Section IV. Section V contains our results, which we discuss in Section VI. Section VII concludes.

II. BACKGROUND

In this section, we review energy markets, transparency obligations and privacy aspects for such markets, and anonymization concepts.

A. Concepts for Local Energy Market

Local markets for renewable energies have a number of positive effects: Community-based funding models increase the acceptance and accelerate the installation of renewable power plants [4]. Local markets seamlessly fit into the distributed structure of renewable power generation [3]. Finally, since the price of renewable energy is falling below the one of conventional energy, private consumption and shared consumption is becoming more important [17].

To efficiently allocate energy in a smart grid environment, a number of energy auction mechanisms [6, 7, 8] have been proposed. Due to the homogeneity of electricity and the high number of buyers and sellers on a competitive market, these mechanisms rely on a double sided auction model with software agents for the bidding process. [8] and [7] use a continuous double auction. That is, the market clears continuously by finding the best matching buy and sell orders each time a new order arrives. While [8] focuses on the optimal behavior of the agents, [7] formulates clearing as an optimization problem that maximizes social welfare. Our market design has been inspired by [6]. The approach is based on the stock exchange model, as compared to the EPEX Spot Market [18]. It uses discrete time intervals of 15 minutes each, and the order book is processed on each time interval in isolation.

B. Restricted Access to Electronic-Commerce Data

Cryptographic approaches for auctions encrypt the bids in order to preserve privacy of traders but on the other hand provide verifiable results of the auctions [11, 12]. These methods propose a verifiable cryptographic protocol for trades and make a trustworthy auctioneer unnecessary. On the other hand, they do not comply with the transparency requirements mainly stemming from the legal framework, which we describe next. This also holds for other approaches proposing to restrict access to such data [15, 19].

C. Transparency and Disclosure Obligations

Data from energy markets must be publicly available for three reasons. First, subsidies for renewable energies in a liberalized market require observing energy production and distribution. In consequence, many national energy laws like the German EEG [10] include the obligation to make details of the commercial use of renewable energies public. Second, the degree of transparency has an impact on the market, in terms of liquidity [13], bid-ask spread, market efficiency and welfare [14]. In particular, with auctions lacking transparency, participants price the risk of fraud and cannot adapt to the bidding behavior of the other bidders. Third, to cope with over- and undersupply of renewable energies, distribution system operators and plant operators must be able to forecast the energy demand and the production of renewable sources [15]. Thus, the production and allocation of renewable energy sources must be openly known.

D. Privacy Aspects of Energy Data

Energy data is sensitive data from a privacy perspective. For example, one can infer the daily routine of a household from the peaks in the energy demand resulting from making breakfast, lunch and dinner. One can conclude personal information

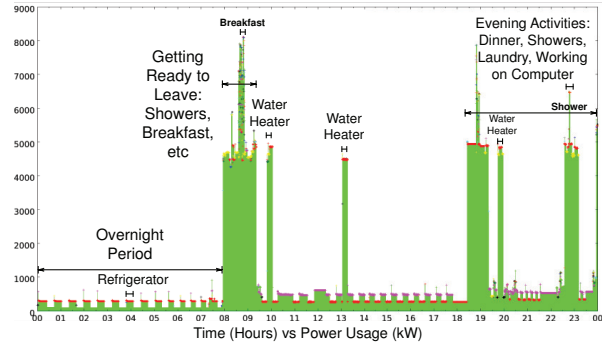


Fig. 1: Smart Meter data, reprinted from [23].

like employment status, vacation timing, church visits or the number of people that live in a household from the distribution of the demand over the day. Fine-grained energy information allows the identification of appliances in the household (Figure 1) or even the television program watched [20]. Energy data also allows re-identification of energy consumers [21], even if the data does not contain unique identifiers. Thus, such data must be protected according to data privacy regulations, e.g., Directive 95/46/EC [9] or national legislation [22]. An important kind of such protection is anonymization, which we discuss next.

E. Anonymization and Perturbation

Anonymization means to protect the privacy of an individual regarding certain features, e.g., his presence in a data set or the fact that he belongs to a distinct group of persons. There is a broad variety of approaches, see [24] for an overview. In this study however, we focus on the effect of an approach that, while being simple, is well-known and intuitive, namely k -Anonymity [16]. While other approaches do offer guarantees beyond the ones of k -Anonymity, they typically do not have those advantages. k -Anonymity makes one individual in a database indistinguishable from $k-1$ others. To this end, it differentiates between quasi-identifiers and sensitive attributes. Quasi-identifiers are attributes allowing to identify some individuals in a database. For example, $\{ZIP, gender, date\ of\ birth\}$ is an identifier for 87% of the US citizens [16]. Sensitive information can be health-status information, religion, etc.

More recent approaches add further privacy guarantees: This includes l -Diversity [25] and t -Closeness [26]. Differential Privacy [27] ensures that the presence or absence of an individual record in a database has a very small impact on the result of a specific database query. Since the local energy market scenario requires publication of the data, as opposed to publishing only specific views on it, Differential Privacy is not applicable here. Another recent approach is geared to energy-data anonymization and divides energy data into high- and low-resolution parts. [28] replaces high-resolution data by pseudonyms, while low resolution data is assigned to identifiers for, say, accounting. A common ground of all these methods is that they modify the data set. k -Anonymity and related approaches make use of generalization and suppression. Differential Privacy relies on adding noise and perturbing the

data set following a Laplace distribution. Thus, anonymization reduces data quality, depending on the extent of anonymity required.

III. A LOCAL MARKETPLACE FOR RENEWABLE ENERGIES

We now describe our local energy market, and how we have realized anonymization.

