SoK: Privacy Technologies for Smart Grids – A Survey of Options.

Marek Jawurek
SAP Research
Karlsruhe, Germany
marek.jawurek@sap.com

Florian Kerschbaum
SAP Research
Karlsruhe, Germany
florian.kerschbaum@sap.com

George Danezis
Microsoft Research
Cambridge, UK
gdane@microsoft.com

***Abstract*—We present the problems associated with protecting privacy in smart grids and survey currently proposed solutions. First we discuss use-cases, the policy environment and the literature on attacks based on monitoring the electricity supply. Then we review the most commonly deployed policy tools, and dive into a detailed survey of privacy technologies for aggregation of readings, billing for consumption and even hiding the consumption from physical adversaries.**

***Keywords*-Smart Grid; Smart meter; Privacy Technology;**

I. Introduction

The Smart Grid modernizes the traditional electricity grid by establishing a communication infrastructure in parallel to the energy delivery network. Household Smart Meters connect to this infrastructure and allow the remote collection of fine-grained readings.

Such deployments have been supported, or even actively driven, by governments in the United States and Europe. On December 19, 2007 the Energy Independence and Security Act was signed into law in the United States which mandates the transition to a new generation "smart grid" architecture. Since September 2009 European Union Directive 2009/72/EC promotes the deployment of smart grids. It states that, where initial pilots are positive, 80% of consumers shall be equipped with smart meters by year 2020.

However, smart meter deployments have raised concerns as being potentially privacy invasive. The consumption data collected by smart meters reflects the use of all electric appliances by inhabitants in a household over time, and allows inferences about behaviours, activities or preferences of inhabitants to be made. Consumer advocates fear that inferred information may be abused for price discrimination, marketing or in unrelated judicial cases.

To date, policy tools, e.g. data protection laws or voluntary regulation, represent the only employed protective measures against privacy invasions by outsiders or insiders of the energy industry. Policy tools, however, only have a deterrent effect and cannot effectively prevent privacy violations. For consumers it would be unfortunate if legislation mandates smart metering, leaving them without choice but to accept the associated privacy threats, as it was the case in the Dutch deployment.

In contrast to policy tools, privacy-enhancing technologies (PETs) can prevent privacy violations before they occur. They implement privacy-preserving protocols that only reveal the minimum amount of information required to achieve a specific purpose.

So far, the research community has provided various and heterogeneous privacy-preserving protocols using different PETs for different applications around smart metering. We believe that, at this point in time, a survey of established research can benefit different stakeholders: It gives practitioners information on viable protocols; it supports researchers in identifying blind spots in existing research and extending it; it informs future standardisation efforts. Furthermore, policy makers can consider the use of PETs, in addition to traditional policy tools.

We provide such a survey with these contributions in particular:

- An overview over the legislative frameworks of the USA and Europe regarding privacy and smart metering.
- A survey of research on inferring behaviour or preferences from the household's use of electricity.
- An extensive survey and systematisation of privacy-preserving aggregation and billing protocols in the field of smart metering. Systematisation includes the employed PET, their assumptions and the system model they operate in, their achieved privacy level and the achieved functionality in comparison to the currently prevailing non-privacy-preserving scenario.

The remainder of this paper is structured as follows: Section II introduces the legitimate uses of readings, the threats and the legal and policy tools that protect privacy. Section III surveys literature around behaviour deduction from household's electricity consumption. Section IV introduces an abstract system model that enables a homogeneous comparison of protocols and the different employed PETs. Section V and Section VI provide overviews and comparisons of privacy-preserving protocols for aggregation and for billing on smart meter data respectively. Section VII surveys alternative approaches that employ batteries at households to mask the consumption of electricity. Finally, we conclude with a summary of our findings in Section VIII.

II. Legitimate uses, threats and policy tools

A. Uses of smart meter personal information

While in the traditional electricity grid monopolies controlled generation, distribution and supply of energy, liberalisation efforts have led to a multitude of interacting roles in an energy market and increased competition between utility companies. Increased interaction between companies has led to increased demand for data to support this interaction. The collection and use of frequent and remote smart meter readings represents the key privacy concern.

The US National Institute for Standards and Technology (NIST) has published an extensive privacy impact assessment as part of the standardisations process surrounding smart grids [45]. Uses and risks relating to detailed energy usage data (in real or delayed time), charging of electric vehicles and energy usage of individual appliances are considered. There is also a distinction between primary and secondary uses of smart meter data. Primary uses are necessary for electricity provision, while secondary uses are value-added services or outright harmful abuses.

The UK Department of Energy and Climate Change (DECC) also enumerates uses of data by actors within the energy sector, as part of its consultation on smart metering and privacy [15]. The structure of the UK energy market is very flexible and as such additional uses were identified. In particular a distinction is drawn between readings collected at monthly, daily and half-hourly intervals.

We present here a merged list of purposes for which meter readings may be collected and used:

Load monitoring and forecasting; Spinning up energy generation is time-consuming and expensive, therefore it is important to generate accurate, data-driven predictions of future consumption. Suppliers need such forecast to buy energy generation contracts that cover their clients, and network operators (or Distribution Supply Operators) require longer term forecasts to ensure the necessary network capacity is available. As such poor forecasts have a direct financial impact.

Efficiency analysis, monitoring and advice; Consumer engagement is necessary to reduce energy consumption. One way to achieve it, is by presenting to consumers real time information about household consumption. For a vibrant market to exist, consumers must also be able to compare tariffs of different suppliers on their consumption. Both functions may require detailed per-consumer readings, but are not absolutely necessary to provide electricity.

Billing; Smart meters enable accurate bills on arbitrary short periods to be computed, and moving away from estimated bills. Providing correct bills provides immediate feedback on household energy efficiency. Furthermore, time-of-use bills can expose the consumers to a more accurate price of electricity at different times in the day. Micro-generation can be supported by measuring household energy input into the grid.

Demand response; Demand-response technologies allow specific domestic devices, like fridges or air-conditioning, to reduce their energy consumption at peak times. Smart meter readings can be used to direct centrally demand-response signals; are natively purely device side approaches can be used to regulate consumption.

Fraud detection; All meters offer a certain degree of tamper resistance and detection. Fraud can also be detected statistically, for instance, by comparing an aggregate consumption with sums of several smart meter readings, or measuring unexpected voltage drops on a line. Tracing the source of the leak or theft might be necessary.

Settlement; An energy market requires suppliers to pay for any difference between contracted generation and the actual energy used by their customers. In theory accurate settlement would require smart meter data at the granularity of the settlement mechanism.

Bill energy consumption to owner of the PEV; It is necessary to authorise and bill for charging electric vehicles away from home, which leak both readings and the location data.

In Section V we list a number of privacy-preserving aggregation protocols, that support the aforementioned activities. They can be used to calculate statistics by aggregating across consumers for load monitoring and forecasting or settlement, for instance. Simple sum-aggregation protocols support the required comparison for fraud detection.

Billing is a special case of aggregation, i.e. the result of aggregation over one consumer's data must be attributable to this consumer, thus Section VI presents privacy-preserving billing protocols separately.

B. Perceived threats

While many threats are associated with smart metering, e.g. confusing tariffs and security of critical infrastructure [**?**], this work focuses on *privacy threats*. Smart meter data (as described by Section II-A) can be misused by legitimate receivers inside the energy industry, unauthorised or unauthorized third parties, or governments. Actors in the energy industry may use household data to price discriminate or profile households to maximise revenue. This has implications for consumer protection and fair competition. Third party industries, such as building & insulation companies, may use energy consumption data to target marketing material. Finally, governments may require access for law enforcement purposes, or as evidence in criminal proceedings. Thus privacy concerns exist about the use of the data by industry insiders, but also outside parties that may be able to access the data legally and legitimately.

These privacy concerns are worsened by fears that smart meter data may fall in the hand of non-authorized parties,

through hacking or accidental data loss. For example, a burglary ring may use smart-meter data to get a real-time map of which households are unoccupied. Frequent readings can be used to infer the lifestyle or religion of occupants, which can be used at a large scale for discrimination. Finally, smart-meter data from high value households, of politicians or security staff, may be the target of foreign intelligence agencies.

Not only the frequency of collection but the scale of collection and retention periods add to the perceived threats. Although, the observation of single homes may be easy when meters are located outdoors, smart metering may enable the mass surveillance of very large number of homes at very low cost. The long-term retention of readings also represents a significant problem, and interferes with nascent concepts of "the right to forget". It allows the discovery of readings in later legal proceedings and mining long periods can uncover indications of illness or changing lifestyle.

C. Privacy-related legal frameworks

The protection of privacy – which includes the privacy of smart meter related data – is supported by law in different ways in different jurisdictions.

