DOMINIK ENGEL

METHODS FOR USER-CENTRIC SMART GRID PRIVACY, SECURITY AND CONTROL

METHODS FOR USER-CENTRIC SMART GRID PRIVACY, SECURITY AND CONTROL

DOMINIK ENGEL

Habilitationsschrift

zur Erlangung der Venia Docendi
für das Fach Angewandte Informatik
an der Naturwissenschaftlichen Fakultät der
Universität Salzburg

Mai 2017

ABSTRACT

The transition from traditional energy grids to so-called *smart grids* is an enabler of the important societal goal to turn from fossil energy sources to renewable energy sources. The vision is to build intelligent energy grids that harness the insights of information and communication technology (ICT) to allow widespread integration of renewable energy sources, self-healing grids, connection of smart homes and smart electric vehicles, synchronization of demand and response, and many other use cases. Spreading such Smart Grid technologies will be inherently difficult without addressing user concerns and actively managing user acceptance by providing secure methods and demonstrating safety of user data and privacy. Privacy, security, and user control in the smart grid user domain are critical for establishing end user *trust* and enabling end user participation.

This habilitation treatise proposes methods for the smart grid user domain to facilitate end-user acceptance of smart grid technology. The main focus is on methods that safeguard end-user data, both in terms of preserving end-user privacy in face of potentially curios or malicious insiders and providing security to fend off outside attackers. Apart from privacy and security, general methods for data handling, such as compression, are discussed. Furthermore, the important topic of user control is introduced, i.e., methods that aim at a two-fold benefit: incentivizing users to participate in smart grid optimization schemes and allowing informed interaction with the new technology. Both, of course, under the assumption that appropriate methods for preserving privacy and providing security are in place.

Overall, the contribution presented in this treatise aims at enabling user acceptance of smart grid technologies by providing methods for the important factors of privacy, security and user control.

ACKNOWLEDGMENTS

A major part of the work constituting this treatise has been done in the context of the *Josef Ressel Center for User-Centric Smart Grid Privacy, Security and Control.* This research center has not only provided a fruitful environment for my own research, but also, and more importantly, gave the great opportunity to work with a research team to explore in depth many of the different aspects of the field. I am very happy to acknowledge the great impact the collaboration with the Josef Ressel research team has made, as co-authors, colleagues, sparring partners and friends – it has been a pleasure and a privilege.

Thanks are due to the Christian Doppler Research Foundation for funding and powering the research program, to the Austrian Ministry of Science, Research, and Economy and the Federal State of Salzburg for funding, and to the company partners of the center – Salzburg AG, Salzburg Wohnbau and Siemens AG – not only for funding, but also for providing the interesting questions and for collaborating to answer them.

Some of the publications contained in this treatise originated from follow-up projects of the Josef Ressel Research Center – INTEGRA, RASSA, PROMISE. The input of the project partners, not only as co-authors, is gratefully acknowledged, as is the funding of the Austrian Research Promotion Agency FFG.

Some of the publications have been authored with colleagues at the "Visual Computing and Multimedia" research group at the University of Salzburg. I did my PhD in the environment of this research group and want to express my thanks.

Thanks are due to my colleagues at the Salzburg University of Applied Sciences, especially at the degree program "Information Technologies and Systems Management" for support and understanding, especially when close paper deadlines impacted on other important business.

I would like to extend my heartfelt thanks to all of my students at the Universities of Bremen and Salzburg, and at the Salzburg University of Applied Sciences. Teaching has always been a rewarding experience for me and has in many ways provided a sense of grounding. Some of the publications that are contained in this treatise have originated from Master's Theses.

I am grateful to the Marshall Plan Foundation for funding collaboration with colleagues from US universities (Cornell University, Bowling Green State University and University of Southern California), whose input I gratefully acknowledge.

Finally, I want to thank my family for their love and great support, and for providing a pillar that is not connected to research or academia.

CONTENTS

1 INTRODUCTION

The term "smart grids" is used to describe the next-generation energy systems – digitized systems of systems that are an important enabler for turning from fossil energy sources to renewable energy sources. Smart grids employ state-of-the-art information and communication technology to control generation, distribution and consumption of energy. With smart grids the power network organization moves from a hierarchical to a decentralized structure and communication flow moves from largely uni-directional to bi-directional. The degree of information needed on network status is vastly more accurate compared to traditional power networks is made available to different stakeholders in fine granularity, sometimes in near real-time.

One of the most important goals of smart grids is the accommodation of environmentally sustainable energy sources. Traditional, non-renewable energy sources can be controlled with a hierarchical network structure, with energy sources and energy sinks at opposing levels of the hierarchy. Many types of renewable energy sources, such as photovoltaics or wind power, generate power at the distribution level. In order to integrate a high number of these renewable sources to produce in bulk quantities an evolution of the network infrastructure and is necessary.

A significant portion of (potential) end-users at this point in time are wary about possible disadvantages of the new smart grid technology, like the uncertainty regarding the level of privacy and possible security breaches [20, 17]. Apart from privacy and security concerns, end-users are skeptical regarding possible benefits of smart grids as a new technology, such as cost savings [17]. End-users have difficulties understanding the level of control they can exert in a smart grid environment [53].

It is valid to diagnose a severe lack of trust towards smart grids on the end-user level. If this lack of trust were to persist, it would prevent many important features of the smart grid: intelligent power pricing plans, distributed energy management, adaptive load balancing, e-mobility, private renewable energy sources and the usage of smart grid infrastructure for other areas, such as home automation.

In order to establish the needed degree of trust in the end-user domain, providing a visible level of both, security and privacy, are imperative. In addition, a further component is necessary, namely: user control. In order to alleviate concerns of a lack in benefits (especially on a personal level) a sufficient degree of understanding of possible interaction is required. Correspondingly, user interaction needs to be reflected accurately on the system side.

Users need to be informed of what choices can be made (user information), how these choices influence smart grid processes (functional transparency), what data items at what granularity need to be disclosed for this purpose (user-managed privacy) and that processes and data transfer are operated in a secure way (traceable security). On all items, systemic feedback to the user needs to be provided consistently in order to raise user awareness and ultimately create a level of user trust that allows meaningful interaction.

The goal of the methods proposed in this treatise is enabling trust in the smart grid end-user domain, by improving existing approaches and develop new mechanisms to establish secure, private and informed – i.e., trusted – user interaction with smart grid technology. The fundamental question addressed is: How can a sustainable level of trust be established in smart grid user domain entities (applications, processes, data protocols, and interfaces) in order to invite and further user participation in smart grid technologies?

The main focus is on methods that safeguard end-user data, both in terms of preserving end-user privacy in face of potentially curios or malicious insiders and providing security to fend off outside attackers. Apart from privacy and security, general methods for data handling, such as compression, are discussed. Furthermore, the important topic of user control is introduced, i.e., methods that aim at a two-fold benefit: incentivizing users to participate in smart grid optimization schemes and allowing informed interaction with the new technology. Both, of course, under the assumption that appropriate methods for preserving privacy and providing security are in place.

Overall, the contribution presented in this treatise aims at enabling user acceptance of smart grid technologies by providing methods for the important factors of privacy, security and user control.

The rest of this treatise is organized as follows: The present chapter introduces the topic and the state of the art in Section 1.1 and puts the general contribution of this habilitation treatise in the context of related work. Section 1.2 details the contributions of the individual papers that cumulatively constitute this treatise. In Chapter 2 the fulfillment of the requirements for a habilitation endeavour are discussed: Percentage of author contribution, teaching, given talks and further aspects for review, such as personal impact factors. Finally, in Chapter 3, the individual publication are printed in their original form.

1.1 BACKGROUND

The move towards smart grids has spawned a large number of industry initiatives, research programmes and standardization efforts, see, e.g., [41] for an overview. Many of the earlier contributions focused on the smart grid ecosystem at a larger scale, without exploring in detail the ramifications of the move towards smart grid technology for the end-user.

More recent programmes increasingly accommodate the user perspective, cf., [41, 42, 102, 78, 55, 23, 53, 76]. Addressing the topic of user acceptance is pointed out as a key issue by almost all authors.

In the following, an overview of what is defined as the smart grid user domain is given, followed by an overview of current contributions in the fields of privacy, security, and user control.

1.1.1 *Smart Grid User Domain and Smart Metering*

Smart grid communication networks have to support a wide range of applications like Advanced Metering Infrastructure, Automated Supply and Demand Response, Feeder Automation, Mobile Workflow Management, etc. [88]. In [101] a communication oriented framework for smart grid is introduced, consisting of three entities, namely, Operation, Business and Consumer Network. A more application-oriented approach especially for the Smart Metering Infrastructure distinguishes between Home Area Network (HAN), Neighbourhood Area Network (NAN) and Wide Area Network (WAN) (e.g., [41]).

In terms of use cases, this treatise will focus on *smart metering*, which is a critical part of smart grids and the area, on which most of the privacy debate centered. Smart meters are the central measuring units in the low-voltage grid and in combination with consumer energy management systems (CEMS) provide an interface to home and building automation. They are involved in use cases such as *Demand Response Management*, *Demandside Management*, *Energy Feedback* and *Electric Vehicle Charging*. Smart Metering therefore exhibits high demands with regard to security and privacy requirements.

Smart meters form a core component of smart grids. Each smart meter contains a processor, as well as storage and communication facilities and is capable of measuring and transmitting detailed usage data. The regulation of the intervals in which the measurements are to be provided differs by country. In many countries, 15-minute intervals seem to be the choice of the regulators. Almost every smart meter model will be capable of measuring energy usage in much higher resolutions (one-second granularity or less).

The current focus in smart metering is electricity metering. There are also proposals to include the metering of water, heat, and gas for multi-utility metering. In many approaches, the electricity meter serves as the main external communication gateway, other utility meters provide data to the power gateway that relays the information to the utility providers. Other approaches have all meters report to a metering gateway that in turn relays the information, see, e.g., [41].

Feedback on energy (or other utility) usage can be provided to the user through various channels, such as in-home displays, but also through web and smart phone applications.

The so-called *customer gateway* [85] is the gateway between the custome premises and the utility provider and other third parties. It is involved in all interactions between the smart grid and devices in customer premises and may support home automation, demand response management. The customer gateway functionality can be integrated into the smart meter unit.

The communication architecture of the future smart grid is not yet finally defined. Due to the manifold needs in various parts of this heterogeneous network multiple standards and protocols will be relevant. According to [41], the focus of research in this area will not be on a particular technology but more on interoperability between different system components like meters, devices and protocols. Nevertheless some typical and well-known communication and data exchange standards applicable in different segments of the smart grid communication network can be identified, such as DLMS/COSEM.

1.1.2 *Privacy*

Privacy can be defined as "the right of the individual to determine when, how, and to what extent he or she will release personal information"[1]. Technological privacy approaches can be seen as tools at the individual's disposal to enforce this right.

It has been argued that both security and privacy need to be addressed from the earliest stages in the development and standardization process for smart grid technology. The terms "security by design" and "privacy by design" [13] are used to describe principles to allow security and privacy to be built into the system, rather than be treated as add-ons.

There are two kinds of privacy approaches: regulatory-based and technology-based [41]. As our focus is a technological one, regulatory-based approaches will only be taken into account as far as they provide requirements for technological solutions. An important source for regulatory scenarios and recommendations are the reports of the European Commission Smart Grid Expert Group Two for regulatory recommenda-

[1] R. v. Duarte, Supreme Court of Canada, 1990-25-01

tions for data safety, data handling and data protection, e.g., [40], as well as the results of the CEN/CENELEC/ETSI workgroups addressing EU Mandate M/490 "Standardization Mandate to support European Smart Grid deployment". Other sources include Common Criteria for Information Technology Security Evaluation (ISO/EIC 15408) and country-specific recommendations, such as the Federal Office for Information Security (BSI) in Germany [12] or the Requirements Catalog for End-to-End Security for Smart Metering commissioned by the Association of Austrian Energy Providers and co-authored by the European Network for Cyber Security (`https://www.encs.eu`).

Overviews of privacy issues and privacy enhancing are given by [11, 43, 76]. There is a number of contributions that deal with technological approaches to end-user privacy in general, for an overview see [47]. In the context of smart grid privacy, a major part of current proposals is focused on smart metering and the load profiles generated by smart meters.

Extractable Information

There is a lot of public concern and discussions on the privacy impact of smart grid technologies. However, the discussion is led without knowing the extent of personal information that can be read out of the involved data. Even more so, there is nearly a complete lack of knowledge about how the amount of personal information relates to the measured time interval.

The area of the smart grid, where most contributions to the privacy debate, are centered on, is Smart Metering. In many countries in Europe it is planned that smart meters will deliver load data in 15 minute time intervals [93]. This has raised privacy concerns (cf., [74, 13, 81]). However, to our knowledge, no one has tried to assess the amount of personal information that can be extracted on 15 minute time interval load profiles, or how, in general, data granularity relates to the amount and nature of extractable personal data.

Information is usually extracted from the load profiles by means of so-called "non-intrusive load monitoring analysis" (NILM). There is a lot of literature on NILM algorithms, e.g., [46, 103, 9, 5, 100, 63, 56, 87]. The goal of these algorithms is the disaggregation of the total load into the individual appliances loads, e.g., for sake of providing energy feedback to the end-user. From the privacy viewpoint, such NILM analyses can be seen as a first step of attacking methods, which aim at the unauthorized extraction of personal information.

To date, there is little systematic research on this subject in the context of smart grids. In [91] an information theoretic approach to abstract privacy and utility requirements is used. The authors aim at providing a measure for the amount of information leaked, and also for the utility that is retained in the data at different levels of abstraction.

There are a number of approaches for matching appliance signatures to load profiles to determine which appliances were used at what time and for how long, e.g. [46, 65, 73]. As mentioned above these types of approaches are usually refered to as "non-intrusive load monitoring" (NILM).

Detection based on NILM is remarkably accurate: In [75] over 90% accuracy are reported in detecting presence and sleep cycle intervals. The results show that "personal information can be estimated with a high degree of accuracy, even with relatively unsophisticated hardware and algorithms" [75, p. 2]. The authors of [71] use genetic algorithms for identification and report flawless identification for up to 10 types of appliances. In [52] successful identification of appliances in low resolution load profiles is reported, e.g., 30min intervals, with the use of data-mining techniques.

Smart Meter Privacy

The aforementioned privacy concerns that arise through the availability of detailed load profiles per customer is documented by a variety of studies, e.g., [90, 78, 55, 75, 74]. In [75] results of a collaboration between researchers from law and engineering are reported. The authors argue that there "exist strong motivations for entities involved in law enforcement, advertising, and criminal enterprises to collect and repurpose power consumption data" [75, p. 1]. For example, burglars could use the data to determine occupancy patterns of houses to time break-ins. Marketing agencies could identify specific brands of used appliances, which could then be used for targeted advertising. In summary, while there are many useful applications of smart meter data, such as energy saving and tailor-made energy rates, the privacy of this kind of data needs to be secured.

It has been argued, that approaches relying on policy alone, may prove inadequate to provide a sufficient level privacy and that technological methods that enforce privacy by virtue of "strength of mechanism" need to be employed [49]. Indeed, a number of such technological approaches have been suggested to remedy the (perceived) loss in privacy and still enable smart metering functionality on a broad basis.

Privacy Enhancing Technologies

There are a number of contributions for privacy-enhancing technologies for smart metering. Jawurek et al. [49] argue, that approaches relying on policy alone, may prove inadequate to provide a sufficient level of privacy and that technological methods that enforce privacy by virtue of "strength of mechanism" need to be employed. Indeed, a number of such technological approaches have been suggested to remedy the loss in privacy and still enable smart metering functionality on a broad basis. In the following, we give a brief overview of these contributions, based on own work published in [30]. More detailed surveys can be found in [49, 38, 43, 76, 11].

The only approach that is widely used in the real world at this point in time, is **anonymization or pseudonymization** of smart metering data. Consumption data and the personal data are split and stored separately. Methods for de-anonymization are a major threat for these types of approaches. It has been shown that even after anonymization or pseudonymization, data items can still be attributed to the individual that originated them. Jawurek et al. [48] show that de-anonymization can also be done in the smart grid user domain. This structural traceability is a problem for schemes that rely on anonymization or pseudonymization only without the use of additional encryption.

Simple aggregation tries to hide data related to individuals by aggregating over a number of house-holds, e.g., all households in a neighborhood are network (NAN). For example, Bohli et al. [10] propose a privacy scheme in which high resolution smart meter readings are aggregated at NAN level and only the aggregate is sent to the utility. They introduce two solutions both with and with-out involvement of trusted third parties.

Due to the inherent link between load data resolution and privacy, splitting the load data into a variety of **different resolutions**, each associated with different authorization levels, has been proposed by a number of contributions. For example, the anonymization scheme proposed by Efthymiou and Kalogridis [24] is based on two different resolutions: a low resolution that can be used for billing purposes, and a high resolution that allows further investigation. This scheme employs a trusted third party escrow service. In the papers included in this treatise, we propose wavelet-based multi-resolution privacy (e.g., [33]).