A. Market Architecture

Local renewable energy sources such as photovoltaic roof installations are connected to the power grid on the level of low-voltage (below 1500 Volts) distribution circuits. Typically, a low-voltage circuit distributes electricity in an urban district or in a village [5]. Since a step-up transformation to higher voltages for long-distance lines would decrease energy efficiency, the local low-voltage circuit specifies the participants for our local energy market [6]:

- **Distribution System Operator (DSO)** In our market, the DSO balances over- and undersupply of locally renewable energy. For example, if the photovoltaic installations cannot meet the demand at night, the DSO transfers energy from the higher voltage grid levels. In case of oversupply, the DSO exports surplus energy.
- **Energy Producer (EP)** Each EP offers an amount of energy for a future period of time. Since larger sources, e.g., wind turbines usually are connected to higher-voltage grids, we consider only micro-combined heat and power plants (CHP) [29] and small photovoltaic sites (PV) [30].
- **Energy Consumer (EC)** The EC places bids for energy at certain time intervals. The typical ECs on a low-voltage circuit are private households. This is realistic, since large enterprises are connected to the higher-voltage grid.
- **Local Market Operator (LMO)** The LMO matches the demand of the ECs with the supply from the EPs. To do so, the LMO manages an order book containing the bids from the ECs and EPs. The order book is public for all market participants and the DSO. To ensure privacy, the LMO anonymizes the bids of the consumers.

Figure 2 depicts our architecture. Each EC predicts its future energy demand. Likewise, the EPs predict their energy production, e.g., based on the weather forecast. The ECs and EPs then place bids for a certain amount of energy at a certain period of time. The LMO computes the outcome of the auction and communicates this information to the EC, EP and DSO. Each EP feeds all energy from its plant to the local low-voltage circuit. If the demands cannot be fulfilled or supply is not consumed, i.e., if the order book cannot be cleared, the DSO imports or exports electricity from higher voltage grid levels. Each EC is charged (a) for electricity provided locally and (b) for conventional energy drawn from the DSO. The price of local electricity is the market price computed by the LMO. Conventional energy is charged at a fixed rate.

B. Model Assumptions

To focus on the influence of anonymization, our architecture simplifies some details.

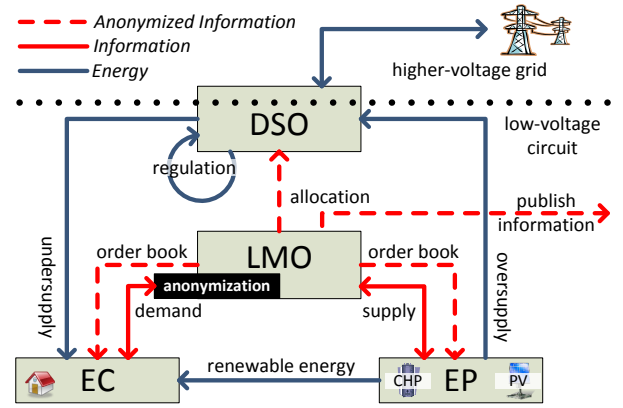


Fig. 2: Architecture of our local energy marketplace

- We assume that EC and EP can exactly forecast their demands and supplies. Incorrect forecasts could be modeled by including stochastic volatility. However, it would be difficult to distinguish the effects of anonymization from incorrect forecasts.
- Since we want to investigate the impact of anonymizing the buy orders, we do not calculate the economic performance of the EPs. Thus, we leave aside that EPs could obtain a compensation for selling oversupply to the DSO.
- We leave aside energy storage and demand response approaches, e.g., Peak Shifting [5]. We plan to address these issues as future work. [31] has shown that it is possible to meet 80% of the demand with renewable energies by smart control strategies without storage.
- We deem CHP a renewable source, because CHP sites generate electricity as a byproduct of heat, and only a small fraction of the emissions can be related to electricity. Furthermore, CHP sites could run on CO₂-neutral energy sources like biogas.

C. Auction Mechanism

Our market is organized in disjunct time slots $t \in \mathcal{T}$ from the time domain \mathcal{T} . To comply with current market standards, e.g., the EPEX Spot Market [18], we use time slots of 15 minutes. The set of market participants \mathcal{M} consists of the energy consumers \mathcal{C} and producers \mathcal{P} , i.e., $\mathcal{M} = \mathcal{C} \cup \mathcal{P}$. To ease presentation, we assume that $\mathcal{C} \cap \mathcal{P} = \emptyset$. Each EC $c \in \mathcal{C}$ has a certain demand $d_c(t)$ for energy at t . Likewise, each EP $p \in \mathcal{P}$ has a certain supply $s_p(t)$.

The order book \mathcal{O} contains the orders of the participants \mathcal{M} . In particular, at a certain time slot t , the order book \mathcal{O}_t stores the buy order $o_c^b(t)$ of a consumer c and the sell order $o_p^s(t)$ of a producer p . $\mathcal{O}_t^s, \mathcal{O}_t^b$ are the sets of all sell- and buy-orders for t . Each order $o \in \mathcal{O}$ stores an amount of energy $o.amount$, an amount of energy that has been allocated from/at the market $o.alloc$ and a limit price $o.limit$. This is the highest price a buyer is willing to pay or the lowest price a seller is willing to get. This results in an order book as shown in Table I.

We apply the *Discrete Time Double Auction*, which is state-of-the-art for energy markets [6]. With this auction type, buyers and sellers specify the amount of the good they want to buy or sell and the limit price. To balance over- and undersupply,

Trader	Sell	Limit	Buy
PV3	3 kWh	0.250 €/kWh	
PV1	5 kWh	0.210 €/kWh	
CHP4	3 kWh	0.190 €/kWh	
PV15	4 kWh	0.160 €/kWh	
HH1		0.240 €/kWh	3 kWh
HH2		0.235 €/kWh	5 kWh
HH3		0.230 €/kWh	1 kWh
HH4		0.220 €/kWh	6 kWh

TABLE I: Order book for single timeslot

the DSO needs to know the allocation in advance. Thus, we use the discrete time variant of the double auction, where the market is cleared at fixed points in time. Each time slot t in the order book \mathcal{O}_t is evaluated independently from all others. Our auction matches the orders from one time slot (1) to compute an equilibrium price that is valid for all orders in this slot, and (2) to allocate the sales and purchases of energy. Note that sell orders or buy orders may be executed only partially. For example, a buy order could be fulfilled partially if there is not enough energy for sale at its limit price. It is also possible that several sell orders partially satisfy a buy order, and vice-versa.

The allocation for a set of orders \mathcal{O}_t works as described in Algorithm 1. Intuitively, the algorithm matches the sell order with the lowest limit and the buy order with the highest limit. This process is repeated until there is no unmatched sell order or buy order any more, or the limits of the remaining orders do not match. The allocation price is the average limit of the last matching orders.