The United States currently takes a sectoral approach to legislation relating to privacy, by protecting specific sensitive items in specific industries. The policy framework for securing the 21st century grid has been published in June 2011 by the executive office of the president of the United States. It highlights the lack of federal sectoral regulation protecting the privacy of smart meter readings, while acknowledging that detailed reading information can be sensitive. It calls for such information to be processed according to the Fair Information Practice Principles [13] that are based on the five principles of (1) Notice/Awareness (2) Choice/Consent (3) Access/Participation (4) Integrity/Security and (5) Enforcement/Redress. The policy framework also expresses the interest of the presidency in including the protection of smart meter data in any future general statute concerning privacy protection.

While there is a lack of U.S. federal privacy regulations for smart meter data, state level sectoral regulation have been introduced including in California, Colorado, Ohio and Oklahoma [49] through their respective public utilities commissions. Janice Tsai [56] describes in detail the political and legislative process of the California Senate Bill 387 on the subject of smart meter privacy protection that ultimately failed to become law in August 2010. The subsequent California Senate Bill 1476 on privacy in smart metering prohibits the sharing, disclosing or selling of a customer's personal data or consumption data. It became law on January 2011. Furthermore, in July 2011 the California Public Utilities Commission issued a decision that broadly imposed the Fair Information Practice Principles on the processing of smart meter data by the local utilities, as well as third parties that come to contact with utilities to access data [1]. This decision was the fist to provide comprehensive protections and has been used as a template in other states.

The European Union and Canada benefit from a *horizontal* data protection regime, i.e. a set of laws and regulations that impose restrictions and obligations when processing any type of personal information. Over the years it has been established that the data protection regime applies to smart meter data.

To fall under the remit of data protection data first must be considered personally identifiable – this means that it can be linked with a living individual. Some data, such as subscriber information are by definition personal. Other information may become personal by virtue of its association with a living individual. For example, in the context of smart metering, detailed half hourly consumption readings are considered private [2].

On the one hand, in a large household or communal dwelling meter readings cannot be directly attributed to single individuals and thus automatic classification as personal data can be controversial. On the other hand, information contained in smart meter readings may disclose attributes that are so sensitive, e.g. religion (praying patterns, following the Sabbath), that they enjoy a special status in some data protection laws.

A second policy question relates to "anonymous" consumption data from smart meters. One could argue that severing the link between the consumption profiles and the subscription information makes the data unlinkable to a living individual, and therefore not subject to data protection regulation. The German data protection law, for example, explicitly excludes anonymized data. However, it defines anonymous data as data that can only be de-anonymized with "extensive, disproportionate effort". This effort is not further specified. In fact such anonymization before further processing was proposed as an early privacy technology for smart metering [18]. Yet, naive anonymization of smart meter data proves to be fragile [30] (see Section III-E). Similar spectacular results in deanonymizing social network privacy [44] have cast doubts on whether it is ever sufficient to simply remove obvious identifiers to anonymize data safely.

How smart metering data can be lawfully processed within Data Protection frameworks is studied in detail by Knyrim and Trieb [35] and also discussed by the EU article 29 Decision [2]. Interestingly, consumer consent can be considered meaningless in most cases as Data Protection legislation makes processing legal anyway for the fulfilment of legal or contractual obligations – such as for billing or ensuring network stability. Use of smart meter data for value-added capitalization services, on the other hand, requires consent as those are not strictly necessary for the provision of electricity. The DECC devoted a large part of consultation on the problem of distinguishing the latter from the former.

The DECC proposal [15] is to require a certain data item to be made available to the energy industry, when it relates to the fulfilment of "regulated duties". Other information, including detailed consumption profiles that could be used to implement time-of-use tariffs, would only be available if the customer has entered into such a specific contract that requires them.

Other principles of data protection have ramifications for the processing of smart meter data and technologies around it. Data should only be collected and used for specific and well-defined purposes; it should be accurate; it should be kept securely; and the data subjects have some rights concerning the processing.

Most interestingly, the minimization principle states that "Personal data shall be adequate, relevant and not excessive in relation to the purpose or purposes for which they are processed" (from the UK Data Protection Act). This is viewed as mandating the least privacy invasive option when considering technical alternatives – possibly opening the door to privacy technologies. In fact the principle is interpreted as: *all other things being equal* the most privacy friendly option should be used. In practice other things are hardly ever equal. One could argue, for example, that the cost of developing, implementing or maintaining a more privacy-friendly system may be higher, just because those systems are less common.

D. Policy Tools

In current smart grid deployments the protection of personal information is overwhelmingly achieved through "policy tools", i.e. organisational mechanisms and rules, rather than through Privacy Technologies that attempt to embed privacy protections (at some level) within the smart grid technology.

Policy tools, in contrast to privacy technologies, have been considered as sole solutions by regulators and industry for two reasons: First, they are simple, thus enabling regulators, industry lobbyists, consumer advocates, politicians and the public at large to discuss and comprehend what the rules are about. Privacy policies, and the respective policy tools that implement them, are described at a very high level and are often one and the same. Secondly, policy tools are more flexible in contrast to privacy technologies, since they can be changed without the need to change expensive technical systems.

However, the key weakness of policy tools relates to what is called in security engineering "strength of mechanism": The aforementioned ease with which policies can be changed also makes it easier for authorised or unauthorised parties to violate the policy. Policy tools rely on the honesty of all parties that come into contact with data to enforce the policy, while privacy technologies prevent potential abuse in the first place through technical protections.

Since our main focus is on privacy technologies, we only provide here a short description of policy tools that have been used to protect privacy in smart metering initiatives.

The first policy tool to regulate the collection and processing of personal data is to ensure *consent*. If a subject gives consent to the processing of their personal data there are few limits on what can be done with it. Free consent is quite problematic as in some jurisdictions either the whole (Canada) or parts (network operators in Germany) of energy provision are licensed monopolies and thus choice is limited. The aforementioned (Section II-C) concept of "regulated duties" in the UK stipulates that in exchange for basic electricity provision consumers must supply a basic set of data items, one reading a month. Consumers can opt-out of daily readings and have to opt-in for more frequent readings that allow time-of-use tariffs. Readings for fraud detection can be collected without explicit consent of suspected households.

Generic policy tools are also inherited when smart meters are deployed in countries with a horizontal data protection framework, such as the EU and Canada. Such jurisdictions require data controllers to allow customers to access and correct faulty data. Furthermore, only a minimum amount of data must be collected for, and deleted after, the completion of a specific purpose. Appropriate measures need to be taken to avoid unauthorised access to data during transmit and storage. Although all these could be supported by privacy technologies, mostly procedural measures are taken, e.g. storing data on servers behind some access control, rather than laptops, or sending data by courier, rather than the post system. Most importantly, the data minimization principle has so far not been interpreted, neither by businesses nor regulators, to mandate privacy technologies that collect less data than procedural approaches.

The office of the privacy commissioner of Ontario in Canada has spearheaded the approach of "privacy by design" within the local smart grid deployments [28]. As their report describes, the key approach is to analyse end-to-end all processes within the smart grid that collect or operate on private data and ensure that appropriate policies are applied to them at each stage. Broadly, this approach is a policy tool applied to the design of the smart grid architecture. Similarly to the drafting of a formal security policy, or a threat model (as part of a secure design methodology like Microsoft's SDL) it allows designers to keep track of where the private data is, and ensure that only the appropriate operations are performed on it. Yet, it does not mandate particularly strong technical privacy protections for the final system.

Finally, sectoral and industry codes of practises are popular when a voluntary, self-regulation is preferred. However, the strength of mechanism they provide is uncertain, since the standards for auditing or enforcement of compliance with these codes are uncertain.

III. Electricity usage - a covert channel

Smart meter deployments are young, and there is little computer security literature on realized threats, such as those listed in Section II-B, on real smart meter data.

Yet, the oldest usage of data mining of electricity consumption for mass surveillance dates back to Germany 1970s. Police used records of low energy consumption and cash based payment to find the safe-house of Rolf Heissler, a member of the Red Army Faction [5]. At the time this practice was ruled illegal. Since then monitoring the energy supply of homes has been associated with the detection of cannabis plantations [16].

On the technical side, Anderson *et al.* first raised questions about the security of prepaid meters in South Africa [6]. More recently, deployed automatic meter reading (AMR) systems have been shown to be susceptible to simple eavesdropping attacks by Rouf *et al.* [52]. These have been limited to illustrating technical ways to extract information from meters, rather than inferring household activities.

However, several works in related research fields have shown how, in theory and in practice, information can be inferred from a household's electricity consumption depending on the granularity of the readings.

A. Non-intrusive load monitoring (NALM)

The idea behind NALM is to identify appliance usage in households without the need for expensive, laborious and intrusive sensor deployments in homes. A single non-intrusive electricity sensor located anywhere in the house or at the electricity meter provides the necessary data. Although, NALM relies on higher resolutions readings (kHz - hz range), in contrast to the 1hz resolution of current smart meters, technically it is very similar to smart metering.