Masking relates to approaches which add numerical artifacts, e.g., random sequences to the original load data to obfuscate individual contributions. The added artifacts are constructed in such a way that they cancel each other out upon aggregation. The aggregator can therefore combine the data values of all participant to create an accurate aggregation, but cannot gain access to individual contribution. For example, Kursawe et al. [64] propose such an aggregation protocol, which compared to other approaches has the advantage of relatively low computational complexity. Defend and Kursawe [21] further improve on this idea. Danezis et al. [19] present another low-overhead protocol for aggregation of smart meter data, which puts minimal computational demands on the smart meter hardware.

Differential privacy, as Dwork [22, p. 1] puts it, roughly speaking, "ensures that (almost, and quantifiably) no risk is incurred by joining a statistical database". Adding or removing an item from the database will not (or only to a very limited degree) affect the result of statistical computations. This is commonly achieved by the distributed generation of noise which is added to the individual data contribution. Shi et al. [94]

propose a scheme for adding random noise to time series data using a symmetric geometric distribution. An advantage of this scheme is that the participants need not trust each other, nor rely on a trusted aggregator. As another example, Ács and Castelluccia [1] obscure individual data sets by adding Laplacian noise, which is jointly generated by the participants. Apart from the obvious drawback that the data is no longer exact after differential privacy is applied, data pollution by malicious participants is another issue with this approach [94].

Secure Signal Processing (SSP) refers to the possibility to perform certain computations, such as aggregation in the encrypted domain. A commonly employed mechanism in SSP is additively homomorphic encryption, which allows some specific manipulations of the ciphertext to be reflected in the plaintext domain. For example, Li et al. [72] propose an overlay network in a tree-like topology and the use of a Paillier cryptosystem. Garcia and Jacobs [45] combine secret sharing with a Paillier cryptosystem to add flexibility in the aggregation (at the expense of additional computational complexity). Erkin and Tsudik [39] extend the idea of homomorphic encryption of smart meter readings by splitting the module into random shares, which, in combination with a modified Pailler cryptosystem, allows flexible spatial and temporal aggregation for different use cases, such as billing or network monitoring. In our own work, we propose the combination of SSP with multi-resolution methods to increase customer privacy choices [32, 59].

1.1.3 *Security*

An overview of research in smart grid security in general can be found in [8]. An overview of smart grid communication security challenges and risks can be found in [70]. There is a large number of publications that discuss security issues in the smart grid user domain specifically, e.g. [2, 3, 6, 23, 55, 96]. In [16] security requirements in an advanced metering infrastructure are discussed. In [41] top-down and bottom-up approaches to smart grid security are distinguished. The top-down approaches focus on user scenarios, such as smart meter reading and billing. The bottom-up approach focuses on security features such as integrity, authentication, authorization, key management and intrusion detection.

There are a number of proposals for **communication security** specific to the smart grid end-user domain. Secure transmission of smart meter data is a key topic addressed by virtually all contributions. The authors of [104] propose a secure multi-cast protocol that automatically derives group memberships and verifies configuration performance. A security protocol for smart meter aggregation that provides hop-by-hop security, while still providing end-to-end security, is proposed by [7]. In [79], a comprehensive proposal for securing smart grid infrastructure is given, including a proposal for a key infrastructure. The authors of [15] propose a scheme for authentication in the smart grid that is privacy aware.

In [57] a secure transport protocol for smart grid data collection in general is presented. The authors of [14] propose a model-based access control system. In [95] a zero-configuration identity-based signcryption scheme for the smart grid is proposed.

General-purpose schemes for securing communication and authenticating communicating parties are of course also valid candidates for the end-user domain of smart grids as well. The German Federal Institute of Information Security (BSI) defines a detailed protection profile for the communication gateway in a smart meter system [12, 67]. The Austrian Association of Energy Providers has commissioned and co-authored a document detailing the requirements for end-to-end security in Smart Metering [50].

Apart from communication security, the protection of the actual data content from unauthorized access, even within a secured communication environment, is an important topic which is addressed by **content security**. The authors of [41] make the case for a system in which insiders will access "data in an authorized manner and will only use this data in an *acceptable* manner" [41, p. 8]. They propose to use a digital rights management system to ensure that data is only accessed in an acceptable manner. Traditionally, the topic of content security has been widely discussed in the context of multimedia data, e.g., [44].

Intrusion detection for smart grid has been identified as an important and critical topic to be addressed, e.g., the authors of [41] point out that mechanisms should be put into place that allow to detect attacks and misuse of data. On the regulatory side, these mechanisms should have counterparts and allow action against malicious parties. However, as rightly argued by [80], "Intrusion detection system (IDS) techniques for this domain are still in their infancy with very little work reported in the literature" [80, page 1]. Of the few contributions that are available, many suffer from high rate of false positives (which, of course, is a death sentence for any IDS). See [80] for a discussion of issues with previously propsed IDS for smart grids. There are a few promising approaches which aim at avoiding false positives. [106, 105] propose an architecture for a distributed IDS in smart grids, based on the artificial immune system approach, which relies on intelligent analyzing modules deployed on various levels of the grid. In [80], a behavior-rule based intrusion detection system for smart grids is proposed, which in initial evaluations has been shown to give a high performance at relatively low false positive rates. Of course, with the vast number of data observations encountered in the smart grid, false positive rates need to be kept at an absolute minimum. As an example for the end-user domain, an intrusion detection system for the HAN in smart grids is proposed by [51]. Our own work, which is included in this treatise, focuses on anomaly detection using a combination of classifiers [77].

1.1.4 *User Control*

In pilots and studies, end users frequently report doubts on the benefits of smart grid technologies on a personal level and difficulties in understanding the level of control they can exert in a smart grid environment. For example, a survey on user preferences in smart metering in Switzerland found among the strongest user concerns regarding smart metering was the lack of control of data and pricing and little potential for savings [54].

Smart grid user control is not as clearly defined a research field as are smart grid privacy and security. To date there is only a limited number of contributions that deal specifically with the user integration into smart grid processes in general. For example, in [13] a number of possible user interactions with smart grid technologies are listed:

- Understanding of how households use energy, better management of energy, and reduction of the carbon footprint,
- Control of expenditure on electricity,
- Experience of fewer and shorter power outages, and notification when the power will come back on, and
- Control energy devices in the home.

As the smart grid is envisioned to interface with home and building automation processes, the list can be arbitrarily expanded.

It is clear that a field like "user control" is necessarily broad. Our own work focuses on agent-based frameworks that uses a game-theoretic approach to improve user control in demand response scenarios [68, 69] and models for users' privacy requirements [60, 61].

1.2 CONTRIBUTION

The primary goal of the methods proposed in this treatise is to improve existing approaches and develop new mechanisms to establish secure, private and informed – i.e., trusted – user interaction with smart grid technology. The fundamental question addressed is: How can a sustainable level of trust be established in smart grid user domain entities (applications, processes, data protocols, and interfaces) in order to invite and further user participation in smart grid technologies? The contribution of this cumulative habilitation treatise has been published in a number of peer-reviewed journal publications as well as a number of publications in conference proceedings. In the following, the contribution of each of these publications is summarized briefly and put into context to each other.

[Eibl14a]

- G. Eibl and D. Engel. Influence of data granularity on nonintrusive appliance load monitoring. In *Proceedings of the Second ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec '14)*, pages 147–151, Salzburg, Austria, 2014. ACM.

Smart metering has been in the focus of the discussion on privacy issues in smart grids. This discussion has been led without taking the resolution of the underlying data into account, also in the academic discourse. In two papers, we have presented a formal analysis on the impact of resolution on smart meter privacy.

In this first paper, we analyse the impact of resolution on a the first step of non-intrusive load monitoring (NILM), i.e., the identification of individual appliances in a household load profile. We show that decreasing resolution has an impact mainly on recall (rather than on precision) in NILM.

[Eibl15a]

- G. Eibl and D. Engel. Influence of data granularity on smart meter privacy. *IEEE Transactions on Smart Grid*, 6(2):930–939, March 2015.

In this second paper, we extend on the previously presented NILM results and discuss the the influence of load profile resolution on the degree of extractable personal information. The intuitive claim that lowering the resolution will increase privacy has been studied systematically. We show that this is indeed the case and that a dyadic series of decreasing resolutions is suitable for providing a series of privacy levels to the end-user. Although this paper was published after our work on wavelet-based smart meter privacy, it provides the formal basis for the utility of this approach.

[Engel11a]

- D. Engel. Conditional access smart meter privacy based on multi-resolution wavelet analysis. In *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies*, pages 45:1–45:5, New York, NY, USA, 2011. ACM.

In terms of methods for privacy enhancement in smart metering, our contributions focus on the idea of lowering the resolution of the available data, providing a wavelet-based data representation, that integrates all resolutions in a single bitstream without data expansion and providing conditional access combined with hierarchical key management.

In this paper, the initial idea of wavelet-based data representation of load profiles is proposed. We have shown that the wavelet transform is a suitable tool to provide the aforementioned dyadic decreases in resolution, as each iterative application of the wavelet transform effectively halves the resolution. A first proof-of-concept implementation is presented and evaluated on inexpensive hardware (a *Beagleboard* in this case).

[Engel13a]

- D. Engel. Wavelet-based load profile representation for smart meter privacy. In *Proceedings IEEE PES Innovative Smart Grid Technologies (ISGT'13)*, pages 1–6, Washington, D.C., USA, Feb. 2013. IEEE.

In this paper, the idea of multi-resolution representation is elaborated. Two wavelet filters are investigated for this purpose: the simple Haar Wavelet and the LeGall 5/3 Wavelet. We show that while the integer-based LeGall 5/3 Wavelet is faster, due to the necessary border handling and the ensuing errors, it is not a possible choice. The Haar Wavelet is well suited for application, as it is also fast in application and provides lossless transformation in practice. Furthermore, we discuss the fact that the original sum is preserved over multiple wavelet decomposition, which is an important property for use-cases such as billing (which also has to work at the lowest resolution). The performance of the approach is evaluated at the example of an extended prototypical implementation. It is shown that the performance these environments offer is sufficient for use in the field.

[Engel13e]

- D. Engel and G. Eibl. Multi-resolution load curve representation with privacy-preserving aggregation. In *Proceedings of IEEE Innovative Smart Grid Technologies (ISGT) 2013*, pages 1–5, Copenhagen, Denmark, Oct. 2013. IEEE.

In this paper, we show that our approach can be combined with additively homomorphic encryption to provide additional degrees of freedom. Formal proof is given that wavelets are fully compatible with additively homomorphic encryption in the context of a Paillier cryptosystem [86]. A proof-of-concept implementation is presented and the high computational demands are discussed.

[Peer14a]

- C. Peer, D. Engel, and S. Wicker. Hierarchical key management for multi-resolution load data representation. In *Proceedings of 5th IEEE International Conference on Smart Grid Communications (SmartGridComm 2014)*, pages 926–932, Venice, Italy, Nov. 2014. IEEE.

Privacy-aware data representations need to be secured for access by authorized parties. In order to facilitate secured, authorized access, different keys need to be provided for different privacy levels. For example, for multi-resolution load profile data, a different key is required for each resolution. For this setup, in this paper we propose a hierarchical key scheme, based on previous work by Lamport [66]. In order to limit the number of keys a user needs to handle, a hierarchical key generation scheme is used. With this key scheme it is ensured that a user who has the key for the highest resolution can derive all necessary sub-keys to the lower resolutions (which are, of course, needed to reconstruct the highest resolution)

[Engel16a]

- D. Engel and G. Eibl. Wavelet-based multiresolution smart meter privacy. *IEEE Transactions on Smart Grid*, PP(99):1–12, 2016. preprint.

In this paper, all of the components for wavelet-based multi-resolution data representation in Smart Metering are integrated into system of a whole. The underlying protocols are presented and discussed in detail, as well as the different communication paths from Smart Meter to distribution system operator, energy provider or another third party. The degrees of freedom are addressed, e.g., the alternatives to run the protocols with or without a data concentrator. For the first time, a comprehensive and detailed security and privacy analysis is conducted, including the basic security assumptions, different attacker models and the discussion of privacy properties from an information-theoretic point of view.

[Knirsch17a]

- F. Knirsch, G. Eibl, and D. Engel. Multi-resolution privacy-enhancing technologies for smart metering. *EURASIP Journal on Information Security*, 2017(1):6, 2017.

Based on the idea of multi-resolution data representation in smart metering, this paper explores the options to combine the previously proposed wavelet-based representation with other privacy enhancing technologies (PETs). The paper discusses the applicability of multi-resolution methods to three PETs: masking protocols, differential privacy and secure aggregation.

[Knirsch15b]

- F. Knirsch, D. Engel, C. Neureiter, M. Frincu, and V. Prasanna. Model-driven privacy assessment in the smart grid. In *Proceedings of the 1st International Conference on Information Systems Security and Privacy (ICISSP)*, pages 173–181, Feb 2015. Best Paper Award.

This paper marks a turn from multi-resolution privacy methods to the field of "user control", namely the users' need for privacy and the requirements in terms of smart grids use cases. These two are often in a counterposition, e.g., a DSO who wants to employ user data for network planning is interested in as high a data resolution as possible, whereas a user may be reluctant to provide this high resolution data, as it would allow the DSO to deduce personal information.

This paper aims at assessing privacy implication of data transfer in the smart grid user domain based on a model of data flows. Based on an ontology that captures the possible threats, the paper develops a first conceptual model of how privacy assessment in the smart grid user domain can be done in a formal way. A toy example is given for illustrative purposes, in which the threat of user presence at home is modelled. The paper won the Best Paper Award at the International Conference on Information Systems Security and Privacy.

[Knirsch16a]

- F. Knirsch, D. Engel, C. Neureiter, M. Frincu, and V. Prasanna. Privacy assessment of data flow graphs for an advanced recommender system in the smart grid. In O. Camp, E. Weippl, C. Bidan, and E. Aïmeur, editors, *Information Systems Security and Privacy – Revised and Selected Papers of ICISSP 2015*, volume 576 of *Communications in Computer and Information Science*, pages 89–106. Springer International Publishing, 2016. Best Paper Award.

This paper extends the previous paper and details the previously presented approach. The concept of data flow graphs as the basis for privacy threat assessment is detailed, and the ontolgy for threat assessment is extended. A real-life example of the privacy implications of smart metering is presented and discussed in detail. Finally, the idea of a recommender system based on the privacy assessment approach is developed. Based on the well-known policy decision point and policy enforcement point patterns, the privacy assessment approach is used to automatically provide guidelines for users who access new applications (such as an application for energy saving tipps provided by a third party supplier that would request fine-grained energy consumption data).

[Knirsch15a]

- F. Knirsch, D. Engel, M. Frincu, and V. Prasanna. Model-based assessment for balancing privacy requirements and operational capabilities in the smart grid. In *Proceedings of the 6th Conference on Innovative Smart Grid Technologies (ISGT)*, pages 1–5, Feb 2015.

In this paper, a model for balancing users' requirements regarding privacy with functional requirements of operational use cases is presented. It is aimed to find an optimal balance between these two, often conflicting, requirements, automatically. This can be achieved to a certain degree, as is demonstrated at the example of a demand-response use case in the University of Southern California microgrid.

[Unterweger15a]

- A. Unterweger and D. Engel. Resumable load data compression in smart grids. *IEEE Transactions on Smart Grid*, 6(2):919–929, March 2015.

Privacy measures try to assess the amount of information contained in data. While the previous papers tried to assess privacy from a model-based view, there are also some ideas to come up with an entropy-like meausre for privacy. While exploring options in this direction, we found that the compressibility (which, of course, is related with entropy) of load profiles had not been studied systematically before.

In this paper, an approach for compression of load data is proposed, that is based on ideas for compression of multimedia data. It is shown that the approach is lightweight on the side of the smart meter and that it is resumable, which is an important property if a smart meter loses connectivity for a period of time.

[Unterweger15b]

- A. Unterweger, D. Engel, and M. Ringwelski. The effect of data granularity on load data compression. *Springer Lecture Notes in Computer Science – Energy Informatics 2015*, 9424:69–80, 2015.

In this paper, we extend the previous work on load data compression and data granularity and explore the effect of data granularity on compression results. For this investigation we joined forces with Martin Ringwelski the main author of the only other algorithm specifically designed for load data compression. We investigate the properties of our algorithm compared to Ringwelski's approach. It turns out that depending on data resolution, one or the other may be preferable.

[Eibl15b]

- G. Eibl, D. Engel, and C. Neureiter. Privacy-relevant smart metering use cases. In *Proceedings of IEEE International Conference on Industrial Technology (ICIT) 2015*, pages 1387–1392, Seville, Spain, 2015. IEEE.

In discussion with the company partners of the Josef Ressel Research Center, it became evident that there is no clear view on the privacy relevance of different smart metering use cases and how this could be addressed by specific PETs. Such an account is also missing in literature. In this paper, we aim at bridging the gap between privacy requirements of smart metering use cases and the features different PETs have to offer.

[Lueckenga16a]

- J. Lückenga, D. Engel, and R. Green. Weighted vote algorithm combination technique for anomaly based smart grid intrusion detection systems. In *Proceedings of International Joint Conference on Neural Networks (IJCNN) 2016*, pages 2738–2742, Vancouver, Canada, July 2016.