ALGORITHM 1: Discrete Time Double Auction

```

Input: Order book for time slot  $t$ :  $\mathcal{O}_t = \mathcal{O}_t^b \cup \mathcal{O}_t^s$ 
Result: Allocation for time slot  $t$  and price
List sellOrders = sortAscendingByLimit( $\mathcal{O}_t^s$ );
List buyOrders = sortDescendingByLimit( $\mathcal{O}_t^b$ );
Order sell = sellOrders.first();
Order buy = buyOrders.first();
// stop if orders do not match
while sell.limit  $\leq$  buy.limit do
    // match the remaining order volumes
    Int volume = min(sell.amount - sell.alloc, buy.amount - buy.alloc);
    sell.alloc = sell.alloc + volume;
    buy.alloc = buy.alloc + volume;
    // fetch next unmatched orders if necessary and possible
    if sell.amount == sell.alloc then
        if not sellOrders.hasNext() then break;
        sell = sellOrders.next();
    end
    if buy.amount == buy.alloc then
        if not buyOrders.hasNext() then break;
        buy = buyOrders.next();
    end
end
// compute price from last matching orders
price =  $\frac{\text{sell.limit} + \text{buy.limit}}{2}$ ;

```

Note that our auction model is deterministic. Thus, if we run experiments with sufficiently many EPs and ECs, we obtain reliable results.

D. Anonymous Auctions

To investigate the impact of anonymization, we define an anonymization method $\text{anon}(f, \mathcal{O})$ that takes the order book \mathcal{O} and an anonymization approach f as input parameters and returns an anonymized copy \mathcal{O}' of the order book. Since we only anonymize the demand from the ECs, $\text{anon}()$ modifies only the buy orders $\mathcal{O}^b \in \mathcal{O}$.

Applying anonymization implies additional monetary costs and additional emissions. Since anonymization modifies the buy orders, it affects the market allocation. In particular, the ECs might bid for more or less energy than actually needed. If an EC buys too much electricity, some of it will not be consumed. In consequence, the EC pays more. Since this quantity is not assigned to another consumer, the CO₂ efficiency of the market decreases. If an EC buys too little electricity from the local market, the DSO balances the difference to the demand with energy from higher voltage grid levels, at a higher price. Since the way the orders are modified is known to the market participants, this may change their bidding behaviour. This needs further investigation. However, the issue exceeds the scope of this paper and is future work, c.f. Section VI for a discussion. In the following, we assume that consumers bid their true demand and valuation.

Note that the market allocation must be computed on the anonymized order book, since the auction must be executed on the data that is also published. The rationale is to convey to auction participants and other interested parties that the allocation has not been manipulated.

In the following, we describe two implementations of $\text{anon}(f, \mathcal{O})$ which we have used for our experiments, namely k -Anonymity and random noise. We have included random noise in order to have another intuitive reference point albeit it does not provide any guarantees.

1) *k-Anonymity*: k -Anonymity modifies the demand (the value $o_c(t).amount$) of all buy orders so that (1) the sequence of buy orders over all timeslots from one consumer is identical to the sequences of $k - 1$ other consumers (k -Anonymity), and (2) the anonymized demand is as similar to the original demand as possible. We measure the similarity by calculating the distance between original and anonymized order sequences according to the L2 Norm, i.e. for one consumer c ,

$$\text{dist}(o_c, o'_c) = \sqrt{\sum_{t \in \mathcal{T}} (o_c(t).amount - o'_c(t).amount)^2}$$

Since creating a k -anonymous database with minimal changes is NP-hard [32], our Algorithm 2 relies upon a well-known heuristic [33]: At first, it randomly selects a sequence of buy orders o_c from one consumer. It then chooses $k - 1$ sequences of buy orders from other consumers that have the the smallest distances to the first sequence. In particular, method $\text{selectGroup}(k, c, \mathcal{C})$ computes a set of consumers P so that $P \subseteq \mathcal{C}$, $c \in P$, $|P| = k$ and $\max_{i \in P} (\text{dist}(o_c, o_i)) \leq \min_{j \in (P \cap \mathcal{C})} (\text{dist}(o_c, o_j))$. The algorithm repeats until less than k sequences of buy orders are not assigned to a k -group. They are assigned to existing k -groups by method $\text{findGroup}(K, c)$, which returns a k -group J from a set of k -groups K so that $J \in K$ and $\max_{i \in J} (\text{dist}(o_c, o_i)) \leq \min_{j \in (J \cap K)} (\text{dist}(o_c, o_j))$. Finally,

the algorithm replaces $o_c(t).amount$ with the group average for each slot.

ALGORITHM 2: $anon()$ implementation for k-Anonymity

Data: $k, \mathcal{T}, \mathcal{C}, \mathcal{O} = \mathcal{O}^b \cup \mathcal{O}^s$
Result: Anonymized order book \mathcal{O}'
Set $\mathcal{O}' = \mathcal{O}^s$;
Set consumers = \mathcal{C} ;
Set kGroups = \emptyset ;
// create groups of k consumers with similar buy orders
while $consumers.size() \geq k$ **do**
 Consumer $c = consumers.getRandomly()$;
 Set $s = selectGroup(k, c, consumers)$;
 $consumers.subtract(s)$;
 $kGroups.add(s)$;
end
// assign the remaining consumers
foreach $c \in consumers$ **do** $findGroup(kGroups, c).add(c)$;
// anonymize $o.amount$ for each k-group
foreach $kGroup \in kGroups$ **do**
 foreach $c \in kGroup, t \in \mathcal{T}$ **do**
 $o_c(t).amount = \frac{\sum_{j \in kGroup} o_j(t).amount}{kGroup.size()};$
 $\mathcal{O}'.add(o_c(t))$;
 end
end

2) *Random noise:* Random noise means adding random numbers. In this study, they are uniformly distributed in the interval $[-\mu, +\mu]$ kWh and are added to the demand of the buy orders. We ensure that the demand does not fall below 0 kWh, see Algorithm 3. To keep the approach simple, we only vary the absolute value of μ .