Data gathered with NALM can be used for load forecasting, help appliance designers, provide information for energy saving audits, help detect appliance failure, support security monitoring and home automation in general and enable demand side load management. The key challenge of the single point sensing is to disaggregate the influences of different electrical appliances on the power line to identify the respective appliances.

The NALM research field dates back to the early 1980. G.W. Hart [25], a particularly active researcher in this field, presents a comprehensive list of publications in this field that emerged between 1983 and 1995. Prudenzi [47] also provides a good overview of the field.

Hart [26] presents a two-phased approach to solve the problem of disaggregating appliances from consumption data: First, a training phase and then an actual working phase. The training phase learns the electricity profile of different appliances, either by manually or automatically turning them on and off. The training can also be augmented or replaced by a database of known appliance signatures. After the training the profiles are used to identify devices in the aggregate consumption.

In contrast, Molina-Markham *et al.* [43] use inhabitants' diaries as auxiliary information to correlate appliance usage and power consumption data. Similar power consumption events are clustered and tagged with duration, power-step at beginning, average power and shape. Then, automated appliances are filtered out. Under the assumption that low power consumption means low human activity all power events occurring in lower power consumption periods are associated with automated appliances. Remaining power events are then correlated with appliances using inhabitants' activity diaries.

In 1992, G.W. Hart [27] first raised concerns about the impact of NALM on household privacy. He identifies NALM as a potential surveillance technology and suggests legal frameworks to limit the information gathering with NALM technology to legitimate purposes. Specifically, he suggests that burglars might use NALM to time their break-ins when occupants are not at home or when they are showering. Furthermore, he identifies that information collected through NALM can be used for direct marketing to consumers with "junk mail" or targeted advertisements.

While the detection of appliance activation, appliance class (oven, fridge, ...) or even appliance make and type allows the deduction of inhabitant behaviour to some extent, the use mode, i.e. how appliances are used, provides even more detailed private information about inhabitants.

B. Use mode detection

Use mode detection goes one step further than detecting which electric device is active: it tries to infer the actual activity performed with the device.

Enev *et al.* [20] analyse electro-magnetic interference (EMI) of television sets on the power line and how it varies with the content displayed. First in a lab environment and then in homes they analyse 8 different television sets from 3 manufacturers and 3 different sizes. EMI traces are measured with a 2hz resolution on the power line with other appliances also attached and in use. After digitalization and transformation to the frequency domain only frequencies of 0-250kHz were considered for analysis. They show that the same video produces repeatable EMI traces on the same television set and highly correlated EMI traces between different television sets. Furthermore, they show that movies can be identified from EMI traces of 1200 known movies with up to 92% accuracy for some sets. This supports the identification of television channels for a known programme and time. Finally, the EMI signature model of an arbitrary television can be learned from known video content and does not require the knowledge of the television make and experiments in the lab. The remote analyses of EMI traces of unknown television sets makes this attack applicable to any household. As countermeasures they propose either the

injection of high-energy broadband noise onto the power line or a power line isolator for every EMI emitting device.

Greveler *et al.* [24] also describe an approach to identify displayed TV channels. First, they note that the smart meter measurements of their sealed smart meter (part of a commercial smart meter deployment) are neither encrypted nor cryptographically signed. Second, they use the 0.5 hz smart meter measurements to create a prediction function that predicts the energy consumption of a dynamic back lighting LCD depending on the brightness of displayed content. Effectiveness of their approach is documented with three movies where prediction and actual consumption match with a Pearson coefficient of 0.93/0.94/0.98 respectively. Further, they extend their analysis to several TV sets with dynamic back lighting and describe how their approach can be automated and used in TV channel identification.

Clark *et al.* [54] employ direct load monitoring at a computer with a sampling frequency of 1khz to identify which website, out of a pool of 8 popular websites, the computer is downloading and rendering. The authors apply different classification techniques like length, motif and cross correlation of consumption traces during web page download and rendering. They find that combinations of different classification techniques (length and cross correlation) produces good accuracy (approx. 60%) and no false positives. Furthermore, they analyse how susceptible the different classification techniques are to website changes. As potentially effective defence strategies they propose either excessive CPU utilisation or random delays which both would imply inconveniences for the computer user.

Bauer *et al.* [9] also employ direct load monitoring at different appliances. In a first training phase, they record feature vectors of all appliances in their different use modes. That means, they measure different amounts of water in the water boiler, fridge with open door/close door, bread cutter cutting bread or salami and so on. From these feature vectors they derive rules that allow to infer the use mode. With their approach they achieve prediction accuracy of 80-90% for use modes of isolated appliances while each use mode can still produce different consumption traces depending on the appliance operator.

While the attacker model behind use mode detection at the appliance is more powerful than the NALM attacker this advantage might be reduced in the future with improved techniques for disambiguation of load signatures in NALM approaches. Then, NALM approaches might achieve a similar level of detail in activity recognition as (intrusive) appliance load monitoring approaches.

C. Behavior deduction

The analysis of electricity consumption allows one to infer inhabitants' behavior, either directly or indirectly from used appliances or their use modes.

Lisovich *et al.* [40] attempt to predict inhabitants behavior directly from smart metering data. Over the course of two weeks they conducted an experiment that collected electrical data and video surveillance in a student flat. Their developed system detected load events from the electrical data and predicted behaviours. Then they evaluate the performance of their behavior prediction with manually extracted control data from the video surveillance. Finally, they construct a sample disclosure metric that divides their behavior deductions into categories like presence, sleep schedule and others and rates the disclosure in those categories according to the accuracy of their behaviour extraction system and the sensitivity of the categories.

D. Other utilities or combinations of utilities

Smart metering has also been discussed to be used for other utilities than electricity, e.g. gas or water. Their consumption traces can also be used to infer inhabitant behaviour as has been shown by Froehlich *et al.* [22]. The authors employ a single pressure sensor to the water infrastructure of a home to identify water usage events and associate them with a distinct fixture/appliance. First in a manual training phase, data of pressure events and flow rate for the opening and closing of all valves are collected in ten different homes. Automated analysis consists of first identifying the start and end of events, then distinguishing between open/close events and finally attributing events to a specific fixture. Although in their experiments all valve events were identified in isolation of each other the authors note that in a realistic setting concurrent valve events may occur and that valves can also only be opened partly or slower/faster than in their experiments.

Kim *et al.* [34] describe how measurements of two utilities (water and electricity) can be used to disambiguate findings of one another. This way a single sensor infrastructure-mediated water pressure approach can be amended by auxiliary information from an electricity meter. In particular the authors also mention potential privacy problems. During their experiments they observe one inhabitant that flushes a toilet twice and another that only flushes once but does not use the sink afterwards and does not switch the bathroom light on.

E. Attacks on anonymized smart metering consumption traces

All previously presented approaches to infer appliance use, use modes or behavior prediction require identity information to play to their full potential. First, identity information about the inhabitants is useful if one wants to apply auxiliary information during use mode deduction like in Molina-Markham *et al.* [43]. Second, even if auxiliary information is not required, inferred behavior information cannot be attributed to an individual if identity information are not available.

Jawurek *et al.* [30] argue that pseudonymized consumption traces, that means consumption traces that have been separated from identity information, can still be attributed to individuals. Using a support vector machine approach on real, but pseudonymized smart meter data of 53 households over 221 days they create a classification that allows to distinguish consumption traces of different households. Their results indicate that the consumption trace represents a fingerprint of its household. Furthermore, they argue that using auxiliary information like household observation correlations between power events and physical events can lead to de-pseudonymization of the household's consumption trace. Finally, the authors identify that only very low smart metering resolutions or frequent re-pseudonymizations can mitigate these attacks.

IV. Privacy Enhancing Technologies

Policy tools, as described in Section II-D, act deterrently but cannot effectively prevent actual privacy violations. In contrast to that, privacy-enhancing technologies (PETs) provide strength of mechanism to support privacy policies at the technical level.

First, for the differentiation of PETs and for the systematisation of aggregation (Section V) and billing (Section VI) protocols we define an ideal (w.r.t. utility) but not privacy-preserving scenario:

Assume that a multitude of data producers (e.g. smart meters) is connected to one service provider (e.g. supplier). Every data producer continuously measures a private data item that represent the amount of service consumption (e.g. energy consumption) by the associated household for a specific atomic time slot (e.g. every 15 minutes). Furthermore assume, that the service provider (WLOG) is interested in calculations over these data items, hence hereafter called data consumer. For this reason, data producers send all their private data items to the data consumer who subsequently performs arbitrary aggregation over arbitrary subsets of data items which yields the desired aggregate. Note, that in this ideal scenario the data consumer also plays the role of an aggregator for which some surveyed protocols explicitly introduce a third party.

This scenario's potential privacy problem is that, in addition to the actually required aggregate, the data consumer also receives private data items that are not required per se. Here, policy tools can act deterrently but cannot actually prevent privacy violations once the data consumer has obtained private data items.