In this paper, the issue of anomaly detection in smart grids is discussed. Intrusion Detection Systems (IDS) are a crucial and necessary aspect of the smart grid, particularly when considering the possible attack vectors and their consequences. While there are many different proposals for IDS in smart grids, the benefits of an anomaly detection technique is still in discussion, due to its capability of detecting zero-day attacks and misuse. This paper proposes a weighted vote classification approach and a general weight calculation function to improve the detection performances of anomaly IDS systems. Initial results show that a combination technique is able to improve classifier performance by several percent.

[Lausenhammer15a]

- W. Lausenhammer, D. Engel, and R. Green. A game theoretic software framework for optimizing demand response. In *Proceedings of the 6th Conference on Innovative Smart Grid Technologies (ISGT)*, pages 1–5, Feb 2015.

Particularly with respect to coordinating power consumption and generation, demand response (DR) is a vital part of the future smart grid. Even though, there are some DR simulation platforms available, none makes use of game theory. While many benefits of DR are currently under study, an issue of particular concern is optimizing end-users' power consumption profiles at various levels. This study proposes the concept for a fundamental, game theoretic, multi-agent software framework for DR simulation that is capable of investigating the effect of optimizing multiple electric appliances by utilizing game theoretic algorithms. Initial results show that by shifting the switch-on time of three household appliances provides a savings of up to 6%.

[Lausenhammer16a]

- W. Lausenhammer, D. Engel, and R. Green. Utilizing capabilities of plug in electric vehicles with a new demand response optimization software framework: Okeanos. *International Journal of Electrical Power and Energy Systems*, 75:1–7, 2016.

This paper extends our previous work on DR simulation. The previously proposed software framework for DR simulation is detailed and extended to address an evaluation of real-world use cases. While initial use cases were based on game theoretic algorithms and focus on consumption devices only, further use cases evaluate the effects of plug in electric vehicles (PEVs). Results with consumers show that the number of involved households does not affect the costs per household. Further evaluation involving PEVs demonstrates that with an increasing penetration of PEVs and feed-in tariffs the costs per household per month decrease.

1.3 REFERENCES

[1] G. Acs and C. Castelluccia. I have a DREAM! (DiffeRentially privatE smArt Metering). In *Proc. Information Hiding Conference*, pages 118–132, 2011.

[2] R. Anderson and S. Fuloria. On the Security Economics of Electricity Metering. In *Proceddings of the Ninth Workshop on the Economics of Information Security (WEIS 2010)*, 2010.

[3] R. Anderson and S. Fuloria. Who Controls the off Switch? In *Proc. First IEEE Int Smart Grid Communications (SmartGridComm) Conf*, pages 96–101, nov 2010.

[4] S. Auer, A. Bliem, D. Engel, A. Uhl, and A. Unterweger. Bitstream-based JPEG encryption in real-time. *International Journal of Digital Crime and Forensics*, 5(3):1–14, 2013.

[5] M. Baranski and J. Voss. Genetic algorithm for pattern detection in NIALM systems. In *IEEE International Conference on Systems, Man and Cybernetics*, 2004.

[6] A. Barenghi and G. Pelosi. Security and Privacy in Smart Grid Infrastructures. In *Proc. 22nd Int Database and Expert Systems Applications (DEXA) Workshop*, pages 102–108, 2011.

[7] A. Bartoli, J. Hernández-Serrano, M. Dohler, A. Kountouris, and D. Barthel. Secure Lossless Aggregation for Smart Grid {M2M} Networks. In *Proceedings of First IEEE International Conference on Smart Grid Communications*, pages 333–338, Gaithersburg, Maryland, USA, 2010.

[8] T. Baumeister. Literature Review on Smart Grid Cyber Security. Technical report, University of Hawaii at Manoa, 2010.

[9] D. C. Bergman, D. Jin, J. Juen, N. Tanaka, C. Gunter, and A. Wright. Distributed Non-Intrusive Load Monitoring. In *Proceedings of the IEEE/PES Conference on Innovative Smart Grid Technologies (ISGT 2011), Anaheim, CA, USA, January 2011*, 2011.

[10] J.-M. Bohli, C. Sorge, and O. Ugus. A Privacy Model for Smart Metering. In *2010 IEEE International Conference on Communications Workshops (ICC)*, pages 1–5, 2010.

[11] F. Borges de Oliveira. *On Privacy-Preserving Protocols for Smart Metering Systems*. Springer International Publishing, 2017.

[12] Bundesamt für Sicherheit in der Informationstechnik (German Federal Office for Information Security). Protection Profile for the Gateway of a Smart Metering System – V1.3. `https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Zertifizierung/Reporte/ReportePP/pp0073b_pdf.pdf?__blob=publicationFile&v=1`, 2014.

[13] A. Cavoukian, J. Polonetsky, and C. Wolf. SmartPrivacy for the smart grid: embedding privacy into the design of electricity conservation. *Identity in the Information Society*, 3(2):275–294, 2010.

[14] H. Cheung, A. Hamlyn, T. Mander, C. Yang, and R. Cheung. Role-based model security access control for smart power-grids computer networks. In *Proc. IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, pages 1–7, 2008.

[15] T. W. Chim, S. M. Yiu, L. C. K. Hui, and V. O. K. Li. PASS: Privacy-preserving authentication scheme for smart grid network. In *Proc. IEEE Int Smart Grid Communications (SmartGridComm) Conf*, pages 196–201, 2011.

[16] F. M. Cleveland. Cyber security issues for Advanced Metering Infrasttructure (AMI). In *Proc. IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, pages 1–5, 2008.

[17] C. Cuijpers and B.-J. Koops. Smart metering and privacy in europe: lessons from the dutch case. In S. Gutwirth, R. Leenes, P. de Hert, and Y. Poullet, editors, *European data protection: coming of age*, pages 269–293. Springer, 2013.

[18] C. Dänekas, C. Neureiter, S. Rohjans, M. Uslar, and D. Engel. Towards a model-driven-architecture process for smart grid projects. In P. Benghozi, D. Krob, A. Lonjon, and H. Panetto, editors, *Digital Enterprise Design & Management*, volume 261 of *Advances in Intelligent Systems and Computing*, pages 47–58. Springer International Publishing, 2014.

[19] G. Danezis, C. Fournet, M. Kohlweiss, and S. Zanella-Béguelin. Smart Meter Aggregation via Secret-sharing. In *Proceedings of the First ACM Workshop on Smart Energy Grid Security*, SEGS '13, pages 75–80, New York, NY, USA, 2013. ACM.

[20] S. J. De and D. L. Metayer. Privacy Harm Analysis: A Case Study on Smart Grids. In *2016 IEEE Security and Privacy Workshops (SPW)*, pages 58–65. IEEE, may 2016.

[21] B. Defend and K. Kursawe. Implementation of privacy-friendly aggregation for the smart grid. In *Proceedings of the first ACM workshop on Smart energy grid security - SEGS '13*, pages 65–74, 2013.

[22] C. Dwork. Differential Privacy: A Survey of Results. In M. Agrawal, D. Du, Z. Duan, and A. Li, editors, *Theory and Applications of Models of Computation*, volume 4978 of *Lecture Notes in Computer Science*, pages 1–19. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[23] C. Eckert and C. Krauß. Sicherheit im Smart Grid – Herausforderungen und Handlungsempfehlungen. *Datenschutz und Datensicherheit*, 8:535–541, 2011.

[24] C. Efthymiou and G. Kalogridis. Smart Grid Privacy via Anonymization of Smart Metering Data. In *Proceedings of First IEEE International Conference on Smart Grid Communications*, pages 238–243, Gaithersburg, Maryland, USA, 2010.

[25] G. Eibl and D. Engel. Influence of data granularity on nonintrusive appliance load monitoring. In *Proceedings of the Second ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec '14)*, pages 147–151, Salzburg, Austria, 2014. ACM.

[26] G. Eibl and D. Engel. Influence of data granularity on smart meter privacy. *IEEE Transactions on Smart Grid*, 6(2):930–939, March 2015.

[27] G. Eibl and D. Engel. Differential privacy for real smart metering data. *Computer Science – Research and Development*, 32(1):173–182, 2017.

[28] G. Eibl, D. Engel, and C. Neureiter. Privacy-relevant smart metering use cases. In *Proceedings of IEEE International Conference on Industrial Technology (ICIT) 2015*, pages 1387–1392, Seville, Spain, 2015. IEEE.

[29] D. Engel. Conditional access smart meter privacy based on multi-resolution wavelet analysis. In *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies*, pages 45:1–45:5, New York, NY, USA, 2011. ACM.

[30] D. Engel. Privacy-preserving smart metering: Methods and applicability (invited talk). In *Proceedings of the fourth Workshop on Communications for Energy Systems*, pages 9–16, Vienna, Austria, Sept. 2013. Austrian Electrotechnical Association.

[31] D. Engel. Wavelet-based load profile representation for smart meter privacy. In *Proceedings IEEE PES Innovative Smart Grid Technologies (ISGT'13)*, pages 1–6, Washington, D.C., USA, Feb. 2013. IEEE.

[32] D. Engel and G. Eibl. Multi-resolution load curve representation with privacy-preserving aggregation. In *Proceedings of IEEE Innovative Smart Grid Technologies (ISGT) 2013*, pages 1–5, Copenhagen, Denmark, Oct. 2013. IEEE.

[33] D. Engel and G. Eibl. Wavelet-based multiresolution smart meter privacy. *IEEE Transactions on Smart Grid*, PP(99):1–12, 2016. preprint.

[34] D. Engel, T. Stütz, and A. Uhl. Evaluation of JPEG2000 hashing for efficient authentication. In *Proceedings of International Conference on Multimedia & Expo, ICME '09*, pages 1728–1731, New York, NY, USA, June 2009.

[35] D. Engel, T. Stütz, and A. Uhl. A survey on JPEG2000 encryption. *Multimedia Systems*, 15(4):243–270, 2009. Springer.

[36] D. Engel, T. Stütz, and A. Uhl. Assessing JPEG2000 encryption with key-dependent wavelet packets. *EURASIP Journal on Information Security*, 2012(1):1–16, 2012.

[37] D. Engel, A. Uhl, and A. Unterweger. Region of interest signalling for encrypted JPEG images. In *Proceedings of the first ACM workshop on Information hiding and multimedia security (IHMMSEC '13)*, pages 165–174, Montpellier, France, 2013. ACM.

[38] Z. Erkin, J. R. Troncoso-Pastoriza, R. L. Lagendijk, and F. Perez-Gonzalez. Privacy-preserving data aggregation in smart metering systems: an overview. *IEEE Signal Processing Magazine*, 30(2):75–86, mar 2013.

[39] Z. Erkin and G. Tsudik. Private computation of spatial and temporal power consumption with smart meters. In *Proceedings of the 10th international conference on Applied Cryptography and Network Security*, ACNS'12, pages 561–577, Berlin, Heidelberg, 2012. Springer-Verlag.

[40] European Commission Task Force Smart Grids, Expert Group 2: Regulatory Recommendations for Data Safety, Sata Handling and Data Protection.

Report. http://ec.europa.eu/energy/gas_electricity/smartgrids/doc/expert_group2.pdf, Feb. 2011. Online.

[41] Z. Fan, P. Kulkarni, S. Gormus, C. Efthymiou, G. Kalogridis, M. Sooriyabandara, Z. Zhu, S. Lambotharan, and W. H. Chin. Smart grid communications: Overview of research challenges, solutions, and standardization activities. *IEEE Communications Surveys and Tutorials*, 15(1):21–38, 2013.

[42] X. Fang, S. Misra, G. Xue, and D. Yang. Smart grid - The new and improved power grid: A survey. *IEEE Communications Surveys and Tutorials*, 14(4):944–980, 2012.

[43] S. Finster and I. Baumgart. Privacy-Aware Smart Metering: A Survey. *IEEE Communications Surveys & Tutorials*, 16(3):1732–1745, jan 2014.

[44] B. Furht and D. Kirovski. *Multimedia Security Handbook*. CRC Press, Inc., Boca Raton, FL, USA, 2004.

[45] F. Garcia and B. Jacobs. Privacy-Friendly Energy-Metering via Homomorphic Encryption. In J. Cuellar, J. Lopez, G. Barthe, and A. Pretschner, editors, *Security and Trust Management*, volume 6710 of *Lecture Notes in Computer Science*, pages 226–238. Springer Berlin Heidelberg, Berlin Heidelberg, 2011.

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

[47] G. Iachello and J. Hong. End-user privacy in human-computer interaction. *Found. Trends Hum.-Comput. Interact.*, 1(1):1–137, 2007.

[48] M. Jawurek, M. Johns, and K. Rieck. Smart metering de-pseudonymization. In *Proceedings of the 27th Annual Computer Security Applications Conference*, ACSAC '11, pages 227–236, New York, NY, USA, 2011. ACM.

[49] M. Jawurek, F. Kerschbaum, and G. Danezis. Privacy Technologies for Smart Grids – A Survey of Options. Technical report, Microsoft Research, 2012.

[50] M. John, K. Kursawe, C. Peters, W. Löw, T. Aichholzer, J. Bernhardt, C. Eberl, B. Egger, M. Farthofer, B. Haberler, B. Korittnig, P. Meyer, B. Morscher, F. Neurater, and R. Schmid. End-to-end security for smart metering. Technical report, Oesterreichs Energie, 2014.

[51] P. Jokar, H. Nicanfar, and V. C. M. Leung. Specification-based Intrusion Detection for home area networks in smart grids. In *Proc. IEEE Int Smart Grid Communications (SmartGridComm) Conf*, pages 208–213, 2011.

[52] G. Kalogridis and S. Z. Denic. Data Mining and Privacy of Personal Behaviour Types in Smart Grid. In *Proc. IEEE 11th Int Data Mining Workshops (ICDMW) Conf*, pages 636–642, 2011.

[53] L. Karg, K. Kleine-Hegermann, M. Wedler, and C. Jahn. E-Energy Abschlussbericht – Ergebnisse und Erkenntnisse aus der Evaluation der sechs Leuchtturmprojekte. Technical report, Bundesministerium für Wirtschaft und Technologie (German Federal Ministry for Economy and Technology), 2014. In German.

[54] S. Kaufmann, M. Loock, and R. Wüstenhagen. Kundenpräferenzen für Smart Metering in der Schweiz: Ergebnisse einer explorativen Studie. IWö Arbeitspapier, 2011.

[55] H. Khurana, M. Hadley, N. Lu, and D. A. Frincke. Smart-Grid Security Issues. *IEEE Security & Privacy*, 8(1):81–85, 2010.

[56] H. Kim, M. Marwah, M. F. Arlitt, G. Lyon, and J. Han. Unsupervised Disaggregation of Low Frequency Power Measurements. In *The 11th SIAM International Conference on Data Mining*, pages 747–758, 2011.

[57] Y.-J. Kim, V. Kolesnikov, H. Kim, and M. Thottan. {SSTP}: A scalable and secure transport protocol for smart grid data collection. In *Proc. IEEE Int Smart Grid Communications (SmartGridComm) Conf*, pages 161–166, 2011.

[58] F. Knirsch, G. Eibl, and D. Engel. Error-resilient Masking Approaches for Privacy Preserving Data Aggregation. *IEEE Transactions on Smart Grid*, PP(99):1–12, 2016. Preprint.

[59] F. Knirsch, G. Eibl, and D. Engel. Multi-resolution privacy-enhancing technologies for smart metering. *EURASIP Journal on Information Security*, 2017(1):6, 2017.

[60] F. Knirsch, D. Engel, M. Frincu, and V. Prasanna. Model-based assessment for balancing privacy requirements and operational capabilities in the smart grid. In *Proceedings of the 6th Conference on Innovative Smart Grid Technologies (ISGT)*, pages 1–5, Feb 2015.

[61] F. Knirsch, D. Engel, C. Neureiter, M. Frincu, and V. Prasanna. Model-driven privacy assessment in the smart grid. In *Proceedings of the 1st International Conference on Information Systems Security and Privacy (ICISSP)*, pages 173–181, Feb 2015. Best Paper Award.

[62] F. Knirsch, D. Engel, C. Neureiter, M. Frincu, and V. Prasanna. Privacy assessment of data flow graphs for an advanced recommender system in the smart grid. In O. Camp, E. Weippl, C. Bidan, and E. Aïmeur, editors, *Information Systems Security and Privacy – Revised and Selected Papers of ICISSP 2015*, volume 576 of *Communications in Computer and Information Science*, pages 89–106. Springer International Publishing, 2016. Best Paper Award.

[63] J. Z. Kolter and T. Jaakkola. Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation. *Journal of Machine Learning Research - Proceedings Track*, 22:1472–1482, apr 2012.

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

[65] H. Y. Lam, G. S. K. Fung, and W. K. Lee. A novel method to construct taxonomy of appliances based on load signatures. *IEEE Transactions on Consumer Electronics*, 53(2):653–660, 2007.

[66] L. Lamport. Password Authentication with Insecure Communication. *Commun. ACM*, 24(11):770–772, nov 1981.

[67] D. Laupichler, S. Vollmer, H. Bast, and M. Intemann. Das BSI-Schutzprofil: Anforderungen an den Datenschutz und die Datensicherheit für Smart Metering Systeme. *Datenschutz und Datensicherheit - DuD*, 35(8):542–546, 2011.