ALGORITHM 3: $anon()$ implementation for random noise

Data: Maximal random deviation $\mu, \mathcal{O} = \mathcal{O}^b \cup \mathcal{O}^s$
Result: Anonymized order book \mathcal{O}'
Set $\mathcal{O}' = \mathcal{O}^s$;
foreach $o \in \mathcal{O}^b$ **do**
 $o.amount = \max(o.amount + rand(-\mu, \mu), 0)$;
 $\mathcal{O}'.add(o)$;
end

IV. DEMAND, SUPPLY AND LIMIT PRICES

We now describe our model for the extent and costs of the supply and demand of electricity, which we use to compare the difference between a market allocation based on the true orders and an allocation based on anonymized orders.

We have decided to model a small German town in the year 2016, for three reasons: First, a local energy market operates on the low-voltage power circuit, which corresponds to a small town (see Section III). Second, in 2016 a local energy market can be expected to have a significant economic impact, because right now the generation costs of renewable electricity is about to underprice the electricity prices for households (“grid parity”) in Germany [17]. Third, we have decided for a German town, because facts about energy consumption and renewable energy production in Germany are available. We use the following data sources:

- **Smart Meter data.** Our data set consists of the energy consumption data of 19 households over one year, sampled in intervals of 15 minutes. We know the number of persons of each household, which ranges from one to seven [34].
- **Photovoltaic data.** We have the hourly measured power production over a whole month [34] from a PV installation at our campus, which is mounted on the roof of a building in an optimal orientation.
- **Combined-Heat-Power-Plant Data.** We know the power production of a CHP site located in a private household, over a month in five minute intervals [34].
- **Demographic data.** From the German statistics office, we know the distribution of households [35], e.g., the fraction of one-person households.
- **Statistical data.** We know the usual distribution of the PV peak capacities of a town [36].
- **Legal Constraints.** The Renewable Energy Act [10] states a pricing model for various renewable energies, depending on the year of construction of the source. The CHP Act [37] determines the compensation fee for electricity produced by CHP sites.
- **CO₂ intensity.** We know how much CO₂ is emitted per kWh for the energy delivered by the DSO, depending on the mix of primary energy sources at a specific time of the day [18].

To investigate a meaningful scenario with thousands of EC and EP, we have to interpolate our data sources. In the following, we explain how we generate supply and demand for some thousand households, PV and CHP sites.

A. Energy Demand

We use our smart meter data, demographic data and statistical variance to generate the energy demand for a large number of ECs over a period of time. Our generation process consists of four steps:

- 1) Based on the distribution of demographic data, we calculate how many households of a specific size exist in our town of 5000 households.
- 2) For each size of a household, we randomly select the calculated number of data sets from our smart meter data.
- 3) To account for the fact that households have a different daily routine, we shift each energy consumption value in time, by using a uniform distribution with the parameters $+/- 0.8$.
- 4) To mimic that households have different electrical appliances, we multiply each energy consumption value with a random number. We choose the random number from a uniform distribution with the range $[-0.6; -0.4]$.

We have verified that our generation process does indeed deliver realistic data. To this end, we have calculated the average power consumption of the households and have compared it to the standard load profile “H0” of individuals in Germany [38]. Figure 8, which compares our computed demand to that standard profile, indicates that the two are very similar.

B. Renewable Energy Supply

We restrict our model to small PV and small CHP sites. These are the most popular sources that feed into the low-voltage grid.

1) *Photovoltaic Sites*: The energy output of PV sites depends on (1) the peak capacities, (2) the real electricity production depending on the weather, and (3) a spread that reflects the fact that PV sites on roofs have different angles to the sun.

The peak capacity is the maximal capacity of a PV site, and depends on the number and the quality of the solar modules installed. Figure 4 shows the distribution of peak capacities in our scenario for PV sites with a range below 50kW [36]. Our model contains PV sites with a capacity below 11kW. The utilized capacity is the average percentage of the peak capacity per time interval that is actually achieved under real weather conditions. To compute it, we have used our dataset of the energy production of a photovoltaic site, see Figure 3. Finally, we have to consider a spread resulting from the fact that on-the-roof-sites are mounted in different angles to the sun. For example, a PV site on a roof that is mounted eastward produces most energy in the morning, and a smaller amount of energy during the rest of the day. Accordingly, a site which is oriented westward produces most energy in the afternoon. We simulate the different mounting positions relative to the insulation angle by random shifts in the "time" and "produced energy" dimensions. We determine the random shifts with uniform distributions with parameters 0 to 0.1 for the "produced energy" dimension and 0 to 0.3 for the time dimension.

2) *Micro-Combined Heat and Power Plants*: CHP sites are bound to the heating demand (CHP sites are "heat led"). The demand depends on various parameters, e.g., the capacity of the heat storage and isolation of the house. If the heating of a house is activated, CHP sites produce a fixed amount of energy per time interval. If heating is deactivated, a CHP site consumes a small amount of energy for internal operations (see Fig. 9). From our data sources, we have extracted the probability distribution of starting times and durations. We use this distribution to randomly generate start and stop times for a large set of small CHP sites with a capacity below 1kW.

C. Limit prices

Our local energy market needs limit prices for any buy and sell order. Pricing strategies for double auctions have been extensively studied in the past, e.g., in [39]. However, since our objective is to investigate the impact of anonymization, we require basic pricing strategies that do not add statistical variance to the market allocation. For this reason, we use a strategy where the limit prices of the orders are interpolated from current regulations and market prices, as follows.

1) *Buy Orders*: The limit price of a buy order is the upper bound for the price for energy an EC is willing to pay. If a local EP requires a price above the limit of the buy order, a rational EC could simply buy its demand from the DSO. Thus, the price of the DSO is the upper bound of the market price. At this moment, German households can buy electricity from energy providers at a price of approximately $0.25 \frac{\text{€}}{\text{kWh}}$

[40]. End consumer electricity prices are assumed to increase slightly with the rate of 0.6 % p.a. [41], hence we assume a price of $0.26 \frac{\text{€}}{\text{kWh}}$ in 2016.

2) *Sell Orders*: The limit of a sell order is the lower bound for the compensation a local EP demands from the EC. In Germany, the compensation price for electricity produced by small CHP sites is fixed to $0.11 \frac{\text{€}}{\text{kWh}}$ for 10 years by the CHP Act [37]. If an EC required a lower limit, a CHP site operator simply could sell its energy to the DSO at $0.11 \frac{\text{€}}{\text{kWh}}$.