In contrast to policy tools, PETs prevent situations that might result in violation of privacy in the first place. In the following we describe general types of PET that can be employed alone or in combination and provide a mapping (Figure I) of surveyed protocols to the PETs they employ:

Anonymization: The idea of anonymization is, that the data consumer might still be able to perform the calculation although the link between data item and creating data producer has been removed. The privacy-enhancing effect of anonymization is based on the data consumer's inability to attribute information learned from private data items to data producers. However, anonymization may turn out to be ineffective: Depending on the nature/accuracy/frequency of data items, it might still be possible to infer the identity of the data producer, or at least distinguish between data producers ([30]), from information contained in the data items themselves. Obviously, this technique is only applicable if the result of the computation does not have to be attributed to a specific data producers, i.e. it is not suitable for billing.

Trusted computation: In trusted computation one can distinguish between two variants: The data producer himself is trusted to perform the computation himself or an additional trusted third party (TTP) is introduced as external (to the data consumer) aggregator into the scenario. Either way, the data consumer only receives the aggregation results and no private data items. On the one hand, depending on how much trust is given to a TTP it represents a rather strong assumption. On the other hand, it also enables protocols to almost achieve the full flexibility of the ideal scenario.

Cryptographic computation: Cryptographic computation relies on the homomorphic property of encryptions (E) or secret sharing schemes (S). Data items arrive at the data consumer as ciphertexts or secret shares and the protocol ensures that the data consumer can only decrypt the aggregate of data item ciphertexts or secret shares but not individual data items.

Perturbation: By deliberately introducing error into data items or the final aggregate one might preserve the utility of the computation required by the data consumer but sufficiently protect data producer privacy. Protocols that employ this strategy, often aspire to provide differential privacy [17] for the aggregation function. Differentialy private aggregation functions add sufficient amounts of random noise to their result, so that individual input data items cannot be inferred.

Verifiable computation: In verifiable computation the aggregator also provides a proof along the aggregate that the calculation has been performed as claimed. Thus, untrusted aggregators can perform the computation while guaranteeing the integrity of the computation's result. Such proof can be provided by interacting as prover in an zero-knowledge proof system (ZKP) with the data consumer as verifier. In a ZKP a verifier only learns the veracity of the statement to be proven but no information that he has not already known before the proof. Protocols following the verifiable computation strategy can differ in the location where the calculation and proof are computed (on site at the data producer or elsewhere) and in their employed cryptographic

Table I
EMPLOYED PRIVACY TECHNOLOGIES (PET) BY WORK

Paper Employed method	Anonymization	Trusted computation	Cryptographic computation	Perturbation	Verifiable computation
Aggregation					
Efthymiou *et al.*[19]	X				
Molina-Markham *et al.*[43]	X				
Bohli *et al.*[10] TTP	X				
Jeske [32]	X				
Petrlic [46]	X				
Ruj *et al.*[53]		X	E		
Li *et al.*[38]		X	E		
Rottondi *et al.*[51]			S		
Kursawe *et al.*[36]			S/E		
Erkin *et al.*[21]			S/E		
Garcia *et al.*[23]			S/E		
Lin *et al.*[39]		X		Optional	
Bohli *et al.*[10] noisy				X	
Shi *et al.*[55]			S	X	
Rastogi *et al.*[48]			S/E	X	
Chan *et al.*[12]			S	X	
Jawurek *et al.*[31]			E	X	
Acs *et al.*[3]			S	X	
Billing					
Lemay *et al.*[37]		X			
Jeske [32]		X			
Petrlic [46]		X			
Jawurek *et al.*[29]					X
Rial *et al.*[50]					X
Danezis *et al.*[14]					X
Molina-Markham *et al.*[43]					X
Molina-Markham *et al.*[42]					X

primitives.

V. AGGREGATION

As introduced in Section II-A privacy-preserving aggregation protocols can support various activities on smart meter data.

In comparison to the non-privacy-preserving scenario introduced in Section IV privacy-preserving protocols may exhibit reduced utility w.r.t. different dimensions. Thus, we first describe the dimensions that we use to characterise protocols in the remainder of this Section and in the summarising Table II.

Aggregate function: The ideal protocol allows the data consumer to freely choose an *arbitrary aggregation* function. Furthermore, its aggregation function might span *arbitrary subsets* of data items. This is due to the data consumer also being the aggregator at the same time and having all data items in unencrypted form to his disposal. Some protocols offer reduced flexibility w.r.t. the aggregate function. Either, the aggregation function is limited to *weighted sums* or even only supports *sums*. Some protocols might also only allow either a *temporal aggregate*, i.e. an aggregate over a set of one data producer's data items, or a *spatial aggregate*, i.e. an aggregate covering single data items of multiple data producers.

Data producer synchronisation: In the ideal scenario, data producers operate *asynchronously*, i.e. they report their data items independently of each other to the aggregator. That relates to times of sending and the handling of data items. This implies (in the ideal protocol) that the aggregator can use any arbitrary subset of data items received in the system's whole lifetime for aggregations. However, some privacy-preserving protocols require *synchronous* data producers, e.g. if the protocol operates in rounds which requires all data producers to send approx. at the same time or with interdependent key-material. While synchronisation usually reduces the flexibility w.r.t. the aggregation function at the aggregator it also imposes higher complexity of data producers and protocol.

Fault-tolerance: Data producers might either suffer from hardware or software failures or from a failure of their communication link. While some protocols can cope with *unlimited failures* of data producers some protocols only tolerate *limited failures* or *no failures* at all. In case additional entities are introduced into the protocol (e.g. TTP or an aggregator different from the data consumer), their fault-tolerance might also influences the robustness of the entire system.

Communication model: In the ideal protocol data producers merely require a uni-directional ($\rightarrow$) communication link directly to the aggregator. However, some protocols have stronger assumptions, they require bi-directional ($\leftrightarrow$) communication links between protocol entities.
In this respect, it is also of interest, how data producers (P) are connected to the data consumer (C). They might directly report to an aggregating data consumer AC or first report to an individual entity that merely represents an aggregator (A) or a trusted aggregator (TA) or just anonymizes ($Anon.$) data items.

Privacy notions: Protocols may fulfil different privacy notions w.r.t. the data consumer: A protocol is considered *anonymous*, if the data consumer learns data items, but is unable to identify its origin from its meta-information (e.g. source address etc.). Some protocols only give a plain text *aggregate* to the data consumer. Protocols that employ an aggregating data consumer may exhibit the *aggregator oblivious* [55] property if, loosely speaking, the data consumer only learns the aggregate and no individual data items. A protocol yields an (computationally) *differentially private ((C)DP)* aggregate if the aggregate is the result of a differentially private [17] function over the input data items. Some protocol introduce their own, *proprietary* ad-hoc notion of privacy.

Aggregate error: A protocol's resulting aggregate can either be an *exact* or a *noisy* aggregate with error and is either the result of data producer failures or of measures

to achieve a specific privacy notion. A *differentially private* aggregate is *noisy* by definition.

Group management: While all surveyed protocols require a setup-phase, the introduction or removal of data producers might impose additional effort on all participants: Some protocols require another *complete* setup phase while others merely require an *partial* setup, e.g. only distributing/removing keys of added/removed data producers.

We recognise that ciphertext expansion and aggregate error are important criteria as well. However, because of the heterogeneity of employed encryption schemes and supported aggregate functions a meaningful comparison is unfeasible.

The remainder of this Section gives a structured overview over solutions that have been proposed to the problem of privacy-preserving aggregation so far. Some of the surveyed protocols were specifically designed for the purpose of calculating aggregates in the context of smart metering while others provide solutions to a generic privacy-preserving data item aggregation model. For better readability and comparability we renamed the entity names introduced by the protocols to match the roles of our ideal scenario: data producer, aggregator, data consumer.

The solutions are ordered according to their employed PET.

Anonymization: The solution described by Efthymiou *et al.* [19] differentiates between two types of data required by (different) data consumers: Anonymized high-frequency data for the immediate usage (e.g. network operations) and low-frequency attributable data for billing or similar purposes. Their solution assumes that low-frequency data is not privacy-invading and thus they do not provide a privacy-preserving protocol for billing. Their combined solution introduces two different identities that *asynchronous* data producer use to send the different kinds of data to the respective data consumer(s): The low-frequency identity is attributable by the data consumer to a data producer while the connection between the anonymous high-frequency identity and the low frequency identity is only known to an escrow service. After a data consumer receives high-/low-frequency data items from data producers over *unidirectional* communication links it can calculate *exact arbitrary aggregates* over *arbitrary subsets* of data items even in the presence of *unlimited failures* of data producers. Data producers that are added to the protocol require the setup of the two identities (*partial* group-management). Furthermore, their paper also present protocols for distributing identities to data producers and temporarily lifting privacy in case of energy theft.