[68] W. Lausenhammer, D. Engel, and R. Green. A game theoretic software framework for optimizing demand response. In *Proceedings of the 6th Conference on Innovative Smart Grid Technologies (ISGT)*, pages 1–5, Feb 2015.

[69] W. Lausenhammer, D. Engel, and R. Green. Utilizing capabilities of plug in electric vehicles with a new demand response optimization software framework: Okeanos. *International Journal of Electrical Power and Energy Systems*, 75:1–7, 2016.

[70] A. Lee and T. Brewer. Smart Grid Cyber Security: Strategy and Requirements. NISTIR 7628.

[71] S. K. J. Leung, S. H. K. Ng, and W. M. J. Cheng. Identifying Appliances Using Load Signatures and Genetic Algorithms. In *Proceedings International Conference on Electrical Engineering (ICEE)*, Hong Kong, 2007.

[72] F. Li, B. Luo, and P. Liu. Secure Information Aggregation for Smart Grids Using Homomorphic Encryption. In *Proceedings of First IEEE International Conference on Smart Grid Communications*, pages 327–332, Gaithersburg, Maryland, USA, 2010.

[73] J. Liang, S. Ng, G. Kendall, and J. Cheng. Load Signature Study Part I: Basic concept, structure, and methodology. *IEEE Transactions on Power Delivery*, 25(2):551–560, 2010.

[74] M. Lisovich, D. Mulligan, and S. Wicker. Inferring Personal Information from Demand-Response Systems. *IEEE Security & Privacy*, 8(1):11–20, 2010.

[75] M. A. Lisovich and S. B. Wicker. Privacy Concerns in Upcoming Residential and Commercial Demand-Response Systems. In *Clemson Power Systems Conference 2008*, 2008.

[76] R. Lu. *Privacy-Enhancing Aggregation Techniques for Smart Grid Communications.* Springer International Publishing, 2016.

[77] J. Lückenga, D. Engel, and R. Green. Weighted vote algorithm combination technique for anomaly based smart grid intrusion detection systems. In *Proceedings of International Joint Conference on Neural Networks (IJCNN) 2016*, pages 2738–2742, Vancouver, Canada, July 2016.

[78] P. McDaniel and S. McLaughlin. Security and Privacy Challenges in the Smart Grid. *IEEE Security Privacy Magazine*, 7(3):75–77, 2009.

[79] A. R. Metke and R. L. Ekl. Security Technology for Smart Grid Networks. *IEEE Transactions on Smart Grid*, 1(1):99–107, 2010.

[80] R. Mitchell and I.-R. Chen. Behavior-Rule Based Intrusion Detection Systems for Safety Critical Smart Grid Applications. *IEEE Transactions on Smart Grid*, 4(3):1254–1263, sep 2013.

[81] 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, pages 61–66, New York, NY, USA, 2010. ACM.

[82] C. Neureiter, G. Eibl, D. Engel, S. Schlegel, and M. Uslar. A concept for engineering smart grid security requirements based on SGAM models. *Computer Science - Research and Development*, pages 1–7, 2014.

[83] C. Neureiter, G. Eibl, A. Veichtlbauer, and D. Engel. Towards a framework for engineering smart-grid-specific privacy requirements. In *Proc. IEEE IECON 2013, Special Session on Energy Informatics*, pages 4803 – 4808, Vienna, Austria, Nov. 2013. IEEE.

[84] C. Neureiter, D. Engel, J. Trefke, R. Santodomingo, S. Rohjans, and M. Uslar. Towards consistent smart grid architecture tool support: From use cases to visualization. In *Proceedings of IEEE Innovative Smart Grid Technologies (ISGT) 2014*, Istanbul, Turkey, Oct. 2014. IEEE.

[85] T. Otani. A Primary Evaluation for Applicability of IEC 62056 to a Next-Generation Power Grid. In *Proc. First IEEE Int Smart Grid Communications (SmartGridComm) Conf*, pages 67–72, 2010.

[86] P. Paillier. Public-Key Cryptosystems Based on Composite Degree Residuosity Classes. In J. Stern, editor, *Advances in Cryptology — EUROCRYPT '99: International Conference on the Theory and Application of Cryptographic Techniques Prague, Czech Republic, May 2–6, 1999 Proceedings*, volume 1592 of *Lecture Notes in Computer Science*, pages 223–238. Springer Berlin Heidelberg, Berlin, Heidelberg, 1999.

[87] O. Parson, S. Ghosh, M. Weal, and A. Rogers. Non-intrusive load monitoring using prior models of general appliance types. In *Twenty-Sixth Conference on Artificial Intelligence (AAAI-12).*, 2012.

[88] A. Patel, J. Aparicio, N. Tas, M. Loiacono, and J. Rosca. Assessing communications technology options for smart grid applications. In *Proc. IEEE Int Smart Grid Communications (SmartGridComm) Conf*, pages 126–131, 2011.

[89] C. Peer, D. Engel, and S. Wicker. Hierarchical key management for multi-resolution load data representation. In *Proceedings of 5th IEEE International Conference on Smart Grid Communications (SmartGridComm 2014)*, pages 926–932, Venice, Italy, Nov. 2014. IEEE.

[90] E. L. Quinn. Privacy and the New Energy Infrastructure. *Social Science Research Network (SSRN)*, 2009.

[91] S. R. Rajagopalan, L. Sankar, S. Mohajer, and H. V. Poor. Smart meter privacy: A utility-privacy framework. In *Proc. IEEE Int Smart Grid Communications (SmartGridComm) Conf*, pages 190–195, 2011.

[92] S. Schinwald, D. Engel, and M. Seidler. Efficient automated liquid detection in microplates. In *Proceedings of 25th IEEE International Computer-Based Medical Systems (CBMS) Symposium*, pages 1–4, Rome, Italy, June 2012.

[93] R. Segovia and M. Sánchez. Set of common functional requirements of the smart meter. Technical Report 73, DG INFSO and DG ENER, European Commission, Brussels, 2011.

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

[95] H. K.-H. So, S. H. Kwok, E. Y. Lam, and K.-S. Lui. Zero-configuration Identity-based Signcryption Scheme for Smart Grid. In *Proceedings of First IEEE International Conference on Smart Grid Communications*, pages 321–326, Gaithersburg, Maryland, USA, 2010.

[96] S. Tan, D. De, W. Z. Song, J. Yang, and S. K. Das. Survey of security advances in smart grid: A data driven approach. *IEEE Communications Surveys Tutorials*, 19(1):397–422, Firstquarter 2017.

[97] A. Unterweger and D. Engel. Resumable load data compression in smart grids. *IEEE Transactions on Smart Grid*, 6(2):919–929, March 2015.

[98] A. Unterweger, D. Engel, and M. Ringwelski. The effect of data granularity on load data compression. *Springer Lecture Notes in Computer Science – Energy Informatics 2015*, 9424:69–80, 2015.

[99] A. Unterweger, F. Knirsch, G. Eibl, and D. Engel. Privacy-preserving load profile matching for tariff decisions in smart grids. *EURASIP Journal on Information Security*, 2016(1):1–17, 2016.

[100] E. Vogiatzis, G. Kalogridis, and S. Z. Denic. Real-time and low cost energy disaggregation of coarse meter data. In *4th IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, 2013.

[101] M. H. F. Wen, K.-C. Leung, and V. O. K. Li. Communication-oriented smart grid framework. In *Proc. IEEE Int Smart Grid Communications (SmartGridComm) Conf*, pages 61–66, 2011.

[102] Y. Yan, Y. Qian, H. Sharif, and D. Tipper. A survey on cyber security for smart grid communications. *IEEE Communications Surveys Tutorials*, 14(4):998–1010, 2012.

[103] M. Zeifman and K. Roth. Nonintrusive Appliance Load Monitoring: Review and Outlook. *IEEE Transactions on Consumer Electronics*, 57:76–84, 2011.

[104] J. Zhang and C. A. Gunter. Application-Aware Secure Multicast for Power-Grid Communications. In *Proceedings of First IEEE International Conference on Smart Grid Communications*, pages 339–344, Gaithersburg, Maryland, USA, 2010.

[105] Y. Zhang, L. Wang, W. Sun, R. C. Green, and M. Alam. Artificial immune system based intrusion detection in a distributed hierarchical network architecture of smart grid. In *2011 IEEE Power and Energy Society General Meeting*, pages 1–8, 2011.

[106] Y. Zhang, L. Wang, W. Sun, R. C. Green, and M. Alam. Distributed Intrusion Detection System in a Multi-Layer Network Architecture of Smart Grids. *IEEE Transactions on Smart Grid*, 2(4):796–808, 2011.

FULFILLMENT OF REQUIREMENTS 2

The University of Salzburg has published requirements for habilitation treatises in its official bulletin:

- In the statues ("Satzung") of the university, the principle process is outlined[1].
- In a separate document, guidelines regarding the requirements ("Habilitationsrichtlinie") are given[2].

On the basis of the more general documents published by the university, the Department of Computer Sciences has detailed the requirements in a separate document for the discpline of Informatics:

- In the document "Konkretisierung der Habilitationsrichtlinie an der PLUS vom 23.6.2015 für Habilitationen im Fach Informatik" the requirements of the university, especially in the area of "high ranking, refereed publications" are detailed.

The fulfillment of these requirements is detailed below.

2.1 PUBLICATIONS FOR CUMULATIVE TREATISE

In the following two subsections, the publications are listed by rating and author contribution. Subsequently, we will refer to the publications by the key, which is given in the first line as [*Key*], e.g., [Engel16a]. The fulfillment is discussed in all detail in the remaining subsections. In Table 2.1 on page 35, rating and author contributions are summarized.

[1] Bulletin no. 23, Academic Year 2015/16, January 26, 2016
[2] Bulletin no. 13, Academic Year 2015/16, December 1, 2015

2.1.1 *Overview by Rating*

*Rating A**

- [Engel16a]

  D. Engel and G. Eibl. Wavelet-based multiresolution smart meter privacy. *IEEE Transactions on Smart Grid*, PP(99):1–12, 2016. preprint.

- [Eibl15a]

  G. Eibl and D. Engel. Influence of data granularity on smart meter privacy. *IEEE Transactions on Smart Grid*, 6(2):930–939, March 2015.

- [Unterweger15a]

  A. Unterweger and D. Engel. Resumable load data compression in smart grids. *IEEE Transactions on Smart Grid*, 6(2):919–929, March 2015.

Rating A

- [Lausenhammer16a]

  W. Lausenhammer, D. Engel, and R. Green. Utilizing capabilities of plug in electric vehicles with a new demand response optimization software framework: Okeanos. *International Journal of Electrical Power and Energy Systems*, 75:1–7, 2016.

- [Lueckenga16a]

  J. Lückenga, D. Engel, and R. Green. Weighted vote algorithm combination technique for anomaly based smart grid intrusion detection systems. In *Proceedings of International Joint Conference on Neural Networks (IJCNN) 2016*, pages 2738–2742, Vancouver, Canada, July 2016.

- [Knirsch17a]

  F. Knirsch, G. Eibl, and D. Engel. Multi-resolution privacy-enhancing technologies for smart metering. *EURASIP Journal on Information Security*, 2017(1):6, 2017.

- [Knirsch15b]

  F. Knirsch, D. Engel, C. Neureiter, M. Frincu, and V. Prasanna. Model-driven privacy assessment in the smart grid. In *Proceedings of the 1st International Conference on Information Systems Security and Privacy (ICISSP)*, pages 173–181, Feb 2015. Best Paper Award.

Rating B

- [Engel11a]

  D. Engel. Conditional access smart meter privacy based on multi-resolution wavelet analysis. In *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies*, pages 45:1–45:5, New York, NY, USA, 2011. ACM.

- [Engel13a]

  D. Engel. Wavelet-based load profile representation for smart meter privacy. In *Proceedings IEEE PES Innovative Smart Grid Technologies (ISGT'13)*, pages 1–6, Washington, D.C., USA, Feb. 2013. IEEE.

- [Engel13e]

  D. Engel and G. Eibl. Multi-resolution load curve representation with privacy-preserving aggregation. In *Proceedings of IEEE Innovative Smart Grid Technologies (ISGT) 2013*, pages 1–5, Copenhagen, Denmark, Oct. 2013. IEEE.

- [Peer14a]

  C. Peer, D. Engel, and S. Wicker. Hierarchical key management for multi-resolution load data representation. In *Proceedings of 5th IEEE International Conference on Smart Grid Communications (SmartGridComm 2014)*, pages 926–932, Venice, Italy, Nov. 2014. IEEE.

- [Lausenhammer15a]

  W. Lausenhammer, D. Engel, and R. Green. A game theoretic software framework for optimizing demand response. In *Proceedings of the 6th Conference on Innovative Smart Grid Technologies (ISGT)*, pages 1–5, Feb 2015.

- [Eibl15b]

  G. Eibl, D. Engel, and C. Neureiter. Privacy-relevant smart metering use cases. In *Proceedings of IEEE International Conference on Industrial Technology (ICIT) 2015*, pages 1387–1392, Seville, Spain, 2015. IEEE.

- [Knirsch16a]

  F. Knirsch, D. Engel, C. Neureiter, M. Frincu, and V. Prasanna. Privacy assessment of data flow graphs for an advanced recommender system in the smart grid. In O. Camp, E. Weippl, C. Bidan, and E. Aïmeur, editors, *Information*

Systems Security and Privacy – Revised and Selected Papers of ICISSP 2015, volume 576 of *Communications in Computer and Information Science*, pages 89–106. Springer International Publishing, 2016. Best Paper Award.

- [Knirsch15a]

  F. Knirsch, D. Engel, M. Frincu, and V. Prasanna. Model-based assessment for balancing privacy requirements and operational capabilities in the smart grid. In *Proceedings of the 6th Conference on Innovative Smart Grid Technologies (ISGT)*, pages 1–5, Feb 2015.

Rating C

- [Eibl14a]

  G. Eibl and D. Engel. Influence of data granularity on nonintrusive appliance load monitoring. In *Proceedings of the Second ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec '14)*, pages 147–151, Salzburg, Austria, 2014. ACM.

- [Unterweger15b]

  A. Unterweger, D. Engel, and M. Ringwelski. The effect of data granularity on load data compression. *Springer Lecture Notes in Computer Science – Energy Informatics 2015*, 9424:69–80, 2015.

2.1.2 *Overview by Author Contribution*

Author Contribution > 50%

- Engel11a – 100%
- Engel13a – 100%
- Engel16a – 65%
- Engel14e – 60%

Author Contribution ≥ 30%

- Peer14a – 35%
- Eibl14a – 35%
- Eibl15a – 35%

- Unterweger15a – 35%
- Lausenhammer15a – 30%
- Lausenhammer16a – 30%
- Lueckenga16a – 30%
- Knirsch17a – 30%

Author Contribution < 30%

- Eibl15b – 15%
- Knirsch16a – 12%
- Knirsch15b – 11%
- Unterweger15b – 10%
- Knirsch15a – 7%

2.1.3 *Author Contribution to Publications*

According to the requirements specified by the department of Computer Sciences, the cumulative treatise needs to consists of at least 8 highly ranked, peer-reviewed publications. The author contribution needs to be at least 30% to count as a full publication. 4 publications can be substituted by publications where the author contribution is less than 30%, where the number of substituted publications needs to be higher corresponding to the author contribution.

> Typischerweise enthält die Sammelhabilitation mindestens 8 hochrangige referierte Publikationen (oder zur Publikation angenommene Manuskripte), bei denen der/die HabilitandIn einen wesentlichen eigenen Anteil beigesteuert hat. Dieser Anteil sollte bei mehreren Autoren mindestens 30% des methodisch/informatischen Beitrags der Publikation ausmachen. 4 dieser Publikationen können durch eine der Anzahl der beitragenden Autoren entsprechende höhere Anzahl von Publikationen ersetzt werden, bei denen der/die HabilitandIn einen geringeren eigenen Anteil beigesteuert hat. Andererseits sollten mindestens 2 Publikationen vorhanden sein, bei denen der eigene Anteil des/der HabilitationswerberIn 50% übersteigt.

> From: "Konkretisierung der Habilitationsrichtlinie an der PLUS vom 23.6.2015 für Habilitationen im Fach Informatik"

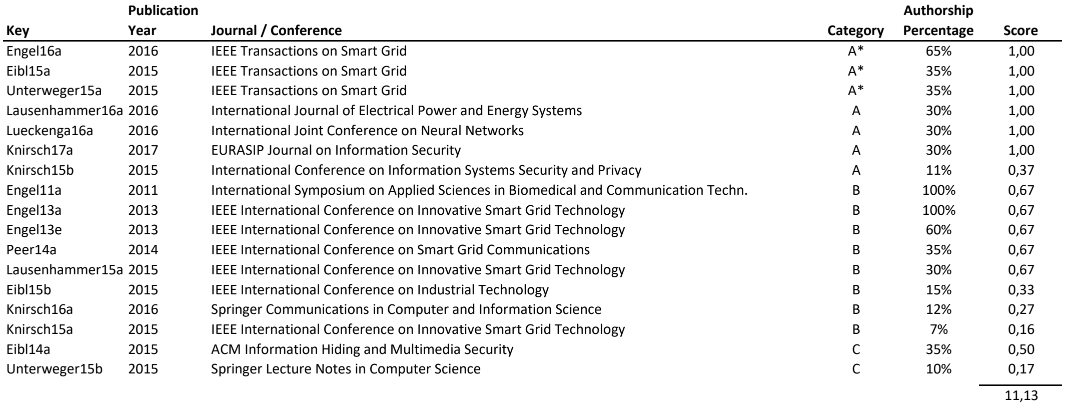
Futhermore, only publications of rating A*/A count as full publications. Publications of rating A*/A can be substituted with a higher number of publications of rating B or C, with a factor of 1.5 and 2, respectively. If the rating for a publication is not available in ERA – CORE, the rating needs to be determined using a comparable ranking (such as MS Academic Research Factors, comarable SCI JIF values, or similar acceptance rates).