The compensation fee for German PV sites has been regulated in the German Renewable Energy Act [10]. As Figure 5 shows¹, the costs depend on the year of construction and degrade annually. The costs from 2013 to 2016 are predicted [42] (dotted line in Figure 5). To assign limit prices to different numbers of PV sites, we use the following model: We assume that no PV sites have been constructed before 2012, since prices would not be competitive. For each year between 2012 and 2016, one fifth of PV sites has been constructed, which offer energy with the limit prices shown in Figure 5.

V. EVALUATION

Our objective is to quantify the influence of anonymization for local energy auctions. We have conducted a series of experiments based on the simulation described in Section IV, and we have computed the differences of the electricity bills and the CO₂ emissions for anonymized and non-anonymized auctions. In the following, we describe how we have computed the costs and the emissions caused by anonymization, followed by a description of our experimental setup and of the results of our experiments.

A. Electricity Costs and CO₂ Emissions

According to our assumptions, each consumer knows its true demand at time slot t . With our formal model, the true demand is the amount of a buy order $o_c^b(t).amount$ in the non-anonymized order book \mathcal{O} . Based on it, we calculate the energy costs and CO₂ emissions for anonymous and non-anonymous order books as follows.

1) *Electricity Costs*: The costs of renewable energy of a consumer c are the product of allocation and allocation price, summed over all time slots:

$$costs_c^r = \sum_{\forall t \in \mathcal{T}} o_c^b(t).alloc * price(t)$$

The costs of traditional energy come from an undersupply of renewable energies, which the DSO must balance at a price of $0.26 \frac{\text{€}}{\text{kWh}}$:

$$cost_c^e(t) = (o_c^b(t).amount - o_c^b(t).alloc) * 0.26 \frac{\text{€}}{\text{kWh}}$$

$$costs_c^e = \sum_{\forall t \in \mathcal{T}} \begin{cases} cost_c^e(t) & \text{if } o_c^b(t).amount > o_c^b(t).alloc \\ 0 & \text{otherwise} \end{cases}$$

¹The German Renewable Energy Act distinguishes between PV sites with different peak capacities. However, only small sites directly feed into low-voltage circuits. Thus, we only have to consider one price.

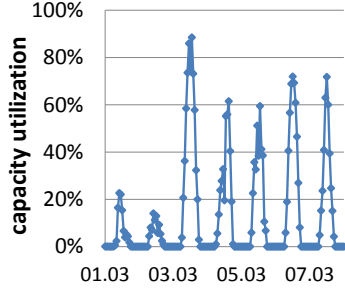


Fig. 3: PV electricity supply

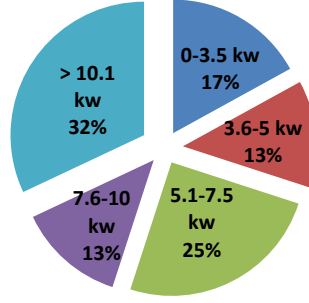


Fig. 4: Distribution of PV capacities

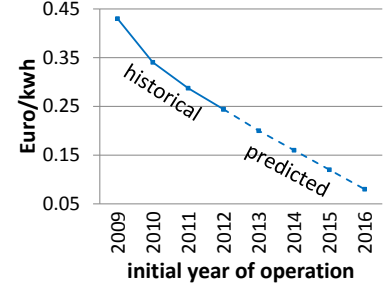


Fig. 5: Compensation for PV sites

The total electricity costs is the sum of the costs of renewable energy and traditional energy, i.e., $costs_c = costs_c^r + costs_c^e$. The energy costs for the anonymized order book \mathcal{O}' , called $costs_c^s$, are calculated in the same way, with one difference: Recall that the DSO balances the true demand. Thus, for the anonymized order book the amount of conventional energy is the difference between the anonymized allocation $o_c^{lb}(t).alloc$ and the non-anonymized order volume $o_c^b(t).amount$.

The CO₂ emissions are produced when consuming conventional energy. Figure 7 shows that the volume of CO₂ emitted per kWh in Germany. It depends on the mix of primary energy sources at a specific time of the day. Our data source is the European Energy Exchange [18]. For traditional energy production, we model the emissions per kWh at time slot t as $\omega(t)$ in $\frac{gCO_2}{kWh}$. For our market, we assume $0\frac{gCO_2}{kWh}$ for PV and CHP sites (c.f. Section III-B).

It is not obvious to compute the amount of additional CO₂ consumption that is caused by anonymization: The anonymized demand $o_c^{lb}(t).amount$ of a consumer c might be higher than the true demand $o_c^b(t).amount$. The difference to the true demand will not be consumed and it does not increase the CO₂ footprint of this consumer. On the other hand, if there is a total undersupply of renewable energy, the other consumers must buy more electricity from conventional sources, which increases the CO₂ emissions. We calculate the additional CO₂ emissions as follows:

$$em(c, t) = (o_c^{lb}(t).alloc - o_c^b(t).alloc) * \omega(t)$$

$$emiss(\mathcal{O}, \mathcal{O}') = \sum_{\substack{\forall t \in \mathcal{T}, \\ \forall c \in \mathcal{C}}} \begin{cases} em(c, t) & \text{if } o_c^{lb}(t).alloc > o_c^b(t).alloc \\ 0 & \text{otherwise} \end{cases}$$

2) CO₂ Emissions:

B. Experimental Setup

Our model town consists of 5000 households. Based on our statistics, this amounts to 9750 inhabitants. In our experiments, we let the number of households constant and vary the number of PV and CHP sites. Thus, the supply of renewable energies varies as well. Table II lists the configurations we have tested. The first three columns show the number of households, PV and CHP sites. The following column quantifies how much

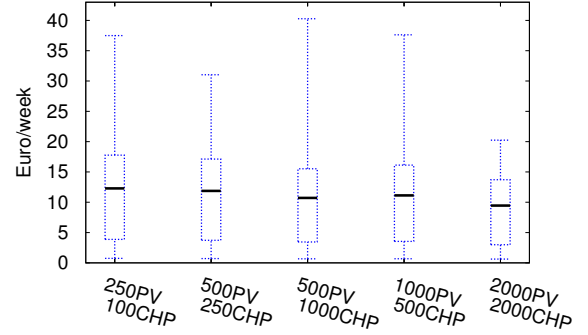


Fig. 6: Total costs for single households

Households	PV	CHP	% regenerative	CO ₂
5000	0	0	0 %	122.19t
5000	250	100	24 %	92.80t
5000	500	250	51 %	70.83t
5000	500	1000	99 %	38.07t
5000	1000	500	103 %	53.22t
5000	2000	2000	268 %	10.89t

TABLE II: Configurations for Simulations and CO₂ emissions

regenerative energy is produced relatively to the demand. We will refer to the configurations with triples containing the number of households, PV and CHP sites, e.g., (5000HH, 250PV, 100CHP) refers to the first line in Table II.