Molina-Markham *et al.* [43] envision a system where *synchronous* data producers send data items without identifying information to so-called neighbourhood gateways. The neighbourhood gateways forward the tuples to the aggregator and data consumer, therefore hiding their origin in the communication network. The paper does not explicitly say whether the neighbourhood gateways introduce pseudonyms for data producers so that the aggregator can calculate sensible temporal aggregates. This only requires minimal changes to the protocol and thus we assume that the data consumer can calculate *exact arbitrary aggregates* over *arbitrary subsets* of data items. This protocol only requires *unidirectional* communication links and tolerates *unlimited failures* of data producers. This protocol only requires *partial* group-management. Furthermore, the authors also indicate how the system could facilitate billing (see Section VI).

Bohli *et al.* [10] present a model for smart meter privacy based on a "smart meter privacy break indistinguishability game" and two solutions to calculate aggregates. The first solution employs anonymization via a TTP to which *asynchronous* data producers report their values over an *unidirectional* communication link. The TTP performs an *arbitrary aggregation* over *arbitrary subsets* of data and forwards the *exact* aggregated values to the data consumer. This solution tolerates *unlimited failures* of data producers and yields an *exact* aggregate.

Their second solution employs perturbation. Every *asynchronous* data producer adds random noise to its readings and sends them over an *unidirectional* communication link to the aggregator. The idea is that every data producer adds enough noise to mask its own values, but that the randomness over all data producers cancels each other out during *sum* aggregation of *arbitrary subsets*. This results in an almost accurate aggregate (in our terms it is *noisy*). The authors note that according to their privacy model the TTP solution provides perfect privacy (*proprietary*) unlike the second solution if high accuracy in the aggregates is required (confidence interval with 0.5Wh width and 99.9% certainty). Both approaches tolerate *unlimited failures* of data producers. Also, both merely require *partial* group-management, added data producers only need to know about the trusted third party and employed random distributions in the first/second approach respectively.

Jeske [32] describes a protocol that combines anonymous reporting and non-anonymous (and non-privacy-preserving) billing of smart metering data items. In every run of the billing protocol with the provider (aggregator) the data producer receives a Camenish Lysyanskaya [11] signature on its committed secret ticket. Later, in the reporting protocol, the data producer performs a signature proof of knowledge for this ticket towards the aggregator and also receives the next ticket for the next reporting. The signature proof of knowledge prevents that the aggregator connects a run of the non-anonymous billing protocol with the subsequent anonymous reporting protocol run. The data producers operate *asynchronously* and use *bi-directional* communication links with the aggregator. The aggregator can aggregate over *arbitrary subsets* of data items and calculate *arbitrary* and

exact aggregates. The protocol tolerates *unlimited failures* of data producers and requires *partial* group-management.

Petrlic [46] proposes to use the grid operator as an anonymizer in a reporting protocol. A trusted third party gives data producers pseudonyms and respective pseudonym certificates. *Asynchronous* data producers send their pseudonymously signed data items to the grid operator. The grid operator verifies the signature (and the TTP certificate) and removes both from the message. Anonymized data items are then forwarded to the aggregator (the data consumer). After the setup-phase, data producers merely require *unidirectional* communication links to their grid operators. Data items that finally arrive at the aggregator enable it to calculate *arbitrary*, *exact* aggregates over *arbitrary subsets*. The protocol only requires *partial* group-management and tolerates *unlimited* data producer failures.

Trusted computation: Ruj *et al.* [53] propose a two-tier system for aggregation of smart metering data and subsequent access by authorised consumers. First, data producers encrypt their measurements using Paillier encryption with the public key of their nearest trusted party and send them over *unidirectional* communication links towards that trusted party at the root of their aggregation tree. Every measurement is also attributed by its creator specifying its nature. Along every node of this aggregation tree, encrypted measurements of the same attribute (supports *temporal* and *spatial subsets*) are homomorphically aggregated (*weighted sum*) to an *exact* aggregate. The trusted party decrypts the aggregated measurements for every attribute and encrypts it with an attribute-based-encryption scheme for encrypted central storage. Multiple key distribution centres distribute public Paillier keys to data producers, private Paillier keys and private attribute-encryption keys and policies to trusted parties and attribute-decryption keys to data consumers. Only consumers with respective keys can decrypt storage contents designated for them. The protocol tolerates *unlimited failures* of data producers and tolerates failures of all key distribution centres after the setup phase. Failure of trusted parties are not tolerated. Addition/removal of data producers/data consumers merely requires *partial* group-management.

Li *et al.* [38] propose aggregation of Paillier-encrypted data items during routing through a minimal-spanning-tree of data producers towards a data consumer. The data consumer starts the creation of a spanning tree among all data producers and also issues the aggregation plans that tell data producers (inner tree nodes) how to aggregate collected data items. Every data producer encrypts its data items using Paillier encryption. Data producer at the leaves of the spanning-tree start sending their encrypted data items up towards the data consumer at the root of the tree. At every data producer that represents an inner-node of the tree previously received aggregation plans are executed on encrypted data items/aggregates from children and their own encrypted data items. The aggregation supports *exact*, *weighted sum* aggregation of *arbitrary subsets* of data items, provided that they are available at aggregation points. This protocol supports *unlimited failures* of data producers. Addition/removal of data producers requires change of the spanning tree and therefore *partial* group-management. The creation of the spanning tree requires *bi-directional* communication links between data producers.

Cryptographic computation: Rottondi *et al.* [51] describe an approach where trusted privacy-preserving nodes and a central configurator are introduced into the smart metering system. *Synchronous* data producers report Shamir secret shares of their measurements over *unidirectional* communication links to privacy-preserving nodes specified by the configurator. The configurator also tells the privacy-preserving nodes how to aggregate incoming measurements and for which consumer. According to these rules privacy-preserving nodes aggregate the shares to shares of the final aggregate. The consumer queries all privacy-preserving nodes that were dispatched by the configurator for the shares of the final aggregate and computes the aggregate from its shares. This use of the Shamir secret sharing scheme enables the computation of *exact*, *weighted sum* aggregates over *arbitrary subsets* of data items. Addition/removal of data producers requires informing them about the specific parameters for the Shamir scheme (*partial* group management).

Kursawe *et al.* [36] propose different protocols for private aggregation or comparison of data item aggregates. Their aggregation protocols allow the *exact* calculation of *sum* aggregates over *spatial* data sets of *synchronous* data producers at the data consumer. The comparison protocols only allow the data consumer to compare a known value to the aggregate or use brute-force to compare the aggregate to neighbouring values of the known value. Four concrete protocols use different approaches to privately compute zero-sum shares among each other to blind their measurements before they are aggregated at the data consumer. All four protocols require *unidirectional* communication links and *complete* group-management as all data producers need to know keys of all other data producers. Also, all four protocols tolerate *no failures* as missing shares of the zero-sum shares would prevent the correct assembly.

Erkin *et al.* [21] describe a system where (*synchronous*) data producers encrypt their data items before sending them over an *unidirectional* communication link to the data consumer. The data producers use Paillier encryption in a special way: The second random parameter to the Paillier encryption is exponentiated with a share of a multiple of the modulus. Thus, only if all ciphertexts from all data producers are homomorphically aggregated the resulting ciphertext has a random parameter that has been exponentiated with a multiple of the modulus. Only then it can be decrypted by the data consumer with the private key. With additional (described in their paper) compensating measures this pro-

tocol allows for *exact*, *weighted sums* of *arbitrary subsets* of data items. *Unlimited failures* of data producer can be compensated with the help of a trusted third party i.e. the manufacturer of the smart meter. Addition/removal of data producers requires re-establishment of the shares and thus a *complete* group-management.

Garcia and Jacobs [23] propose an aggregation protocol that specifically targets the detection of energy theft at substations. *Synchronous* data producers connected to one substation with an *unidirectional* communication link create shares of each measured data item and encrypt all but one share with the public key of another connected data producer. Then they forward these shares over the substation to those respective data producer that can decrypt them. Upon receipt of all shares from all other data producers a data producer sums them up, together with the share it did not send out, and returns the result to the substation. The substation adds all results and determines whether the total sum is different to what it physically measured itself for the set of data producers. This protocol allows for *exact*, *sum* of *spatial subsets* of data items. *No failures* are tolerated as missing shares would not allow for correct summation. Addition/removal of data producer requires distribution of public keys and thus *partial* group-management.

Perturbation: Lin *et al.* [39] describe a system where *asynchronous* data producers continuously create, and additively blind, measurements and store them in a central storage system. Authorised data consumers can access the central storage system and create *exact*, *weighted sum* over *temporal* subsets or *noisy*, *arbitrary aggregates* of *spatial subsets* of blinded measurements. However, subsequently the data consumer has to contact all respective data producers to obtain the blinding factors for the used data items. Thus, this protocol requires *bi-directional* communication links between data producer and data consumers. This approach tolerates *no failures* of data producer and only requires *partial* group-management.