> "Hochrangige referierte Beiträge" sind wie folgt definiert: Beiträge in Zeitschriften und Konferenzen die im jeweiligen Fachgebiet in dem die Habilitation erworben [wird] als "Top Qualität" bzw. "Sehr Gute Qualität" eingestuft werden. Beispielweise sind das Publikationen, die laut ERA – CORE (Excellence in Research for Australia) Ranking als A* bzw. A beurteilt werden. Bei Publikationen in den Kategorien B und C erhöht sich die Anzahl der notwendigen Publikationen um den Faktor 1,5 bzw. 2. Die Sammelhabilitation muss mindestens 2 Publikationen aus der Top Qualitätskategorie (z.B. ERA – CORE A*) und 2 aus der sehr guten Qualitätskategorie (z.B. ERA – CORE A) enthalten. Alternativ ist deren Qualität über vergleichbare andere Rankings (z.B. MS Academic Research Faktoren, vergleichbare SCI JIF Werte, vergleichbare Akzeptanzraten, etc.) nachzuweisen.
>
> From: "Konkretisierung der Habilitationsrichtlinie an der PLUS vom 23.6.2015 für Habilitationen im Fach Informatik"

In Table 2.1 the requirements for counting publications as contributions to the treatise have been formalized as follows: For a publication *x*, the rating factor R(x) is determined as follows, where E(x) is the rating for publication in the ERA–CORE Scale (A*, A, B, C) – either the direct rating or the derived rating based on a comparable score, as discussed in Section 2.1.4.

$$R(x) = \begin{cases} 1 & \text{for } E(x) = \text{A}^* \mid E(x) = \text{A} \\ \frac{2}{3} & \text{for } E(x) = \text{B} \\ \frac{1}{2} & \text{for } E(x) = \text{C} \end{cases} \tag{2.1}$$

The score $S(x)$ indicates the score for a publication x with which it can be counted toward fulfillment of the treatise requirement. A score of $S(x) = 1$ indicates that x counts as a full publication (i.e., is of rating A/A* and the author contribution is at least 30%) and a smaller score indicates a correspondingly smaller contribution. $C(x)$ indicates the author contribution for publication x.



Key	Publication Year	Journal / Conference	Category	Authorship Percentage	Score
Engel16a	2016	IEEE Transactions on Smart Grid	A*	65%	1,00
Eibl15a	2015	IEEE Transactions on Smart Grid	A*	35%	1,00
Unterweger15a	2015	IEEE Transactions on Smart Grid	A*	35%	1,00
Lausenhammer16a	2016	International Journal of Electrical Power and Energy Systems	A	30%	1,00
Lueckenga16a	2016	International Joint Conference on Neural Networks	A	30%	1,00
Knirsch17a	2017	EURASIP Journal on Information Security	A	30%	1,00
Knirsch15b	2015	International Conference on Information Systems Security and Privacy	A	11%	0,37
Engel11a	2011	International Symposium on Applied Sciences in Biomedical and Communication Techn.	B	100%	0,67
Engel13a	2013	IEEE International Conference on Innovative Smart Grid Technology	B	100%	0,67
Engel13e	2013	IEEE International Conference on Innovative Smart Grid Technology	B	60%	0,67
Peer14a	2014	IEEE International Conference on Smart Grid Communications	B	35%	0,67
Lausenhammer15a	2015	IEEE International Conference on Innovative Smart Grid Technology	B	30%	0,67
Eibl15b	2015	IEEE International Conference on Industrial Technology	B	15%	0,33
Knirsch16a	2016	Springer Communications in Computer and Information Science	B	12%	0,27
Knirsch15a	2015	IEEE International Conference on Innovative Smart Grid Technology	B	7%	0,16
Eibl14a	2015	ACM Information Hiding and Multimedia Security	C	35%	0,50
Unterweger15b	2015	Springer Lecture Notes in Computer Science	C	10%	0,17
					11,13

Table 2.1.: Scores counting towards the fulfillment of the treatise requirements for the individual publications, which the cumulative treatise is comprised of (sorted by rating and author contribution).

$$S(x) = \begin{cases} 1 & \text{for } C(x) \geq 0.3 \\ R(x)/0.3 & \text{for } C(x) < 0.3 \end{cases} \tag{2.2}$$

Table 2.1 shows the scores which count toward fulfillment of the treatise requirements for the individual publications, calculated using the formula above.

2.1.4 *Discussion of Ratings*

In the following, the assignment of publication category (A*, A, B, C) of the publication in the submitted treatise is discussed. The requirements state that the ERA – CORE (Excellence in Research for Australia) can be used, or alternatively the quality of the publications can be shown by comparable rankings (such as acceptance rate or SCI Journal Impact Factor JIF). For the purpose of this treatise, the publications of categories A* and A are discussed in more detail with screenshots of the measures from the sources of the rankings included in Appendix A.1. For categories B and C shorter justification are given.

For retrieving the ERA – CORE rankings, the following source was used: `http://mjolnir.lille.inria.fr/~roussel/rankings/era/`. For the Google Scholar h5-Index the following source was used: `https://scholar.google.at/citations?view_op=top_venues&hl=de`. For the Journal Impact Factor, the websites of the respective journals were used.

CATEGORIES A* AND A

- *IEEE Transactions on Smart Grids* was established after the ERA ranking had been discontinued. However, its scope is comparable to the *IEEE Transactions on Power Systems*, which is ranked in ERA as a journal of category A*. The impact factor of *IEEE Transactions on Smart Grids* at the time of publication of the papers included in this treatise was above the impact factor for *IEEE Transactions on Power Systems*, as shown in Figure A.1 on page 226. Therefore, the publications published in *IEEE Transactions on Power Systems* are considered as category A*.

- *International Journal of Electrical Power and Energy Systems* has an impact factor of 3.11, see Figure A.2 on page 227, which is a bit below *IEEE Transactions on Power Systems*, and therefore has been included in category A.

- *International Joint Conference on Neural Networks* is ranked by ERA with category A, see Figure A.3 on page 227.

- *EURASIP Journal on Information Security* does not yet have an impact factor and is not listed in ERA ranking. However, as the journal had an acceptance rate as low as 14% in 2015 (source: personal communication with editor-in-chief Prof. Katzenbeisser) the publications in this journal are assigned category A.

- *International Conference on Information Systems Security and Privacy* had an acceptance rate of 18% in 2015. Furthermore the one paper in this treatise that was published in the proceedings of this conference was awarded the Best Paper Award. Therefore, this paper (Knirsch15b) has been classified as category A.

CATEGORIES B AND C

- *IEEE International Conference Multimedia and Expo* is rated as category B in ERA.

- *International Symposium on Applied Sciences in Biomedical and Communication Technologies* has the same Google H5-Index (11, retrieved May 2017) as the *International Conference on Information Systems Security*, which is ranked as category B in ERA.

- *IEEE International Conference on Innovative Smart Grid Technologies* has a Google h5-Index of 30 (retrieved May 2017), which is above the h5-Index of conferences ranked as B in ERA (e.g., *International Conference on Information Systems Security*). Respective publications have been assigned category B.

- *IEEE International Conference on Innovative Smart Grid Technologies Europe* has a Google h5-Index of 21 (retrieved May 2017), which is also above the h5-Index of conferences ranked as B in ERA (e.g., *International Conference on Information Systems Security*). Respective publications have been assigned category B.

- *IEEE International Conference on Smart Grid Communications* has a Google h5-Index of 31. Respective publications have been assigned category B.

- *IEEE International Conference on Industrial Technology* has a Google Scholar h5-Index of 14, which is comparable to the *International Conference on Information Systems Security*, which is ranked as category B in ERA.

- *Springer Collections, e.g., Lecture Notes in Computer Science (LNCS)* are not rated in ERA and do not have an IF. The requirements of the Department for Computer Sciences mention LNCS explicitly as a renowned publication series, but do not assign a rating. Therefore, we determine the ranking for each of the two publications individually. *Knirsch16a* is an extended version of a conference paper that has won a best paper award at ICISSP 2015, which had an acceptance rate of 18% – we conservatively assign category B. *Unterweger15b* is a publication at a conference ("Energieinformatik" – "Energy Informatics"), for which the proceedings have been published in LNCS. In 2015, this conference had an acceptance rate of 50%, which is below "IEEE International Conference on Intelligent Computer Communication and Processing" (acceptance rate of 63%), which is ranked as category C in ERA. Therefore, Unterweger15b is assigned to category C.

- *ACM Information Hiding and Multimedia Security* is the merger of two conferences, one of which (ACM Information Hiding) is rated as category C in ERA. Therefore publications have been assigned to category C.

2.2 FURTHER PUBLICATIONS

In addition to the "core" publications counting towards the cumulative habilitation treatise, the candidate needs to deliver "further publications" from the general field of the treatise, but not directly related to the "core". The publications selected for this category are listed below. In Table 2.2 the corresponding scores are given, with the same definition of scores counting towards fulfillment (Equation 2.2) as used for the "core" publications.

Note that an exhaustive list of publications authored and co-authored by the candidate is contained in the CV in Section A.2.

> Im Sinn der in §1 vorgenommenen Spezifikation sollen 3 weitere "hochrangige referierte Beiträge" vorliegen, die überwiegend ausserhalb des Themas der Habilitationsschrift, aber im Fach, für das die venia docendi beantragt wird, liegen.
>
> From: "Konkretisierung der Habilitationsrichtlinie an der PLUS vom 23.6.2015 für Habilitationen im Fach Informatik"

The following publications form the part of "further publications" for this habilitation endeavour:

- [Knirsch16b]

  F. Knirsch, G. Eibl, and D. Engel. Error-resilient Masking Approaches for Privacy Preserving Data Aggregation. *IEEE Transactions on Smart Grid*, PP(99):1–12, 2016. Preprint.

- [Engel12a]

  D. Engel, T. Stütz, and A. Uhl. Assessing JPEG2000 encryption with key-dependent wavelet packets. *EURASIP Journal on Information Security*, 2012(1):1–16, 2012.

- [Unterweger16a]

  A. Unterweger, F. Knirsch, G. Eibl, and D. Engel. Privacy-preserving load profile matching for tariff decisions in smart grids. *EURASIP Journal on Information Security*, 2016(1):1–17, 2016.

- [Engel09a]

  D. Engel, T. Stütz, and A. Uhl. A survey on JPEG2000 encryption. *Multimedia Systems*, 15(4):243–270, 2009. Springer.

- [Engel09b]

  D. Engel, T. Stütz, and A. Uhl. Evaluation of JPEG2000 hashing for efficient authentication. In *Proceedings of International Conference on Multimedia & Expo, ICME '09*, pages 1728–1731, New York, NY, USA, June 2009.

- [Neureiter14a]

  C. Neureiter, D. Engel, J. Trefke, R. Santodomingo, S. Rohjans, and M. Uslar. Towards consistent smart grid architecture tool support: From use cases to visualization. In *Proceedings of IEEE Innovative Smart Grid Technologies (ISGT) 2014*, Istanbul, Turkey, Oct. 2014. IEEE.

- [NEUREITER14B]

  C. Neureiter, G. Eibl, D. Engel, S. Schlegel, and M. Uslar. A concept for engineering smart grid security requirements based on SGAM models. *Computer Science - Research and Development*, pages 1–7, 2014.

- [DAENEKAS14A]

  C. Dänekas, C. Neureiter, S. Rohjans, M. Uslar, and D. Engel. Towards a model-driven-architecture process for smart grid projects. In P. Benghozi, D. Krob, A. Lonjon, and H. Panetto, editors, *Digital Enterprise Design & Management*, volume 261 of *Advances in Intelligent Systems and Computing*, pages 47–58. Springer International Publishing, 2014.

- [EIBL17B]

  G. Eibl and D. Engel. Differential privacy for real smart metering data. *Computer Science – Research and Development*, 32(1):173–182, 2017.

- [SCHINWALD12A]

  S. Schinwald, D. Engel, and M. Seidler. Efficient automated liquid detection in microplates. In *Proceedings of 25th IEEE International Computer-Based Medical Systems (CBMS) Symposium*, pages 1–4, Rome, Italy, June 2012.

- [AUER13A]

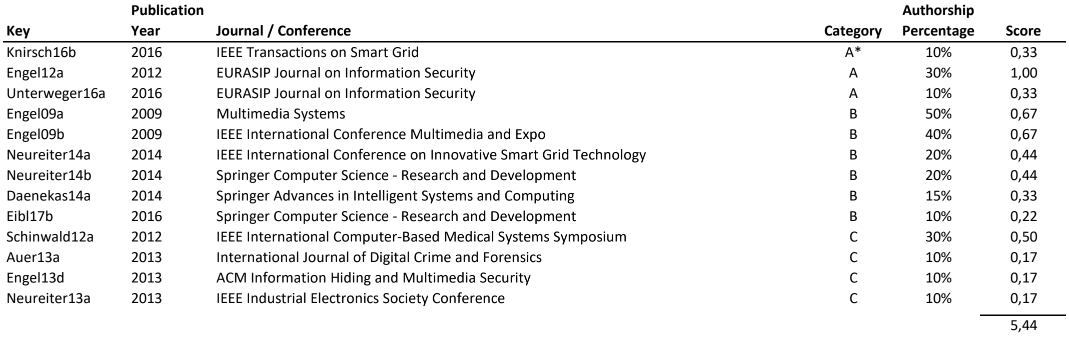
  S. Auer, A. Bliem, D. Engel, A. Uhl, and A. Unterweger. Bitstream-based JPEG encryption in real-time. *International Journal of Digital Crime and Forensics*, 5(3):1–14, 2013.

- [ENGEL13D]

  D. Engel, A. Uhl, and A. Unterweger. Region of interest signalling for encrypted JPEG images. In *Proceedings of the first ACM workshop on Information hiding and multimedia security (IHMMSEC '13)*, pages 165–174, Montpellier, France, 2013. ACM.

- [NEUREITER13A]

  C. Neureiter, G. Eibl, A. Veichtlbauer, and D. Engel. Towards a framework for engineering smart-grid-specific privacy requirements. In *Proc. IEEE IECON 2013, Special Session on Energy Informatics*, pages 4803 – 4808, Vienna, Austria, Nov. 2013. IEEE.



Key	Publication Year	Journal / Conference	Category	Authorship Percentage	Score
Knirsch16b	2016	IEEE Transactions on Smart Grid	A*	10%	0,33
Engel12a	2012	EURASIP Journal on Information Security	A	30%	1,00
Unterweger16a	2016	EURASIP Journal on Information Security	A	10%	0,33
Engel09a	2009	Multimedia Systems	B	50%	0,67
Engel09b	2009	IEEE International Conference Multimedia and Expo	B	40%	0,67
Neureiter14a	2014	IEEE International Conference on Innovative Smart Grid Technology	B	20%	0,44
Neureiter14b	2014	Springer Computer Science - Research and Development	B	20%	0,44
Daenekas14a	2014	Springer Advances in Intelligent Systems and Computing	B	15%	0,33
Eibl17b	2016	Springer Computer Science - Research and Development	B	10%	0,22
Schinwald12a	2012	IEEE International Computer-Based Medical Systems Symposium	C	30%	0,50
Auer13a	2013	International Journal of Digital Crime and Forensics	C	10%	0,17
Engel13d	2013	ACM Information Hiding and Multimedia Security	C	10%	0,17
Neureiter13a	2013	IEEE Industrial Electronics Society Conference	C	10%	0,17
					5,44

Table 2.2.: Scores counting towards the fulfillment of the treatise requirements for the individual publications comprising the set of "further publications" (sorted by rating and author contribution).

2.3 TEACHING

2.3.1 *Past Courses*

The applicant has taught courses at the Universities of Bremen and Salzburg, as well as the Salzburg University of Applied Sciences. In the following, a list of these courses is given, where "UPW" stands for "units per week" (in German: "Semesterwochenstunde") and refers to the number of units (45 minutes) taught per week per semester for a single course (1 UPW corresponds to a total of 14 units in a semester).