For all configurations we have chosen two different k and μ parameters: $k = 10, k = 20, \mu = 0.01kWh$ and $\mu = 0.1kWh$. On average, a household consumes $0.070kWh$ in 15 minutes. Thus, $\mu = 0.1kWh$ already results in a high deviation of the demand.

C. Experiment Results

We now describe the results of our experiments. As a baseline, we have quantified the costs and CO₂ emissions of our market place without anonymization. We will have successfully shown that anonymization is affordable if our experiments acknowledge that energy costs and CO₂ emissions caused by anonymization are small, as compared to this baseline.

1) *Baseline*: Figure 6 shows the total costs in € per week (y-axis) for each household in each configuration (x-axis), for the non-anonymized case ($costs_c$). We have computed

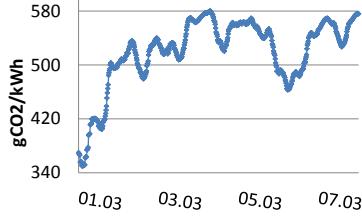


Fig. 7: CO₂ intensity of DSO energy

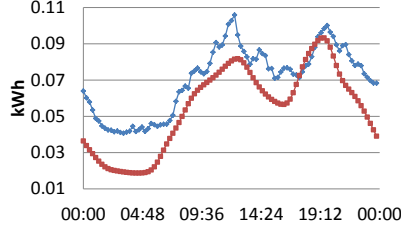


Fig. 8: Validation of our setup

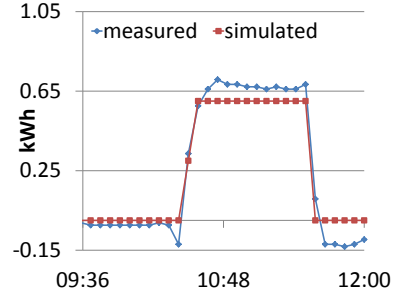


Fig. 9: CHP electricity supply

the maximal, minimal, average and quartile of the costs. For instance, consider configuration (5000HH, 250PV, 100CHP). With this configuration, the average costs of a household is approximately 12.28 €. Since our setup contains households of up to seven members, the maximal costs of a household is 37.49 €. Note that, due to collective use of devices, a seven-person household consumes much less energy than seven times a one-person household. Figure 6 shows that the higher the coverage of regenerative energy, the lower the total costs for a household on average. The relation between coverage and total costs is not linear since demand and supply may not match, and the prices at the energy market are lower compared to the conventional energy providers.

Table II shows the baseline CO₂ emissions and the coverage. Recall that supply and demand fluctuate over time. Thus, a total oversupply (a coverage with regenerative energy with more than 100%) does not necessarily mean that no conventional energy will be consumed. For example, in the configuration (5000HH, 2000PV, 2000CHP), regenerative sources produce 268% of the demand of electricity. Nevertheless, 10.89t of CO₂ will be emitted.

Total costs and CO₂ emissions also depend on the temporal distribution of regenerative energy production. The configurations (5000HH, 500PV, 1000CHP) and (5000HH, 1000PV, 500CHP) nearly have the same coverage, but the configuration with more PV sites has higher total costs and higher CO₂ emissions. This is because CHP sites tend to produce continuously over the whole day, while PV sites have high peaks at midday, leading to a temporary oversupply.

2) *Additional CO₂ Emissions:* Figure 10 shows the additional CO₂ emissions $emiss(\mathcal{O}, \mathcal{O}')$, as defined in Subsection V-A, for each of our configurations and for four different anonymizations. The higher the coverage with regenerative energy, the higher is the impact of anonymization. If the coverage is high, more regenerative energy can be bought from the market and reduces the CO₂ emission (c.f. Table II). However, anonymization modifies the buy orders, i.e., the market allocation deviates from the real demand. This increases the CO₂ emissions, as explained in Subsection V-A.

Nevertheless, leaving random noise anonymization with $\mu = 0.1$ aside, the additional CO₂ emission is lower than 8t. Furthermore, the absolute emission of configuration (5000HH, 2000PV, 2000PV) is one twelfth of the one of a configuration without regenerative energies, i.e., the absolute emissions in this case are low anyway.

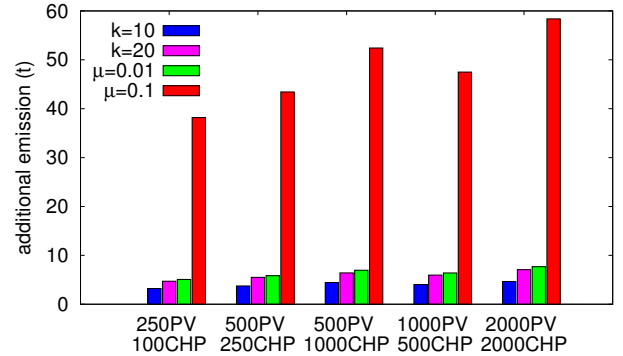


Fig. 10: Additional CO₂ emissions with anonymization

3) *Additional Costs of Anonymization:* To analyze the additional costs that are imposed by anonymization, we subtract for each configuration the energy costs of our baseline (cf. Figure 6) from the energy costs when applying anonymization: $costs'_c - costs_c$. We have calculated the maximal, minimal, average and quartile of the additional costs.

Figure 11a and Figure 11b display the additional energy costs for perturbation with random noise. Random noise with $\mu = 0.01$ incurs additional costs of less than 0.50€/week in the worst case. On average, the additional costs are even lower. The additional costs for $\mu = 0.1$ are approximately ten times higher. Recall that the total energy costs of a household are approximately 12.28 €/week on average, with outliers up to 40.00 €/week per week. Thus, even with this basic method, the additional costs are low. Figure 11c and Figure 11d show the additional costs for applying k-anonymity with $k = 10$ and $k = 20$. For both parameters, in an average case and for the whole quartile, the anonymization leads to additional costs of less than 0.35€/week for $k = 20$ and less than 0.23€/week for $k = 10$. However, in the extreme case, the additional costs are much higher than with random noise, up to 7.65€/week. This is caused by k-anonymity, which averages the demand of sets of k individual households at a time. If a household consumes a lot of energy at unusual times, the distance to the average of the k -group is high. In consequence, the household bids for an amount of energy that is much too high or too low, with high additional costs.