Differentially private perturbation: Shi *et al.* [55] first define aggregator obliviousness by the means of an indistinguishability game. Subsequently, the authors present an aggregation protocol in which *synchronous*, individual data producers add random noise from a geometric distribution in such a way, that the sum of the random noise of all data producers will guarantee differential privacy for the aggregate result. Every data producer encrypts its noisy measurement with an individual share of zero in a Diffie-Hellman based encryption scheme before sending it over an *unidirectional* communication link to the aggregator. The aggregator owns a final share that eventually allows him to decrypt (brute force/Pollard's lambda method) the *differential private* aggregate of all measurements. This approach supports *noisy*, *differentially private sum* aggregates over *spatial subsets* of data items. However, because of the use of the secret sharing of zero, this protocol tolerates *no failures* and *complete* group-management.

Rastogi *et al.* [48] argue that time series data might be highly correlated and that therefore the sensitivity of aggregates over these data is higher and thus bigger noise is required to guarantee differential privacy. In their proposed protocol, individual data producers calculate aggregates over *temporal* subsets of their data items and then calculate the discrete Fourier transform of the aggregate. The discrete Fourier transform requires the addition of noise with lower variance than the original aggregate result. After a cut off of high frequencies they add Gaussian noise and a random blinding factor to the result and send it Paillier-encrypted to the aggregator. The aggregator homomorphically aggregates all data producer's aggregates and returns them to the data producers for removal of the random blinding factor. Every data producer removes his previously added blinding factor and creates a decryption share. Finally, the aggregator receives all decryption shares and can recover a differentially private discrete Fourier transform which he inversely discrete Fourier transforms to obtain the final result. This protocol supports the calculation of *sum* aggregates of *arbitrary, temporal aggregates* and requires *synchronous* data producers for the decryption phase. It tolerates *limited failures* of data producers and requires *complete* group-management because of its use of decryption shares across data producers.

In Chan *et al.* [12] a protocol is proposed, where a system of intersecting user groups allows the aggregator to compensate for *unlimited failures* of data producers with the help of redundant information. In a setup phase all data producers are organised as the leaves of a binary interval tree. Every node in the tree represents an aggregate of its descendant's data items. Therefore every data producer is member of several such aggregate blocks. The authors propose to run the aggregation algorithm from Shi [55] among members of the same block. In the case of data producer failure and subsequent inability to complete the block aggregation, all other members of that block might still contribute their data items to other aggregates of blocks where they are members, too. These other block aggregates then enable the calculation of overall aggregates of live data producers despite of some failed data producers and thus some incomplete block aggregates. The protocol requires *partial* group-management: Upon addition of new data producers the tree can either be enlarged and only newly created blocks require a new setup or additional data producers are organised in an additional tree altogether. The usage of Shi [55] as block aggregation protocol supports *differentially private sum* aggregates over *spatial subsets* of data items but also requires *synchronous* data producers yet only *unidirectional* communication links. Their paper also describes two sampling protocols that save bandwidth, because the data consumer calculates the aggregate as extrapolation of the sampled values.

Jawurek and Kerschbaum [31] propose a protocol in which all *asynchronous* data producers encrypt their data

items using Paillier encryption and the public key of a key-managing authority. They send their data items over an *unidirectional* communication link to the data consumer. The data consumer collects all encrypted data items and selects *arbitrary subsets* for the homomorphic calculation of *weighted sums*. In order to obtain the *differentially private* plaintext of the aggregate, the data consumer contacts distributed key-managing authority instances for decryption. A zero-knowledge proof between the data consumer and the key-managing authority instances allows the key-managing authority instances to know that the data consumer only used data items in their logical order (to prevent re-use) and a specified aggregation function. Therefore, it can apply the appropriate amount of random noise to the decrypted aggregate to guarantee differential privacy. Furthermore, the protocol employs a threshold secret-sharing that enables key-managing authorities not to require synchronisation of state. This prevents malicious data consumers that attempt to induce different states at instances. Once data producers have contributed their encrypted data items their failure does not impact the protocol and therefore it tolerates *unlimited failures* of data producers. The distribution of key-managing authorities makes the protocol also resilient to failures of key-managing authority instances. As all data producers operate independently of each other the addition/removal of data producer requires *partial* group-management.

Acs *et al.* [3] propose a system where *synchronous* data producers form clusters among each others. All pairs of data producers in every cluster create pairwise keys (by exchange over *bi-directional* communication links across the data consumer) that they add to their measurements and that cancel each other out during additive modular aggregation at the data consumer. Individual data producers also add random noise according to Gamma distributions which yields Laplacian random noise in the aggregate and thus makes the calculated *sums* over *spatial subsets* of data items *differentially private*. This protocol tolerates *limited failures* of data producers, as the data consumer can query the remaining data producers for their part of the respective pairwise keys for recovery of the aggregate. Addition/removal of data producers requires all data producers to re-create their pairwise cluster keys and thus requires *complete* group-management.

VI. Billing & other household computations

As introduced in Section II-A privacy-preserving protocols have been specially designed for the billing on smart meter data: The calculation of a bill over one household's smart meter data according to this household's tariff.

An irrevocable objective of bill computation is correctness, i.e. correct computation on integer tariff and authentic smart meter data. Furthermore, the result of a billing computation has to be attributable to a specific household. Thus, privacy-preserving billing protocols, in contrast to general aggregation protocols surveyed in Section V, can only employ two out of all the aforementioned (see Section IV) PETs to implement the ideal's protocol functionality in a privacy-preserving manner: *Verifiable computation* and *trusted computation*. *Anonymization* is not applicable, because in the end, the supplier has to send an electricity bill to someone. *Perturbation* is generally not applicable, because the bill is usually expected, by consumers or legislation, to be exact.

In the following we characterise protocols by two criteria: Their employed PET and, if they implement *verifiable computation*, the tariffs they support. Tariff support is dependant on the (non-) interactive proof they rely upon. Protocols employing the *trusted computation* strategy support arbitrary tariffs. That means, they may be limited by the computing power of the aggregator and the available time but not by the types of calculations.

In the following we describe the tariff types that surveyed protocols introduce. We assume that a smart meter records a consumption value for every atomic time slot and that time slots can fall into different contexts, e.g. time-of-day. Then the price for consumption of a specific time slot is calculated as follows:

Linear tariff: A consumption value is priced linearly with a unit price. The unit price may be context-dependent.

Interval linear tariff: As extension to the *linear tariff*, the consumption values domain is divided into intervals while each interval indicates a unit price. Depending on which interval a consumption value falls into, it is priced at the unit price indicated by that interval. Intervals and unit price may be context-dependent.

Cumulative tariff: The consumption values domain is divided into intervals. In contrast to *interval linear tariff*, if a consumption value spans several intervals, then every part of that consumption value is priced linearly at the unit price of the interval it falls into. The total price is then the sum over intervals. Intervals and unit price may be context-dependent.

Cumulative polynomial tariff: As extension to the *cumulative tariff*, the individual consumption value is priced polynomially for every interval. Polynomials and intervals may be context-dependent.

Arbitrary tariff: Arbitrary computations.

These are the surveyed billing protocols:

Verifiable computation: Molina-Markham *et al.* [43] proposed in 2012, alongside their aggregation protocol, the idea that "generic ZKPs" could be used in the consumer device to prove honest calculation of the bill. Yet no specific construction was presented.

Jawurek *et al.* [29] propose the introduction of a component between the smart meter and the data consumer to perform the bill calculation and proof generation. It requires

[1] Allows for sums over arbitrary, temporal aggregates

Table II
CHARACTERISTICS OF AGGREGATION PROTOCOLS ACCORDING TO DIMENSIONS FROM SECTION V.