- "Mobile Networks and Security" (Lecture Part on IT-Security, 1 UPW), Salzburg University of Applied Sciences (since 2015)
- "Energy Informatics Fundamentals: Network and Communication Technologies" (Lecture, 1 UPW), Salzburg University of Applied Sciences (since 2015)
- "Network Reliability and Virtualization" (Lecture and Lab, 3 UPW), Salzburg University of Applied Sciences (since 2013)
- "Internet Infrastructure and Security" (Lecture and Lab, 3 UPW), Salzburg University of Applied Sciences (since 2013)
- "Cryptology" (Lecture, 1 UPW), Salzburg University of Applied Sciences (since 2006)
- "Master Seminar" (Lecture 1 UPW), Salzburg University of Applied Sciences (2011)

- “Network Reliability and Security” (Lecture and Lab, 3 UPW), Salzburg University of Applied Sciences (2011–2012)
- “Mobile & Distribution Networks” (Lecture and Lab, 3 UPW), Salzburg University of Applied Sciences (2010-2012)
- “Multimedia Technologies” (Lecture, 3 UPW and Lab, 2 UPW), Salzburg University of Applied Sciences (2010–2013)
- “Media Informatics” (Lecture, 3 UPW and Lab, 2 UPW), Salzburg University of Applied Sciences
- “Distributed and Autonomous Systems” (Lecture and Lab, 2 UPW), Salzburg University of Applied Sciences (2009)
- “Advanced Topics in Databases” (Lab, 1 UPW), University of Salzburg (2008)
- “Database Systems” (Lab, 2 UPW), University of Salzburg (2007–2008)
- “Introduction to Unix Systems” (Lab 1 UPW), University of Salzburg (2006–2008)
- “Software Project” (1 UPW), University of Bremen (2003–2004)
- “Software Development” (Lab, 1 UPW), University of Bremen (2003)

Student evaluation results for more than 50 courses taught at the Salzburg University of Applied Sciences in the degree program *Information Technologies and Systems Management* (ITS) are attached to this document. The attached detailed evaluation results not only include numerical results, but also students’ comments on the course. A summary of the results is included in this document in Table 2.3, which contains the mean of the numerical results over all evaluations for three evaluation categories. The three categories are the ones that directly relate to the course instructor, and include a total of 12 items, namely:

- Category “Organisation and structure of the course”
  - Overview of the goal and contents of the course
  - Communication of exam modalities
- Category “Instructor”
  - Expert knowledge
  - Quality of presentation

  - Clarity of explanation
  - Responsiveness to questions and suggestions
  - Fairness
- Category “Course Results”
  - Achievement of goals
  - Knowledge gain
  - Overall impression of the course

The students can grade each item on a scale from 1 to 5, where 1 stands for “Excellent” and 5 stands for ”Very Poor”. Table 2.3 for each course states the mean over all items in each category. Overall this covers responses by 263 students (out of 797 students, i.e., a response rate of a third). It can be clearly seen, that the candidate has consistently been evaluated very positively by the students.

2.3.2 *Future Courses*

University of Salzburg

For the University of Salzburg University, the following two advanced courses are envisioned.

PRIVACY ENHANCING TECHNOLOGIES: Privacy issues are gaining attention, especially in Europe. The new European General Data Protection Regulation (GDPR), which will be in effect from May 2018 in all EU member states, regulates privacy issues and prescribes state-of-the-art measures to be taken to protect personal and sensitive data. The GDPR also requires “privacy by default” and “privacy by design” in all (software) products dealing with such data. In this light, technical measures to safeguard privacy have come to the focus of many research groups. So-called privacy enhancing technologies allow to balance the individual need for privacy and the functional requirements of data processing tasks. In this course, we first review the requirements put forward by the GDPR and other regulation and we go through fundamental privacy principles. Moving to the technical side, we will first investigate anonymization and pseudonymization and discuss why both techniques fall short of providing proper privacy. We then investigate advanced technical methods to increase privacy in data processing systems. Specifically, we will discuss masking protocols, the principles of homomorphic encryption, various additively homomorphic encryption systems, secure aggregation by homomorphic encryption, differential privacy and the impact of data resolution on privacy. Finally, we will discuss the possibility of privacy measures

						Evaluation Results (1 -- Excellent to 5 -- Very Poor)		
Term	Year	Course Title	Type	Degree Program	Student Evaluation Response Rate	Organisation and Structure of the Course	Instructor (Knowledge, Rhetorics, Transfer, Q&A, Fairness)	Result (Goal, Knowledge Gain, Overall Impression)
WT	2010	Media Informatics	Lecture	ITS-Bachelor (FT)	21,1%	1,13	1,00	1,00
WT	2010	Media Informatics	Lab	ITS-Bachelor (FT)	15,8%	1,00	1,00	1,00
WT	2010	Mobile and Distribution Networks	Lecture and Lab	ITS-Master (FT)	77,8%	1,21	1,08	1,24
WT	2010	Media Informatics	Lecture	ITS-Bachelor (JC)	47,4%	1,00	1,00	1,15
WT	2010	Media Informatics	Lab	ITS-Bachelor (JC)	36,8%	1,00	1,00	1,00
WT	2010	Mobile and Distribution Networks	Lecture and Lab	ITS-Master (JC)	100,0%	1,14	1,08	1,34
ST	2011	Multimedia Technologies	Lecture	ITS-Bachelor (FT)	25,0%	1,00	1,00	1,00
ST	2011	Multimedia Technologies	Lab	ITS-Bachelor (FT)	6,3%	1,00	1,00	1,00
ST	2011	Network Reliability and Security	Lecture and Lab	ITS-Master (FT)	15,4%	1,00	1,00	1,00
ST	2011	Multimedia Technologies	Lecture	ITS-Bachelor (JC)	27,3%	1,00	1,00	1,00
ST	2011	Multimedia Technologies	Lab	ITS-Bachelor (JC)	27,3%	1,00	1,00	1,00
ST	2011	Cryptology	Lecture	ITS-Bachelor (JC)	26,5%	1,03	1,00	1,04
WT	2011	Media Informatics	Lecture	ITS-Bachelor (FT)	58,8%	1,13	1,04	1,13
WT	2011	Media Informatics	Lab	ITS-Bachelor (FT)	47,1%	1,20	1,08	1,19
WT	2011	Media Informatics	Lecture	ITS-Bachelor (JC)	54,5%	1,13	1,10	1,44
WT	2011	Media Informatics	Lab	ITS-Bachelor (JC)	45,5%	1,20	1,28	1,60
WT	2011	Master Seminar	Seminar	ITS-Master (JC)	48,3%	1,25	1,07	1,41
ST	2012	Multimedia Technologies	Lecture	ITS-Bachelor (FT)	11,1%	1,25	1,20	1,00
ST	2012	Network Reliability and Security	Lecture and Lab	ITS-Master (FT)	25,0%	1,08	1,07	1,00
ST	2012	Multimedia Technologies	Lecture	ITS-Bachelor (JC)	30,0%	1,00	1,07	1,00
ST	2012	Multimedia Technologies	Lab	ITS-Bachelor (JC)	20,0%	1,00	1,20	1,00
ST	2012	Cryptology	Lecture	ITS-Bachelor (JC)	20,8%	1,10	1,00	1,07
ST	2012	Network Reliability and Security	Lecture and Lab	ITS-Master (JC)	58,3%	1,04	1,03	1,24
WT	2012	Multimedia Technologies	Lecture	ITS-Bachelor (JC)	40,0%	1,44	1,10	1,42
WT	2012	Multimedia Technologies	Lab	ITS-Bachelor (JC)	30,0%	1,30	1,33	1,56
WT	2012	Internet Infrastructure and Security	Lecture and Lab	ITS-Master (JC)	28,6%	1,00	1,00	1,00
ST	2013	Multimedia Technologies	Lecture	ITS-Bachelor (FT)	21,4%	1,00	1,00	1,00
ST	2013	Network Reliability and Security	Lecture and Lab	ITS-Master (FT)	60,0%	1,10	1,00	1,00
ST	2013	Cryptology	Lecture	ITS-Bachelor (JC)	34,8%	1,18	1,00	1,20
ST	2013	Network Reliability and Security	Lecture and Lab	ITS-Master (JC)	66,7%	1,00	1,00	1,00
WT	2013	Media Informatics	Lecture	ITS-Bachelor (FT)	14,3%	1,00	1,00	1,00
WT	2013	Multimedia Technologies	Lab	ITS-Bachelor (FT)	14,3%	1,00	1,00	1,00
WT	2013	Internet Infrastructure and Security	Lecture and Lab	ITS-Master (FT)	16,7%	1,00	1,00	1,00
WT	2013	Multimedia Technologies	Lecture	ITS-Bachelor (JC)	36,4%	1,30	1,00	1,20
WT	2013	Multimedia Technologies	Lab	ITS-Bachelor (JC)	36,4%	1,20	1,10	1,10
WT	2013	Internet Infrastructure and Security	Lecture and Lab	ITS-Master (JC)	21,4%	1,00	1,00	1,10
ST	2014	Network Reliability and Security	Lecture and Lab	ITS-Master (FT)	20,0%	1,00	1,00	1,00
ST	2014	Cryptology (4th Semester)	Lecture	ITS-Bachelor (JC)	65,4%	1,10	1,10	1,20
ST	2014	Cryptology (6th Semester)	Lecture	ITS-Bachelor (JC)	25,8%	1,00	1,00	1,00
ST	2014	Network Reliability and Security	Lecture and Lab	ITS-Master (JC)	27,3%	1,00	1,00	1,20
WT	2014	Internet Infrastructure and Security	Lecture and Lab	ITS-Master (FT)	60,0%	1,20	1,00	1,00
ST	2015	Network Reliability and Security	Lecture and Lab	ITS-Master (FT)	60,0%	1,00	1,00	1,00
ST	2015	Cryptology	Lecture	ITS-Bachelor (JC)	35,3%	1,10	1,10	1,10
WT	2015	Network Reliability and Security	Lecture and Lab	ITS-Master (FT)	37,5%	1,30	1,10	1,30
WT	2015	Mobile Networks and Security	Lecture and Lab	ITS-Bachelor (JC)	45,5%	1,10	1,00	1,00
WT	2015	Foundations of Energy Informatics	Lecture	ITS-Master (JC)	50,0%	1,00	1,00	1,00
ST	2016	Cryptology	Lecture	ITS-Bachelor (FT)	19,7%	1,10	1,00	1,10
ST	2016	Network Reliability and Security	Lecture and Lab	ITS-Master (FT)	9,1%	1,30	1,00	1,00
ST	2016	Cryptology	Lecture	ITS-Bachelor (JC)	29,0%	1,10	1,00	1,00
ST	2016	Network Reliability and Security	Lecture and Lab	ITS-Master (JC)	27,3%	1,00	1,00	1,10
WT	2016	Internet Infrastructure and Security	Lecture and Lab	ITS-Master (FT)	35,7%	1,20	1,20	1,20
WT	2016	Mobile Networks and Security	Lecture	ITS-Master (JC)	50,0%	1,00	1,00	1,00
WT	2016	Internet Infrastructure and Security	Lecture and Lab	ITS-Master (JC)	16,7%	1,00	1,00	1,10
					Average	**1,09**	**1,04**	**1,11**

WT … Winter Term
ST … Summer Term

FT … Fulltime degree program
JC … Job-compliant degree program

Table 2.3.: Overview of Student Course Evaluation

and how users can best be communicated the consequences of disclosing personal data. This course could be taught as a lecture, but would also be well suited for the format of a seminar.

SMART GRID IT-SECURITY AND PRIVACY: In this course, we look at the move of traditional energy grids towards intelligent systems, so-called "smart grids", from the perspective of IT-security and privacy. The advancement in energy systems is an important enable for achieving the transformation of energy systems from fossil resources to renewable energy sources. However, the introduction of advanced information and communication technology to the energy grids also implicates new possible attack vectors, and also affects the privacy of end users' in many respects. We will first investigate the main use cases envisioned for smart grids in the distribution system: demand response management, direct load control, energy feedback, smart metering, home automation and electric vehical charging. We will then discuss the security implications for each of them, going through a detailed risk analysis. We will then discuss countermeasures that have been proposed by different organization in Europe (e.g., ENISA or CEN/CENELEC/ETSI) and see, how these countermeasures can be integrated into the larger picture of the systems architecture of a smart grid. We will then move the focus to privacy and discuss the data requirements of the aforementioned use cases: what kind of data in which (temporal and spatial) aggregation level and in what time interval is needed to fulfill the use cases? This insights will be counterposed to possible privacy implications: Given the data needed for the use cases, what other information could be deduced? Especially, can we deduce personal data such as lifestyle, personal preferences or even religion? Finally, we will look at privacy enhancing technologies that have specifically been suggested for smart grids to safeguard this personal information (at least to some degree) while still allowing the use cases to function.

Salzburg University of Applied Sciences

The courses at Salzburg University of Applied Sciences, centered on cryptology and security will be continued. In "Cryptology", an introduction to the foundations of cryptology is given. "Mobile Networks and Security" applies these foundation in actual protocols and security measures, such as TLS, IPSec and SSH. Both of these courses are Bachelor level courses. "Internet Infrastructure and Security" is a Master level course, which extends the basics to more comprehensive security approaches and architectures and deals with advanced security measures, such as anomaly detection for IDS systems, biometric authentication and blockchain technology.

2.4 SELECTED TALKS

The candidate has held a number of talks, including keynotes and invited talks. A selection is presented in the following.

- *The Interplay of Data Resolution and Privacy in Smart Metering*, Invited Talk, Department of Electrical Engineering, Cornell University, Ithaca, USA, 2017
- *The Interplay of Data Resolution and Privacy in Smart Metering*, Dagstuhl Seminar 16032 "Privacy and Security in Smart Energy Grids", `http://dx.doi.org/10.4230/DagRep.6.1.99`, Dagstuhl, Germany, 2016
- *Privacy-preserving Smart Metering: Methods and Applicability*, Keynote – Communications for Energy Workshop, Vienna, Austria, 2013
- *Privacy and Security Challenges in the Privacy and Security Challenges in the Smart Grid User Domain*, Keynote – 1st ACM Workshop on Information Hiding and Multimedia Security, Montpellier, France, 2013
- *Privacy Challenges in Smart Grids*, Panel Session on Smart Grid Security, IEEE ISGT EU 2014, Istanbul, Turkey
- Panelist Round Table *Sichere IKT Architektur im Smart Grid* (in German), Session "Sicherheit, Systemkontrolle und Versorgungssicherheit", Smart Grids Week, Salzburg, Austria, 2013
- *Datenschutz im Smart Metering: Herausforderungen und Lösungsansätze* (in German), VDE Smart Grid Forum,
  Hannover Messe (Industry trade show on industrial automation, energy, industrial supply and more), Germany, 2014
- Panelist Round Table *Smart Metering – hemmen Privacy Bedenken den technischen Fortschritt?* (in German), Session "Kunden und Märkte", Smart Grids Week, Graz, Austria, 2014
- *Status der europäischen Standardisierung für IT-Security und Privacy im Smart Grid* (in German), Österreichs Energie, Vienna, Austria, 2014
- *Sichere IKT-Architektur im Smart Grid* (in German), Österreichs Energie, Vienna, Austria, 2013
- *Datenschutz und -sicherheit im intelligenten Stromnetz* (in German), Lecture series "Anwendungen in Wirtschaft und Technik", University of Salzburg, Austria, 2013

- *Video Processing Activities and Applied Research at Sony DADC*, with M. Aster, Invited Talk – 6th International Symposium on Image and Signal Processing and Analysis (ISPA '09), Salzburg, Austria, 2009

2.5 FURTHER ASPECTS FOR REVIEW

According to the requirements, there are other aspects that positively influence the contribution of the candidate. The following of these aspects are fulfilled for the submitted application.

2.5.1 *H-Index*

A personal H-Index of the candidate ≥ 7, according to Google Scholar (without own references), is seen as positive for the review of the application. The H-Index of the candidate is 9 (according to Google Scholar Profile `https://scholar.google.at/citations?user=vbczhIkAAAAJ`, without own references).

2.5.2 *Development of Systems or Software which are Used in Practice*

A major part of this habilitation treatise has been developed in the context of the Josef-Ressel-Center Program. The central idea of this program is to bring together industry and academia to solve problems. The results of this thesis have been included in industry standardization (EU CEN/CENELEC/ETSI Mandate M/490) and have provided a basis for the tender for smart metering in Western Austria.

2.5.3 *Financial Support of Academic Activities*

As already mentioned, the research presented in this habilitation treatise is based on a 5-year research programme: *Josef Ressel Center for User-Centric Smart Grid Privacy, Security and Control*. The total funded budget for the research center is 800.000 €. Furthermore, based on the work and results presented in this thesis, publicly funded follow-up projects have been acquired with the Austrian Research Promotion Agency FFG (Project No.s 838793 – INTEGRA, 848811 – RASSA, 849914 – PROMISE).

2.5.4 *Support for Doctoral Theses*

In the context of this habilitation, the bases for three doctoral theses has been founded, for each of which the habilitation candidate is acting as co-supervisor.