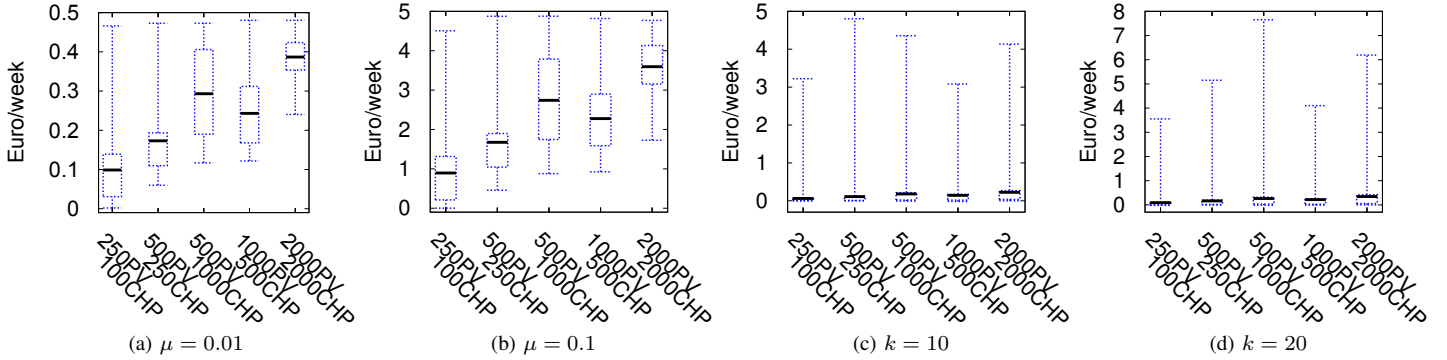


Fig. 11: Additional costs for single households for random noise perturbation and k-anonymity

VI. DISCUSSION

Our objective has been to investigate the impact of privacy enhancement techniques on the performance of local energy markets. We have been particularly interested to find out if applying anonymization is feasible at all, by doing experiments that make use of well-researched anonymization techniques. Our experiments show that the higher the coverage with renewable energies, the more severe is the relative impact of anonymization in terms of CO₂ emission and additional monetary costs. However, in total the absolute additional costs and emissions are still low. Thus, if a society is concerned about privacy to accept a local energy auction as an efficient allocation mechanism for renewable energy, anonymization techniques can certainly be applied.

The results of our experiments have raised a number of directions for future research on anonymous energy markets. First, approaches like k-anonymity do not consider outliers, i.e., households with an unusual distribution of the energy demand. Although the additional costs of anonymity are small for the majority of households, these outliers have to pay a high fee. Thus, other anonymization strategies are needed to give way to low additional fees for all participants. One option might be to mitigate the effect of outliers by insertion of dummy values [43], or by swapping values from different households [44] (see Section II).

Our second observation is that basic anonymization techniques have the effect that the consumers bid for more energy than the actual demand. Thus, due to anonymization a share of renewable energies is not consumed, which increases the CO₂ emissions. One way to reduce this effect would be to adapt anonymization techniques such that the bids are always lower or equal to the actual demand of the households.

Finally, to directly observe the impact of anonymization, we have left aside the effects of sophisticated bidding and pricing strategies, learning effects of the households, effects on properties of the market mechanism (e.g., incentive compatibility), total production costs in fully liberalized energy markets including costs of regulating voltage and frequency in the low-voltage circuit etc. An important future direction would be to find out how these issues interact with anonymization.

VII. CONCLUSION

Since many renewable energy sources are distributed by nature, local energy markets relying on a discrete time double auction model are promising to allocate and distribute renewable energy in low-voltage circuits. However, since the bids reflect the supply and demand of energy of the households, such markets are not in line with privacy legislation: Publishing the data as required by transparency obligations violates the privacy of individuals if anonymization does not take place.

In this paper, we have modeled a local energy market, we have generated a data set that models the supply and demand of energy of a small town in the near future, and we have studied the effects of applying k-Anonymity and of adding random noise. In particular, we have compared the monetary costs and the CO₂ emissions from an anonymized/perturbed local energy market to an unmodified one. We have been able to realistically quantify the influence of anonymization and perturbation on the order book of a local energy auction regarding CO₂ efficiency and monetary costs. For most of the households, the applied methods result in low additional cost. Unusual power consumption profiles however may lead to high costs, especially when using k-anonymity. Furthermore, we have identified promising new lines of research, such as anonymization methods that avoid additional CO₂ emissions and large additional costs when used to anonymize the order book of a local energy auction.

REFERENCES

- [1] Commission of the European Communities, “Renewable Energy Road Map – Renewable Energies in the 21st Century,” Commission Communication COM/2006/0848 final, 2007.
- [2] Bundesministerium für Bildung und Forschung, “Hightech-Strategie 2020 für Deutschland,” <http://www.bmbf.de>, 2010.
- [3] F. Hvelplund, “Renewable Energy and the Need for Local Energy Markets,” *Energy*, vol. 31, no. 13, 2006.
- [4] J. Dobson and C. Julian, “Re-energising Our Communities: Transforming the Energy Market Through Local Energy Production,” ResPublica Green Papers, Available at <http://www.respublica.org.uk>, 2012.
- [5] A. Ipakchi and F. Albuyeh, “Grid of the Future,” *IEEE Power and Energy Magazine*, vol. 7, no. 2, 2009.