Paper	Agg. Function	Subset	Sync.	Fault-tolerance	Comm. Model	Privacy	Error	Group-Mgmt.
Efthymiou *et al.*[19]	Arb.	S/T		Unl.	P→Anon.→C	Anon.	Exact	P
Molina-Markham *et al.*[43]	Arb.	S		Unl.	P→Anon.→C	Anon.	Exact	P
Bohli *et al.*[10]	Arb.	S/T		Unl.	P→TA→C	Aggr.	Exact	P
	Sum	S/T		Unl.	P→AC	Prop.	Noisy	P
Jeske [32]	Arb.	S/T		Unl.	P↔C	Anon.	Exact	P
Petrlic [46]	Arb.	Arb.		Unl.	P→Anon.→C	Anon.	Exact	P
Ruj *et al.*[53]	W. sum	S/T		Unl.	P→TA→C	Agg.	Exact	P
Li *et al.*[38]	W. sum	Arb.		Unl.	P↔P↔C	Agg.	Exact	C
Rottondi *et al.*[51]	Sum	S/T	Y	Lim.	P→A→C	Agg.	Exact	P
Kursawe *et al.*[36]	Sum	S	Y	No	P↔AC	AO	Exact	C
Erkin *et al.*[21]	Sum	Arb.	Y	Unl.	P↔P→C	AO	Exact	C
Garcia *et al.*[23]	Sum	S	Y	No	P↔AC	AO	Exact	P
Lin *et al.*[39]	W. sum	S/T		No.	P↔AC	AO(T),prop.(S)	Noisy(S)	P
Rastogi *et al.*[48]	[1]	T	Y	No	P↔AC	CDP,AO	Noisy	C
Shi *et al.*[55]	Sum	S	Y	No	P→AC	CDP,AO	Noisy	C
Chan *et al.*[12]	Sum	S	Y	Unl.	P→ AC	CDP,AO	Noisy	P
Jawurek *et al.*[31]	W. sum	Arb.		Unl.	P→AC↔T	DP,AO	Noisy	P
Acs *et al.*[3]	Sum	S	Y	Lim.	P↔AC	DP,AO	Noisy	P

software changes to the smart meter, i.e. the creation and cryptographic signing of Pedersen commitments. The plug-in component intercepts all communication coming from the smart meter and only forwards signed commitments. Then, it obtains the current tariff from the software provider to calculate the bill. The proof of correct calculation consists of the final price, all signed commitments from the smart meter and an aggregation of all random parameters that were used by the smart meter for commitment creation. Due to the homomorphic property the data consumer can analogously compute the bill calculation on the signed commitments to verify that the resulting commitment is a commitment to the received final price and aggregated random parameters. Thus it verifies correct computation with the correct tariff and authentic smart meter measurements. Due to the limitations of the semi-homomorphic Pedersen commitments this protocol supports the calculation of *linear tariffs*.

Rial *et al.* [50] use a similar system model as Jawurek *et al.* [29]. A smart meter creates measurements, commitments on these measurements and signatures over the commitments and sends them to a user device. The user device obtains the tariff from the data consumer and calculates the fee and the required proof. In their approach, the user uses Σ-protocols for the implementation of ZKP for various tariffs: *cumulative tariffs*, *interval linear tariffs* and even *cumulative polynomial tariffs*. The authors describe one particularly efficient proof variant for *linear tariffs* that relies on the additive homomorphic commitments as in [29].

Molina-Markham *et al.* [42] implement the cryptographic primitives from [29, 50, 36] for low-cost micro controllers to analyse their run times, memory requirements, energy consumption and economic feasibility. They conclude that elliptic curve variants of these protocols are practical even for very cheap devices in the field.

Danezis *et al.* [14] build upon the work of Rial *et al.* [50] and describe a protocol that limits the information leaked from the final billed amount. Depending on the actual consumer behaviour, the final price allows to infer some information about the behaviour. First, the paper describes the concept of differential privacy [17] and develops a variant that allows them to make a pricing mechanism differentially private for different "privacy units" (i.e. time intervals that should be masked) by only adding positive random noise to the actual final price. Using this variant of differential privacy they extend their protocols from [50] in such a way, that the user pays a differentially private fee for a chosen "privacy unit". Furthermore, the protocol guarantees the provider, that at any time, the amount paid so far exceeds the actual total fees. The user, on the other hand, can use previously paid additional random noise for future payments as long as that guarantee is not violated.

Trusted computation: Lemay *et al.* [37] propose a software architecture consisting of several virtual machines running on one smart meter. They are supervised and separated from each other by one hypervisor that also controls their access to the smart meter's metric hardware as well as their network connections. With the help of a TPM (Trusted Platform Module) the architecture supports remote attestation to verify the integrity of the system and its computations. The authors mention local computation of bills as one virtual machine among others responsible for applications like demand response or a consumer portal.

Both, Jeske [32] and Petrlic [46] propose similar approaches, that also make use of TPMs and software attestation to verify the integrity of the smart meter and therefore the trustworthiness of the local calculation.

VII. Battery Approaches

Battery based privacy technologies attempt to mask the amount of energy consumed, even from adversaries that have physical control of the smart meter, or a direct way to measure the energy input to the home. This is achieved by using a re-chargeable battery to buffer and smoother consumption from the utility.

This approach was first proposed in Kalogridis *et al.* [33]. The authors describe an algorithm to determine the charge and discharge keeping consumption steady. The consumption needs to be adjusted when the battery nears full or empty charge. It is evaluated using real and simulated datasets under realistic battery capacity constraints. Privacy is measured using three (non-standard) metrics based on entropy, clustering and regression. They conclude that load hiding for their real data sets using one battery is feasible.

The proposed algorithm is extended in Varodayan *et al.* [57]. The authors propose randomising the choices of the algorithm instead of trying to achieve a steady load. They show that this improves privacy under the (non-standard) metric of mutual information rate. No evaluation of feasibility is conducted.

An almost identical proposal to Kalogridis *et al.* [33] has been later made by McLaughlin *et al.* [41]. The evaluation uses real data and measures privacy against practical NALM algorithms. Furthermore they evaluate the (non-standard) metric of mutual entropy between unbuffered and buffered consumption. They conclude that almost perfect privacy against NALM is feasible using a set of 10 real batteries.

The battery approach is also used by Acs *et al.* [4]: differential privacy is applied, such that it is hard to distinguish whether a particular power consumer was on during a time interval. This has the positive side effect that the adversary can be even capable of introducing additional power consumers into the household without compromising privacy. They present three different algorithms to determine the (randomized) power consumption from the utility provider. Evaluation is theoretical and the authors conclude that very large, impractical battery capacities are necessary for differential privacy.

This approach has been slightly extended in Backes *et al.* [7] by adding a battery recharging consumption to the overall load. Thus enabling recharging while still providing differential privacy. They conclude that their approach is practical for limited cases of simulated examples using real battery capacities.

VIII. Summary

Smart meter readings are used for a variety of functions, that need to be performed in a timely, reliable, and correct way. Yet, the readings are considered private customer data in most jurisdictions. This is justified by research that demonstrates which information about households can be inferred from readings.

Currently, privacy protection mostly relies on policy, regulation, organizational and legal measures. These lack the strength of mechanism required to ensure that malicious insiders or outsiders do not get access to the data. In contrast we have described three families of privacy technologies that have been proposed specifically for use in smart grids: private aggregation, private billing and battery based approaches. They include multiple trade-off points in terms of functionality, efficiency, ease of deployment and scalability.

The April 2012 DECC consultation on privacy tentatively concluded that "*whilst there were several potential technologies under development that could facilitate data minimisation, aggregation or anonymisation, none had been proven on a commercial scale in the UK market and all required further work*". Interestingly, out of the 22 distinct privacy technologies we surveyed (fig. I) about half originated in part from industrial research organizations [2] , yet *none* of the authors were directly employed by the energy industry. Given the number of options presented in this survey we would argue that the onus is now on the energy industries to take ownership of the privacy implications of the infrastructure they are deploying, invest in privacy research and experiment with deployments of the privacy technologies that have been proposed.

References

[1] Future of privacy summary of california public utilities commission proposed decision on smart grid privacy and security. On-line http://www.futureofprivacy.org/2011/05/09/future-of-privacy-summary-of-california-public-utilities-commission-proposed-decision-on-smart-grid-privacy-and-security/, May 9 2011.

[2] Opinion 12/2011 on smart metering. Article 29 Decision, April 4 2011.

[3] G. Acs and C. Castelluccia. I have a dream!(differentially private smart metering). In *Information Hiding*, pages 118–132. Springer, 2011.

[4] G. Acs, C. Castelluccia, and W. Lecat. Protecting against physical resource monitoring. In *Proceedings of the 10th annual ACM workshop on Privacy in the electronic society*, pages 23–32. ACM, 2011.

[5] B. S. Amador. The federal republic of germany and left wing terrorism. Master's thesis, Naval Postgraduate School, Monterey, CA, December 2003.

[6] R. Anderson and S. Bezuidenhoudt. On the reliability of electronic payment systems. *Software Engineering, IEEE Transactions on*, 22(5):294–301, 1996.

[7] M. Backes and S. Meiser. Differentially private smart metering with battery recharging. *IACR Cryptology ePrint Archive*, 2012:183, 2012.

[8] F. Bao, P. Samarati, and J. Zhou, editors. *Applied Cryptography and Network Security - 10th International Conference, ACNS 2012, Singapore, June 26-29, 2012. Proceedings*, volume 7341 of *Lecture Notes in Computer Science*. Springer, 2012.

[9] G. Bauer, K. Stockinger, and P. Lukowicz. Recognizing the use-mode of kitchen appliances from their current consumption. In *Proceedings of the 4th European conference on Smart sensing and context*, EuroSSC'09, pages 163–176, Berlin, Heidelberg, 2009. Springer-Verlag.