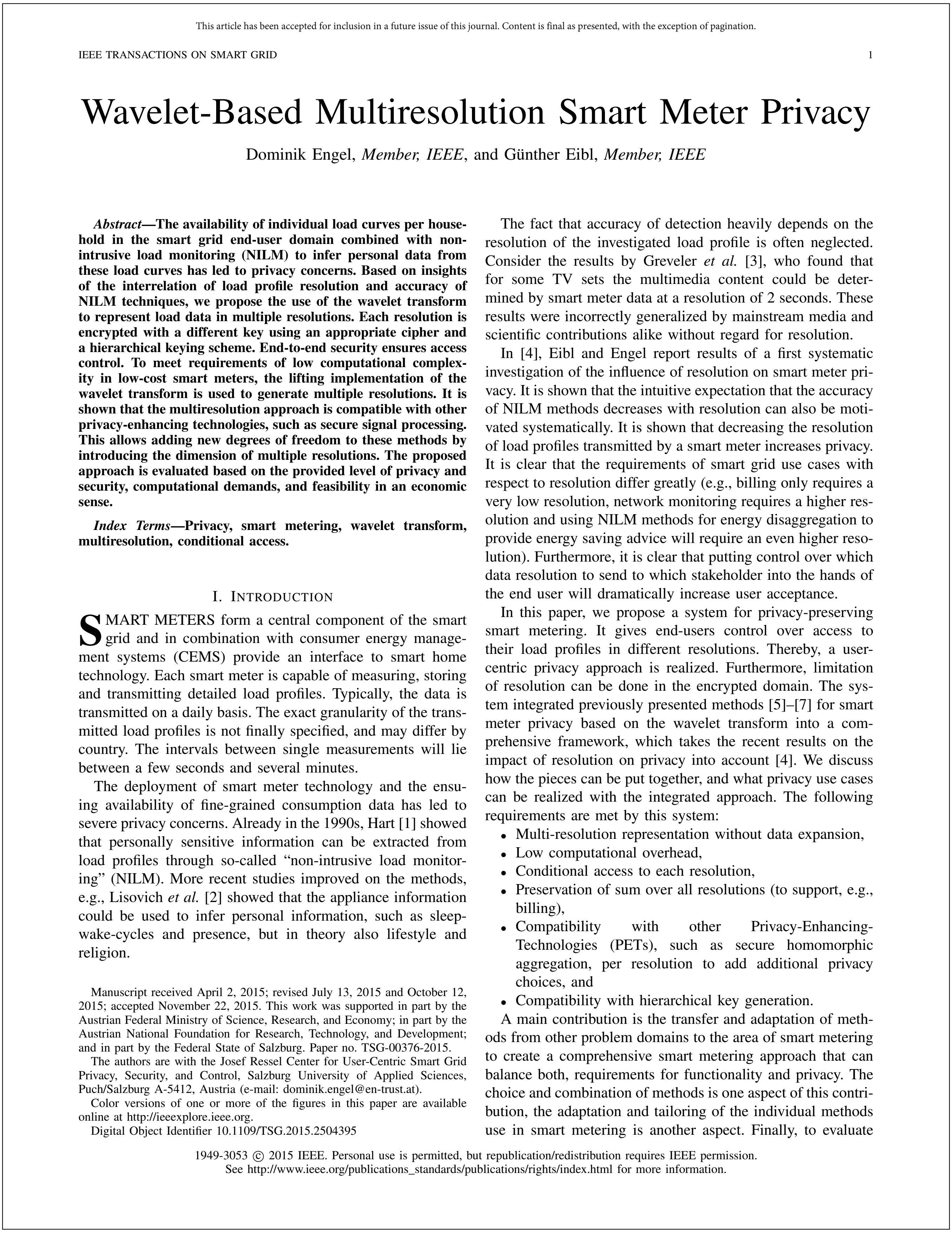
- Fabian Knirsch: *Privacy Enhancing Technologies in the Smart Grid User Domain* (University of Salzburg, Austria)
- Christian Neureiter: *A Domain-Specific, Model Driven Engineering Approach for Systems Engineering in the Smart Grid* (University of Oldenburg, Germany)
- Judith Schwarzer: *Modellierung von Benutzerakzeptanz und -interaktion in Demand Response Szenarien* (University of Oldenburg, Germany)

PUBLICATIONS 3

In the following, the papers that cumulatively consitute this treatise are printed in the form of their original publication.

3.1 ENGEL16A

- D. Engel and G. Eibl. Wavelet-based multiresolution smart meter privacy. *IEEE Transactions on Smart Grid*, PP(99):1–12, 2016. preprint.



Wavelet-Based Multiresolution Smart Meter Privacy

Dominik Engel, *Member, IEEE*, and Günther Eibl, *Member, IEEE*

***Abstract*—The availability of individual load curves per household in the smart grid end-user domain combined with non-intrusive load monitoring (NILM) to infer personal data from these load curves has led to privacy concerns. Based on insights of the interrelation of load profile resolution and accuracy of NILM techniques, we propose the use of the wavelet transform to represent load data in multiple resolutions. Each resolution is encrypted with a different key using an appropriate cipher and a hierarchical keying scheme. End-to-end security ensures access control. To meet requirements of low computational complexity in low-cost smart meters, the lifting implementation of the wavelet transform is used to generate multiple resolutions. It is shown that the multiresolution approach is compatible with other privacy-enhancing technologies, such as secure signal processing. This allows adding new degrees of freedom to these methods by introducing the dimension of multiple resolutions. The proposed approach is evaluated based on the provided level of privacy and security, computational demands, and feasibility in an economic sense.**

***Index Terms*—Privacy, smart metering, wavelet transform, multiresolution, conditional access.**

I. Introduction

SMART METERS form a central component of the smart grid and in combination with consumer energy management systems (CEMS) provide an interface to smart home technology. Each smart meter is capable of measuring, storing and transmitting detailed load profiles. Typically, the data is transmitted on a daily basis. The exact granularity of the transmitted load profiles is not finally specified, and may differ by country. The intervals between single measurements will lie between a few seconds and several minutes.

The deployment of smart meter technology and the ensuing availability of fine-grained consumption data has led to severe privacy concerns. Already in the 1990s, Hart [1] showed that personally sensitive information can be extracted from load profiles through so-called “non-intrusive load monitoring” (NILM). More recent studies improved on the methods, e.g., Lisovich *et al.* [2] showed that the appliance information could be used to infer personal information, such as sleep-wake-cycles and presence, but in theory also lifestyle and religion.

The fact that accuracy of detection heavily depends on the resolution of the investigated load profile is often neglected. Consider the results by Greveler *et al.* [3], who found that for some TV sets the multimedia content could be determined by smart meter data at a resolution of 2 seconds. These results were incorrectly generalized by mainstream media and scientific contributions alike without regard for resolution.

In [4], Eibl and Engel report results of a first systematic investigation of the influence of resolution on smart meter privacy. It is shown that the intuitive expectation that the accuracy of NILM methods decreases with resolution can also be motivated systematically. It is shown that decreasing the resolution of load profiles transmitted by a smart meter increases privacy. It is clear that the requirements of smart grid use cases with respect to resolution differ greatly (e.g., billing only requires a very low resolution, network monitoring requires a higher resolution and using NILM methods for energy disaggregation to provide energy saving advice will require an even higher resolution). Furthermore, it is clear that putting control over which data resolution to send to which stakeholder into the hands of the end user will dramatically increase user acceptance.

In this paper, we propose a system for privacy-preserving smart metering. It gives end-users control over access to their load profiles in different resolutions. Thereby, a user-centric privacy approach is realized. Furthermore, limitation of resolution can be done in the encrypted domain. The system integrated previously presented methods [5]–[7] for smart meter privacy based on the wavelet transform into a comprehensive framework, which takes the recent results on the impact of resolution on privacy into account [4]. We discuss how the pieces can be put together, and what privacy use cases can be realized with the integrated approach. The following requirements are met by this system:

- Multi-resolution representation without data expansion,
- Low computational overhead,
- Conditional access to each resolution,
- Preservation of sum over all resolutions (to support, e.g., billing),
- Compatibility with other Privacy-Enhancing-Technologies (PETs), such as secure homomorphic aggregation, per resolution to add additional privacy choices, and
- Compatibility with hierarchical key generation.

A main contribution is the transfer and adaptation of methods from other problem domains to the area of smart metering to create a comprehensive smart metering approach that can balance both, requirements for functionality and privacy. The choice and combination of methods is one aspect of this contribution, the adaptation and tailoring of the individual methods use in smart metering is another aspect. Finally, to evaluate

Manuscript received April 2, 2015; revised July 13, 2015 and October 12, 2015; accepted November 22, 2015. This work was supported in part by the Austrian Federal Ministry of Science, Research, and Economy; in part by the Austrian National Foundation for Research, Technology, and Development; and in part by the Federal State of Salzburg. Paper no. TSG-00376-2015.

The authors are with the Josef Ressel Center for User-Centric Smart Grid Privacy, Security, and Control, Salzburg University of Applied Sciences, Puch/Salzburg A-5412, Austria (e-mail: dominik.engel@en-trust.at).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSG.2015.2504395

the effectiveness of the approach, a privacy measure for smart metering is introduced.

The remainder of this paper is organized as follows: Section II gives an overview of related work and discusses suggestions for other privacy-enhancing technologies. The impact of resolution on privacy is reviewed in Section III-A. Section III contains all details on the proposed wavelet-based approach for smart meter privacy. This section also shows that the used wavelet transform preserves the sum over all resolutions, which is an important property for use cases like billing. The proposed approach is evaluated in Section IV. In Section V the compatibility with other PETs is discussed. Section VI concludes and gives an outlook on future work.

II. Related Work

There are a number of contributions for privacy-enhancing technologies for smart metering. Jawurek *et al.* [8] argue, that approaches relying on policy alone, may prove inadequate to provide a sufficient level of privacy and that technological methods that enforce privacy by virtue of "strength of mechanism" need to be employed. Indeed, a number of such technological approaches have been suggested to remedy the loss in privacy and still enable smart metering functionality on a broad basis. In the following, we give a brief overview of these contributions, based on [9]. More detailed surveys can be found in [8], [10], and [11].

The only approach that is widely used in the real world at this point in time, is *anonymization or pseudonymization* of smart metering data. Consumption data and the personal data are split and stored separately. Methods for de-anonymization are a major threat for these types of approaches. It has been shown that even after anonymization or pseudonymization, data items can still be attributed to the individual that originated them. Jawurek *et al.* [12] show that de-anonymization can also be done in the smart grid user domain. This structural traceability is a problem for schemes that rely on anonymization or pseudonymization only without the use of additional encryption.

Simple aggregation tries to hide data related to individuals by aggregating over a number of house-holds, e.g., all households in a neighborhood are network (NAN). For example, Bohli *et al.* [13] propose a privacy scheme in which high resolution smart meter readings are aggregated at NAN level and only the aggregate is sent to the utility. They introduce two solutions both with and with-out involvement of trusted third parties.

Due to the inherent link between load data resolution and privacy, splitting the load data into a variety of *different resolutions*, each associated with different authorization levels, has been proposed by a number of contributions. For example, the anonymization scheme proposed by Efthymiou and Kalogridis [14] is based on two different resolutions: a low resolution that can be used for billing purposes, and a high resolution that allows further investigation. This scheme employs a trusted third party escrow service. In the manuscript presented here, we build on previous work on wavelet-based multi-resolution privacy [6], [7].

Masking relates to approaches which add numerical artifacts, e.g., random sequences to the original load data to obfuscate individual contributions. The added artifacts are constructed in such a way that they cancel each other out upon aggregation. The aggregator can therefore combine the data values of all participant to create an accurate aggregation, but cannot gain access to individual contribution. For example, Kursawe *et al.* [15] propose such an aggregation protocol, which compared to other approaches has the advantage of relatively low computational complexity. Defend and Kursawe [16] further improve on this idea. Danezis *et al.* [17] present another low-overhead protocol for aggregation of smart meter data, which puts minimal computational demands on the smart meter hardware.

Differential privacy, as Dwork [18, p. 1] puts it, roughly speaking, "ensures that (almost, and quantifiably) no risk is incurred by joining a statistical database". Adding or removing an item from the database will not (or only to a very limited degree) affect the result of statistical computations. This is commonly achieved by the distributed generation of noise which is added to the individual data contribution. Shi *et al.* [19] propose a scheme for adding random noise to time series data using a symmetric geometric distribution. An advantage of this scheme is that the participants need not trust each other, nor rely on a trusted aggregator. As another example, Ács and Castelluccia [20] obscure individual data sets by adding Laplacian noise, which is jointly generated by the participants. Apart from the obvious drawback that the data is no longer exact after differential privacy is applied, data pollution by malicious participants is another issue with this approach [19].

Secure Signal Processing (SSP) refers to the possibility to perform certain computations, such as aggregation in the encrypted domain. A commonly employed mechanism in SSP is homomorphic encryption, which allows some specific manipulations of the ciphertext to be reflected in the plaintext domain. For example, Li *et al.* [21] propose an overlay network in a tree-like topology and the use of a Paillier cryptosystem. Garcia and Jacobs [22] combine secret sharing with a Paillier cryptosystem to add flexibility in the aggregation (at the expense of additional computational complexity). Erkin and Tsudik [23] extend the idea of homomorphic encryption of smart meter readings by splitting the module into random shares, which, in combination with a modified Pailler cryptosystem, allows flexible spatial and temporal aggregation for different use cases, such as billing or network monitoring.

III. Wavelet-Based Smart Meter Privacy

A. Motivation for Multi-Resolution Privacy

The basis for both, regulatory-based and technology-based approaches, is detailed knowledge of what information can be extracted from the available user data. To date, there is little systematic research on this subject in the context of smart grids.

In [24], Molina-Markham *et al.* investigate the information revealed from load profiles at different granularities.

They show that with off-the shelf statistical methods detailed information on the behavior of users can be inferred from load profiles without prior knowledge or precomputed appliance signatures. They argue that "the information leaks directly correlate with the time granularity that a meter measures power consumption" [24, p. 61] and list a number of privacy-relevant questions that can be answered using load profiles at granularities ranging from hours to seconds.

In [4], Eibl and Engel study the impact of resolution on NILM methods systematically. The authors use the so-called F-Score, a combination for accuracy and precision to evaluate appliance detection in a number of different resolution. The approach is empirically evaluated using the publicly available REDD energy disaggregation data set [25]. The results clearly and systematically show that appliance detection accuracy decreases with resolution.

Based on these insights, it is only logical to aim at differentiating access control to load profile by resolution. The classical wavelet transformation in the lifting implementation is the ideal tool to create integrated, multi-resolution representations of load profiles. In contrast to the Fourier transform, which only provides localization in the frequency domain, the wavelet transform allows to strike a balance between localization in the time domain and the frequency domain. This allows to create a cascade of lower resolution representations of the original sequence, each of which exactly correspond to a subsampled version of the previous resolution with half its size. Each resolution contained in the multi-resolution load profile can be tailored to correspond to a class of detection accuracy (as will be discussed in Section IV-B). Granting access to third party based on this multi-resolution representation allows informed, privacy-aware data exchange to the user.

Multi-resolution privacy, as proposed here can be seen as orthogonal to many of the approaches mentioned in Section II, i.e., approaches such as secure signal processing or masking can be combined with and enhanced by multi-resolution analysis. This leads to more possibilities that can be presented to the user, e.g., a higher resolution could be restricted to be communicated over secure aggregation only, while a lower resolution could be communicated as an individual load profile to a defined external party through end-to-end security. Note that the approach proposed here supports both, billing and aggregation use-cases (which are often distinguished in literature). The combination of multi-resolution privacy with selected other PETs is discussed in more detail in Section V.

B. Multi-Resolution Load Profile Representation

Let $L[i]$ be a tuple of length n, with $i = 1, \dots, n$, containing the data values of the original load profile (without loss of generality, we assume these data values to be of type float). For the sake of readability, we omit the index and write $\mathbf{L}$ whenever we do not need to refer to individual elements of $L[i]$.

A wavelet transform of maximum depth d, denoted as $W_d(\mathbf{L})$, is applied to the original load profile $\mathbf{L}$ by iteratively applying the transformation in d steps to the resulting low pass subbands: In each step k, for $k = 1, \dots, d$, the wavelet transform operates on data of resolution $r := d - k$ and produces the next lower resolution, which is half the size of the data at resolution r. As illustrated in Fig. 1, in each step k, a low-pass subband $\mathbf{L}_{r-1}$ and a high-pass subband $\mathbf{H}_r$ are produced in this way, each half of the size of the input data, i.e., the number of samples after step k is $n \cdot 2^{-k}$. The coefficients contained in $\mathbf{H}_r$ are stored as the wavelet coefficients of resolution r. Then the wavelet transform is applied to $\mathbf{L}_{r-1}$ to produce the next lower resolution, until the lowest resolution $r = 0$ is reached. At the end of the transformation the sequence $\mathbf{L}_0$, $\mathbf{H}_1, \dots,$ $\mathbf{H}_d$ is obtained, see Fig. 1. In the following, we use $\mathbf{w}_i$ as a short hand to refer to each of the individual subbands $\mathbf{L}_0, \mathbf{H}_1, \dots, \mathbf{H}_r\}$, i.e., $\mathbf{w}_0 := \mathbf{L}_0$, $\mathbf{w}_1 := \mathbf{H}_1, \dots, \mathbf{w}_d := \mathbf{H}_d$. Furthermore, we denote the *all* wavelet coefficients necessary for a resolution r by $\mathbf{W}_r$, i.e., $\mathbf{W}_r := \{\mathbf{L}_0, \mathbf{H}_1, \dots, \mathbf{H}_r\}$. For a wavelet transform of depth d, the maximum (i.e., original) resolution is d.