- [6] D. Ilic, P. G. D. Silva, S. Karnouskos, and M. Griesemer, "An Energy Market for Trading Electricity in Smart Grid Neighbourhoods," in *6th International Conference on Digital Ecosystems Technologies*, 2012.
- [7] S. Lamparter, S. Becher, and J.-G. Fischer, "An Agent-based Market Platform for Smart Grid," in *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*, 2010.
- [8] P. Vytelingum, S. D. Ramchurn, T. D. Voice, A. Rogers, and N. R. Jennings, "Trading Agents for the Smart Electricity Grid," in *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*, 2010.
- [9] European Parliament and the Council of the European Union, "Directive 95/46/EC," Official Journal L 281, 11/23/1995, p.31., 1995.
- [10] Bundesrepublik Deutschland, "Gesetz für den Vorrang Erneuerbarer Energien (EEG)," Bundesgesetzblatt I/2008 S.1754, 2008.
- [11] F. Brandt, "Fully Private Auctions in a Constant Number of Rounds," *Lecture Notes in Computer Science (LNCS)*, vol. 2742, pp. 223–239, 2003.
- [12] D. Parkes, M. Rabin, S. Shieber, and C. Thorpe, "Practical secrecy-preserving, verifiably correct and trustworthy auctions," *Electronic Commerce Research and Applications*, vol. 7, pp. 294–312, 2008.
- [13] M. Pagano and A. Röell, "Transparency and Liquidity: A Comparison of Auction and Dealer Markets with Informed Trading," *Journal of Finance*, vol. 51, no. 2, 1996.
- [14] D. Bergemann and J. Horner, "Should Auctions Be Transparent?" Centre for Economic Policy Research, Tech. Rep. DP7989, 2010, available at SSRN: <http://ssrn.com/abstract=1707859>.
- [15] A. Wagner, S. Speiser, O. Raabe, and A. Harth, "Linked Data for a Privacy-Aware Smart Grid," *Lecture Notes in Informatics*, vol. 1, no. P-175, 2010.
- [16] L. Sweeney, "k-Anonymity: A Model for Protecting Privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 5, 2002.
- [17] V. Quaschnig, "Einsatzmöglichkeiten und Potentiale der Photovoltaik in Deutschland ohne erhöhte EEG-Vergütung," in *27. Symposium Photovoltaische Solarenergie*, 2012.
- [18] EEX, "European Energy Exchange," URL: <http://www.eex.com>, 2013.
- [19] G. Federal Office for Information Security, "Protection Profile for the Gateway of a Smart Metering System," Available at <https://www.bsi.bund.de>, 2011.
- [20] U. Greveler, B. Justus, and D. Loehr, "Multimedia Content Identification Through Smart Meter Power Usage Profiles," in *5th International Conference on Computers, Privacy and Data Protection*, 2012.
- [21] E. Buchmann, K. Böhm, T. Burghardt, and S. Kessler, "Re-Identification of Smart Meter Data," *International Journal on Personal and Ubiquitous Computing*, 2012.
- [22] Bundesrepublik Deutschland, "Bundesdatenschutzgesetz (BDSG)," Bundesgesetzblatt I/2003 S.66, 2003.
- [23] 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*, 2010.
- [24] B. Fung, K. Wang, R. Chen, and P. S. Yu, "Privacy-Preserving Data Publishing: A Survey of Recent Developments," vol. 42, 2010.
- [25] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkatasubramanian, "l-Diversity: Privacy Beyond k-Anonymity," in *22nd International Conference on Data Engineering*, 2006.
- [26] L. Ninghui, L. Tiancheng, and S. Venkatasubramanian, "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity," *23rd International Conference on Data Engineering*, Apr. 2007.
- [27] C. Dwork, "Differential Privacy," in *33th International Colloquium on Automata, Languages and Programming*, 2006.
- [28] C. Efthymiou and G. Kalogridis, "Smart Grid Privacy via Anonymization of Smart Metering Data," in *1st International Conference on Smart Grid Communications*, 2010.
- [29] M. D. Paepe, P. D'Herdt, and D. Mertens, "Micro-CHP systems for residential applications," *Energy Conversion and Management*, vol. 47, pp. 3435–3446, 2006.
- [30] J. M. Pearce, "Photovoltaics - a path to sustainable futures," *Futures*, vol. 34, 2002.
- [31] F. Adamek *et al.*, "Energiespeicher für die Energiewende," Available at <http://www.vde.com/etg>, 2012.
- [32] G. Aggarwal, T. Feder, and K. Kenthapadi, "Anonymizing Tables," in *10th International Conference on Database Theory*, 2005.
- [33] M. E. Nergiz, M. Atzori, and Y. Saygin, "Towards Trajectory Anonymization: A Generalization-Based Approach," in *SIGSPATIAL ACM GIS 2008 International Workshop on Security and Privacy in GIS and LBS*, 2008.
- [34] IIP, "Energy Database of the Institute for Industrial Production," Available on request at <http://www.iip.kit.edu/>, 2013.
- [35] Statistisches Bundesamt, "Statistisches Jahrbuch 2012," <http://destatis.de>, 2012.
- [36] Transnet BW, "Statistical PV and CHP data," <http://transnet-bw.de/eeg-and-kwk-g/eeg-anlagendaten/>, 2012.
- [37] Bundesrepublik Deutschland, "Kraft-Wärme-Kopplungsgesetz," Bundesgesetzblatt I/2012 S.1494, 2012.
- [38] VDEW, "Lastprofilverfahren zur Belieferung und Abrechnung von Kleinkunden in Deutschland," VDEW-Materialien M-02/2000 Frankfurt a.M., 2000.
- [39] S. Gjerstadt and J. Dickhaut, "Price Formation in Double Auctions," *Games and Economic Behavior*, vol. 22, 1998.
- [40] Eurostat, "Household electricity prices in the EU27 rose by 6.3% and gas prices by 12.6% (25.05.2012)," <http://epp.eurostat.ec.europa.eu>, 05 2012.
- [41] International Energy Agency, "World energy outlook," <http://www.worldenergyoutlook.org/>, 2012.
- [42] Fraunhofer ISE, "Studie Stromgestehungskosten Erneuerbare Energien (30.Mai 2012)," <http://www.ise.fraunhofer.de>, 05 2012.
- [43] A. R. Beresford and F. Stajano, "Location Privacy in Pervasive Computing," *IEEE Pervasive Computing*, vol. 2, no. 1, 2003.
- [44] V. Estivill-Castro and L. Brankovic, "Data Swapping: Balancing Privacy Against Precision in Mining for Logic Rules," in *1st International Conference on Data Warehousing and Knowledge Discovery*, 1999.