[10] J.-M. Bohli, O. Ugus, and C. Sorge. A privacy model for smart metering. In *Proceedings of the First IEEE International Workshop on Smart Grid Communications (in conjunction with IEEE ICC 2010)*, 2010.

[2] FxPal, Microsoft, NEC, PARC, SAP and Toshiba.

[11] J. Camenisch and A. Lysyanskaya. A signature scheme with efficient protocols. In S. Cimato, C. Galdi, and G. Persiano, editors, *SCN*, volume 2576 of *Lecture Notes in Computer Science*, pages 268–289. Springer, 2002.

[12] T.-H. H. Chan, E. Shi, and D. Song. Privacy-preserving stream aggregation with fault tolerance. In *Proceedings of the 16th International Conference on Financial Cryptography and Data Security*, FC '12, 2012.

[13] F. T. Commission. Fair information practice principles. On-line http://www.ftc.gov/reports/privacy3/fairinfo.shtm, June 25 2007.

[14] G. Danezis, M. Kohlweiss, and A. Rial. Differentially private billing with rebates. In W.-Y. Ma, J.-Y. Nie, R. A. Baeza-Yates, T.-S. Chua, and W. B. Croft, editors, *Information Hiding*, pages 148–162. ACM, 2011.

[15] Smart metering implementation programme data access and privacy consultation document. United Kingdom Department of Energy and Climate Change, Consulation Document, April 2012.

[16] S. Depuru, L. Wang, and V. Devabhaktuni. Electricity theft: Overview, issues, prevention and a smart meter based approach to control theft. *Energy Policy*, 39(2):1007–1015, 2011.

[17] C. Dwork. Differential privacy. In *ICALP*, pages 1–12. Springer, 2006.

[18] C. Efthymiou and G. Kalogridis. Smart grid privacy via anonymization of smart metering data. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 238 –243, oct. 2010.

[19] C. Efthymiou and G. Kalogridis. Smart grid privacy via anonymization of smart metering data. *2010 First IEEE International Conference on Smart Grid Communications*, pages 238–243, 2010.

[20] M. Enev, S. Gupta, T. Kohno, and S. Patel. Televisions, video privacy, and powerline electromagnetic interference. In *Proceedings of the 18th ACM conference on Computer and communications security*, pages 537–550. ACM, 2011.

[21] Z. Erkin and G. Tsudik. Private computation of spatial and temporal power consumption with smart meters. In Bao et al. [8], pages 561–577.

[22] J. E. Froehlich, E. Larson, T. Campbell, C. Haggerty, J. Fogarty, and S. N. Patel. Hydrosense: infrastructure-mediated single-point sensing of whole-home water activity. In *Proceedings of the 11th international conference on Ubiquitous computing*, Ubicomp '09, pages 235–244, New York, NY, USA, 2009. ACM.

[23] F. Garcia and B. Jacobs. Privacy-friendly energy-metering via homomorphic encryption. In *Proceedings of the 6th International Workshop on Security and Trust Management*, 2010.

[24] U. Greveler, B. Justus, and D. Loehr. Multimedia content identification through smart meter power usage profiles.

[25] G. Hart. http://www.georgehart.com/research/nalmrefs.html.

[26] G. Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870 –1891, dec 1992.

[27] G. W. Hart. Residential energy monitoring and computerized surveillance via utility power flows. *IEEE Technology and Society Magazine*, June 1989.

[28] Operationalizing privacy by design: The ontario smart grid case study. Office of the Information & Privacy Commissioner of Ontario, February 2 2011.

[29] M. Jawurek, M. Johns, and F. Kerschbaum. Plug-in privacy for smart metering billing. In S. Fischer-Hübner and N. Hopper, editors, *PETS*, volume 6794 of *Lecture Notes in Computer Science*, pages 192–210. Springer, 2011.

[30] M. Jawurek, M. Johns, and K. Rieck. Smart metering depseudonymization. In *ACSAC*, pages 227–236, 2011.

[31] M. Jawurek and F. Kerschbaum. Fault-tolerant privacy-preserving statistics. In *accepted at PETS*, 2012.

[32] T. Jeske. Privacy-preserving smart metering without a trusted-third-party. In J. Lopez and P. Samarati, editors, *SECRYPT*, pages 114–123. SciTePress, 2011.

[33] G. Kalogridis, C. Efthymiou, S. Z. Denic, T. A. Lewis, and R. Cepeda. Privacy for smart meters: Towards undetectable appliance load signatures. In *2010 First IEEE International Conference on Smart Grid Communications*, pages 232–237. IEEE, 2010.

[34] Y. Kim, T. Schmid, M. B. Srivastava, and Y. Wang. Challenges in resource monitoring for residential spaces. In *Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, BuildSys '09, pages 1–6, New York, NY, USA, 2009. ACM.

[35] R. Knyrim and G. Trieb. Smart metering under eu data protection law. *International Data Privacy Law*, 1(2):121–128, 2011.

[36] K. Kursawe, G. Danezis, and M. Kohlweiss. Privacy-friendly aggregation for the smart-grid. In *PETS*, pages 175–191, 2011.

[37] M. Lemay, G. Gross, C. A. Gunter, and S. Garg. Unified architecture for large-scale attested metering. In *in Hawaii International Conference on System Sciences. Big Island*. ACM, 2007.

[38] F. Li, B. Luo, and P. Liu. Secure information aggregation for smart grids using homomorphic encryption. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 327 –332, oct. 2010.

[39] H.-Y. Lin, W.-G. Tzeng, S.-T. Shen, and B.-S. P. Lin. A practical smart metering system supporting privacy preserving billing and load monitoring. In Bao et al. [8], pages 544–560.

[40] M. A. Lisovich, D. K. Mulligan, and S. B. Wicker. Inferring personal information from demand-response systems. *IEEE Security & Privacy*, 8(1):11–20, January 2010.

[41] S. McLaughlin, P. McDaniel, and W. Aiello. Protecting consumer privacy from electric load monitoring. In *Proceedings of the 18th ACM conference on Computer and communications security*, pages 87–98. ACM, 2011.

[42] A. Molina-Markham, G. Danezis, K. Fu, P. Shenoy, and D. Irwin. Designing privacy-preserving smart meters with low-cost microcontrollers. In *Proceedings of the 16th International Conference on Financial Cryptography and Data Security*, February 2012.

[43] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin. Private memoirs of a smart meter. In *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, BuildSys '10, New York, NY, USA, 2010. ACM.

[44] A. Narayanan and V. Shmatikov. Robust de-anonymization of large sparse datasets. In *IEEE Symposium on Security and Privacy*, pages 111–125. IEEE Computer Society, 2008.

[45] National Institute of Standards and Technology. NISTIR 7628., August 2010.

[46] R. Petrlic. A privacy-preserving concept for smart grids. In *Sicherheit in vernetzten Systemen: 18. DFN Workshop*, pages B1–B14. Books on Demand GmbH, 2010.

[47] A. Prudenzi. A neuron nets based procedure for identifying domestic appliances pattern-of-use from energy recordings at meter panel. In *Power Engineering Society Winter Meeting, 2002. IEEE*, volume 2, pages 941 – 946 vol.2, 2002.

[48] V. Rastogi and S. Nath. Differentially private aggregation of distributed time-series with transformation and encryption. In *Proceedings of the 2010 international conference on Management of data*, SIGMOD '10, pages 735–746, New York, NY, USA, 2010. ACM.

[49] S. Report. Assessment of demand response and advanced metering. Technical report, Federal Energy Regulatory Commission, November 2011.

[50] A. Rial and G. Danezis. Privacy-preserving smart metering. In *Proceedings of the 10th annual ACM workshop on Privacy in the electronic society*, pages 49–60. ACM, 2011.

[51] C. Rottondi, G. Verticale, and A. Capone. A security framework for smart metering with multiple data consumers.

[52] I. Rouf, H. A. Mustafa, M. Xu, W. Xu, R. D. Miller, and M. Gruteser. Neighborhood watch: security and privacy analysis of automatic meter reading systems. In T. Yu, G. Danezis, and V. D. Gligor, editors, *ACM Conference on Computer and Communications Security*, pages 462–473. ACM, 2012.

[53] S. Ruj, A. Nayak, and I. Stojmenovic. A security architecture for data aggregation and access control in smart grids. *Arxiv preprint arXiv:1111.2619*, 2011.

[54] K. F. S. S. Clark, J. Sorber and E. Learned-Miller. Current events: Compromising web privacy by tapping the electrical outlet. July 2011.

[55] E. Shi, T.-H. H. Chan, E. G. Rieffel, R. Chow, and D. Song. Privacy-preserving aggregation of time-series data. In *NDSS*, 2011.

[56] J. Tsai. Privacy and the smart grid: A policy-making case study. Tprc, Carnegie Mellon University, August 15 2010.

[57] D. Varodayan and A. Khisti. Smart meter privacy using a rechargeable battery: minimizing the rate of information leakage. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 1932–1935. IEEE, 2011.