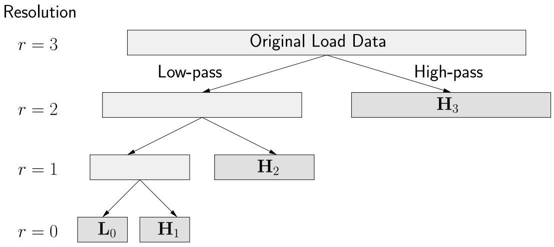


Fig. 1. Wavelet transformation of a smart meter load profile.

The synthesis step of the inverse wavelet transform W_d^{-1} starts with the lowest resolution $r = 0$. To get the next higher resolution of the signal the next higher resolution subband is needed, so that in a series of d steps one finally obtains the original load curve (since we only consider lossless transformations).

In order to provide a load profile $\mathbf{L}_t$ with a target resolution r from coefficients $\mathbf{W}_d$ (where $r \leq d$) only r synthesis steps need to be performed and only the subbands with resolution $t \leq r$, i.e., $\mathbf{W}_r$ are needed. Denoting the selection of the lower r resolutions from $\mathbf{W}_d$ as $T_r(\cdot)$ this can be written as

$$\mathbf{L}_r = W_r^{-1}(T_r(\mathbf{W}_d)) \tag{1}$$

The operator T_r can be generalized to be any linear transformation T of the wavelet coefficients to be used for example for denoising, which could be valuable for transmission of signal aggregations. Representing load profiles in the wavelet domain can also be a basis for signal processing [26] and data compression [27], [28]. As these aspects are out of the scope of this paper, the combination of signal processing and compression with the privacy approach proposed here will be addressed by future work. Another item to be considered for future work is the use of non-uniform sampling: for areas where the signal is smooth the sampling interval could be increased adaptively, whereas for areas where the signal is less smooth, the sampling interval could be increased – in this way privacy could be increased while still retaining the same average sampling rate.

The discrete wavelet transform does not lead to data expansion, and by using a lifting implementation the transformation can even be done in-place: "The lifting scheme also leads to a fast in-place calculation of the wavelet transform, i.e., an implementation that does not require auxiliary memory." [29, p. 4]. Therefore, the number of bits needed to represent the coefficient data of a level d wavelet transform, $\mathbf{W}_d$, is the same number of bits needed to represent the original load data L.

The wavelet coefficients of the different subbands can be represented in a single, embedded bitstream, which correspondingly contains all resolutions. Note that if the appropriate filter is used, the wavelet transform is lossless, i.e., no data loss occurs and the original load curve can be recovered perfectly from the coefficients contained in the embedded bitstream.

To implement multi-resolution analysis in a manner that is suitable for smart metering devices, wavelet lifting [30] provides a helpful perspective: This view on the wavelet transform factors wavelet filters into lifting steps, which for many filters rely on simple operations only.

C. Applying the Haar Wavelet to Load Profiles

In this paper, we use the simple Haar wavelet filter to create multi-resolution load profiles. Other filters have been studied by Engel in [6], where the author came to the conclusion that the Haar wavelet is sufficient for all currently envisioned use cases and at the same time has the important property of very low computational complexity: The Haar wavelet filter realizes low-pass filtering as averaging of the sample values. The high-pass step is realized by calculating the corresponding differences to allow for lossless reconstruction.

Let $\mathbf{L}_r$ be the input signal at resolution r, and $\mathbf{L}_{r-1}$ and $\mathbf{H}_r$ be the low-pass and high-pass output signals, respectively. Further let n be the length of $\mathbf{L}_r$, and for the sake of simplicity, let $n \bmod 2 = 0$. The lifting steps for the forward transform (going from resolution r to the next lower resolution $r-1$) with the Haar wavelet can be written as follows (adapted from [30]):

$$\hat{L}_{r-1}[i] = L_r[2i] \tag{2}$$

$$\hat{H}_r[i] = L_r[2i+1] \tag{3}$$

$$H_r[i] = \hat{H}_r[i] - \hat{L}_{r-1}[i] \tag{4}$$

$$L_{r-1}[i] = \hat{L}_{r-1}[i] + \frac{1}{2}(H_r[i]), \tag{5}$$

with $i = 1, \ldots, \frac{n}{2}$. The inverse transform (going from resolution $r-1$ to $r-1$) correspondingly is given as

$$\hat{L}_{r-1}[i] = L_{r-1}[i] - \frac{1}{2}(H_r[i]) \tag{6}$$

$$\hat{H}_r[i] = H_r[i] + \hat{L}_{r-1}[i] \tag{7}$$

$$L_r[2i+1] = \hat{H}_r[i] \tag{8}$$

$$L_r[2i] = \hat{L}_{r-1}[i] \tag{9}$$

again with $i = 1, \ldots, \frac{n}{2}$.

This transformation is lossless, and the applied operations in each step are equivalent to subsampling. In effect, each iteration of applying the Haar wavelet is equivalent to halving the sampling rate.

The Haar wavelet filter perfectly preserves the first moment in each step of the transform. The sum of the original load profile (i.e., the total consumption) can be accessed at any resolution, because, as can be easily shown, the sum for the load profile at any resolution r can be obtained from the next lower resolution $r-1$ as follows:

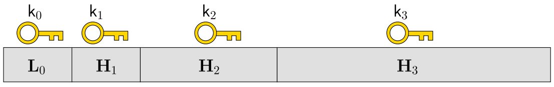


Fig. 2. Final encrypted bitstream produced by the smart meter: The wavelet coefficients of each subband are encrypted with an individual key.

$$\sum_{i=1}^{n} L_r[i] = 2 \cdot \sum_{j=1}^{n/2} L_{r-1}[j]. \tag{10}$$

This is an important property, as it allows the use of any lower resolution for functions like accurate billing, as the sum of the original sequence can be derived from any of the lower resolutions.

D. User-Centric Privacy Through Conditional Access

The idea of conditional access stems from the context of multimedia entertainment data. Entertainment content usually exists in various resolutions (e.g., mobile content, standard definition, high definition), which may be priced differently. A multi-resolution representation of the multimedia data allows the efficient representation of the resolutions in a single bitstream. This is an advantage as only one version of the bitstream needs to be handled and transmitted. Conditional access allows users to pay only for the resolutions they are interesting in. For example, the owner of a standard definition television has no need to pay for the high-definition version of the content. Through conditional access only the bitstream portion relevant for the desired resolution is decrypted, the rest of the bitstream is ignored.

We propose to use the conditional access paradigm for smart metering data in multi-resolution representation. Each subband $\mathbf{L}_0, \mathbf{H}_1, \ldots, \mathbf{H}_d$ is encrypted with a different key (key generation and handling are discussed in Section III-E). The whole datastream is transmitted over a Smart Grid communication infrastructure. Access to the different resolutions is thereby only granted to parties that hold the needed keys, as illustrated by Figure 2.

This scheme allows flexible control by the end-user how access is granted to smart meter data. For example, a particular energy provider may be granted access only to the lowest resolution for billing purposes. A third-party service providing energy saving advice by employing NILM methods may be granted access to the highest resolution by the user. The end-user may further be willing to provide data for network monitoring, but only at a medium resolution. Note that the approach presented here provides all the necessary means and ingredients for multi-resolution privacy, but it does not make the decision on which resolution to choose on the users' behalf. This part could be provided, e.g., by a recommender system, which advises the user on the privacy implications of a certain resolution. First steps into such automated privacy recommendations have been made in [31].

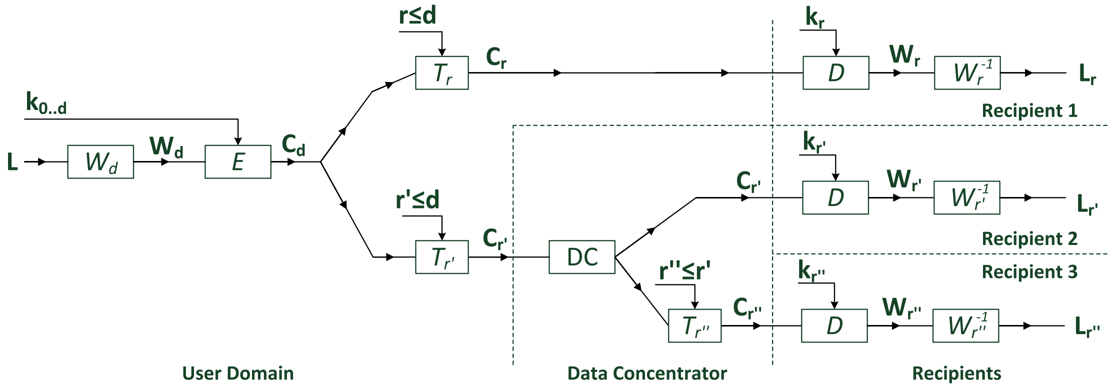


Fig. 3. Illustration of Wavelet-based Multi-Resolution Privacy: Three communication paths from user domain to recipient.

T_t (where $0 \leq t \leq R$), i.e., decreasing the resolution, can be applied before or after encryption. Note that no keys are required to decrease the resolution of the encrypted bitstream, as the operation is a simple truncation: For reducing the resolution from r to $r-1$, discarding the bits of the encrypted coefficients of $\mathbf{H}_r$ is sufficient. As each resolution is encrypted separately, the boundaries are known and the truncation can be done in the encrypted domain (if a cipher is used that preserves the number of bits, the truncation points are known implicitly, otherwise explicit markers can be introduced).

This allows rate adaptation not only by the users themselves, but also at a later point, e.g., through a third party (no keys need to be provided for this third party). In this way, the proposed scheme also enables relaying of data in various resolution. An illustrative example is discussed in Section III-F.

E. Hierarchical Key Generation

In [32], a scheme for generating keys for multi-resolution privacy is proposed, based on previous suggestions stemming from multimedia security (e.g., [33]). The original idea is due to Lamport [34], who already proposed the underlying method in 1981.

For each resolution r, the key k_{r-1} for the next lower resolution $r-1$ is obtained by using a cryptographic one-way hash function h on k_r:

$$\mathsf{k}_{r-1} = h(\mathsf{k}_r). \tag{11}$$

In this way, for each resolution r, all lower resolution keys $\mathsf{k}_{r-1}, \mathsf{k}_{r-2}, \ldots, \mathsf{k}_0$ can always be obtained. Using this hierarchical key generation scheme saves overhead in key management, as only a single key needs to be stored and transmitted.

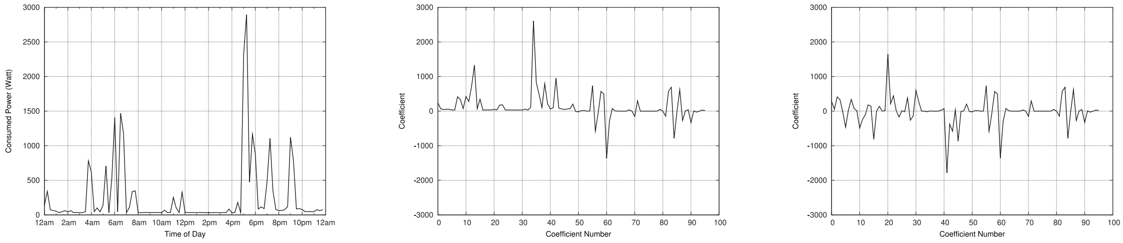
By using a secure cryptographic one-way hash function, the one-way property ensures that inferring keys for higher resolutions from a lower resolution key is extremely difficult.

Note that the hierarchical keys are generated for a symmetric scheme, such as AES. Access to the different resolution to different stakeholders is granted by using the public keys of these stakeholders. For examples, if a user wants to grant access to an external party to resolution $r = 2$, the user encrypts k_2 with the public key pk of this external party, producing a "wrapped" key $\mathsf{wk} = E_{\mathsf{pk}}(\mathsf{k}_2)$. The external party can use its private key sk to obtain the symmetric key: $D_{\mathsf{sk}}(\mathsf{wk}) = \mathsf{k}_2$. Subsequently, keys k_1 and k_0 can be derived by using Equation 11. With these keys, all wavelet coefficients of $\mathbf{L}_0, \mathbf{H}_1$ and $\mathbf{H}_2$ can be decrypted. Finally, the inverse wavelet transform is applied to obtain the load profile in the target resolution.

F. Illustration

Figure 3 illustrates the proposed method. In the user domain, a wavelet transform of depth d is applied to the original load profile $\mathbf{L}$, resulting in the wavelet coefficients $\mathbf{W}_d$. The coefficients for each resolution $r = 0, \ldots, d$ are encrypted with a unique key k_r (using the hierarchical key scheme, as discussed above). Access to different resolutions is granted to recipients based on these keys. In the illustration, there are three recipients, each of which is granted access to a different resolution, r, r', and r'', respectively. Access to resolution r, the highest resolution, is granted to Recipient 1, which could be a third party service provider for energy optimization through NILM. Resolutions r' and r'' are relayed over a data concentrator (typically operated by the DSO). Recipient 2 could be the DSO itself, which is granted access to resolution r' to use the data at this resolution for demand prognosis. Recipient 3 could be the energy provider, which receives a very low resolution r'' through the DSO's data concentrator for billing purposes. (Note that the use of a data concentrator is possible with the proposed scheme, but not required. The regulation in some countries prescribes the role of a data concentrator, in other countries no data concentrators are used.)

Before the data leaves the user domain, by applying T the user can decrease the data to be transmitted. In the illustration, there are two communication channels which leave the user domain: the upper communication channel is a *direct channel* to Recipient 1 (e.g., via the user's Internet connection). For this recipient, the user grants access to resolution r by providing key k_r. Furthermore, only the coefficient data up to resolution r needs to be transmitted. This can be achieved



(a) Original Data (b) Level 1 Decomposition (c) Level 5 Decomposition

Fig. 4. Example for Wavelet Decomposition of Smart Meter Data.

by creating $\mathbf{C}_r$ through applying T_r to $\mathbf{C}_d$. Recipient 1 can decrypt $\mathbf{W}_r$ from C_r and by applying the inverse wavelet transform can reconstruct $\mathbf{L}_r$.

The lower communication channel leaving the user domain transmits encrypted wavelet coefficient data to Recipient 2 and Recipient 3, via a *data concentrator* (DC). As the maximum resolution for all recipients behind the DC is r', the transmitted data can be reduced to $C_{r'}$ before it is passed on to the DC.

The DC does not have any keys. Nevertheless, the DC can apply T to $\mathbf{C}_{r'}$ to reduce the amount of data transmitted to Recipient 3 to the lower resolution r''. The option of nodes in the network being able to perform rate adaption is advantageous in networks with low bandwidth links, such as narrow-band powerline communication (PLC). Note that the reduction of resolution by the data concentrator DC by applying $T_{r''}$ is not relevant for security. Even if DC had passed on $\mathbf{C}_{r'}$ to Recipient 1 instead of $\mathbf{C}_{r''}$, Recipient 1 would still lack the key to decrypt $\mathbf{W}_{r'}$.

Both, Recipient 2 and Recipient 3 can obtain the load profiles $\mathbf{L}_{r'}$ and $\mathbf{L}_{r''}$, respectively, in the resolutions intended by the user.

To also illustrate the data perspective, Figure 4 shows an example of the wavelet decomposition of actual data. The original sampling interval in this example is 15 minutes, i.e., 96 values per day, as shown in Fig. 4(a). The coefficients of a level-1 decomposition with the Haar wavelet is shown in Fig. 4(b). The left half of these coefficients represent an lower resolution of the original sequence with a sampling interval of 30 minutes. Fig. 4(c) shows a level-5 wavelet decomposition. From this representation, 6 different resolution (including the original resolution) can be obtained, with the lowest resolution being composed of the three left-most values and representing a sampling interval of 8 hours.

IV. Evaluation

In this section, the proposed scheme is evaluated with respect to computational demands, security and privacy and real-world feasibility.

A. Complexity

As discussed above, implementing the wavelet transform as lifting steps is computationally inexpensive. Generally, the discrete wavelet transform has a complexity of $\mathcal{O}(n)$. Due to the simple operations used in the lifting implementation, the transformation part can be realized by inexpensive smart meter hardware.

The computational demands for encryption depends on the used encryption scheme. For standard encryption schemes, efficient implementations exist that can be integrated into smart meter hardware. Some overhead is introduced for key management, and potentially for the creation of session keys.

The proposed method has been implemented as a proof of concept in Java (Oracle Java v8 with ARM-extensions, version 1.8.0 with hard float). The method was evaluated in a low-cost ARM-based environment (Raspberry Pi 2, featuring a 900MHz quad-core ARM Cortex-A7 CPU and 1 GB RAM at a cost of \$35) running Raspbian Linux (in the version released on February 15, 2015, based on Debian Wheezy). The choice of this hardware platform is sensible, as it reflects the computing capabilities smart metering hardware will most likely provide. Rather than choosing the current solution of a single smart meter manufacturer, we use the open Raspberry Pi platform in combination with Linux to provide a test environment that is representative of future smart meters in a more general way.

To evaluate the method, we use the publicly available REDD data set [25]. This data set contains load data for a number of houses over the period of several weeks at a measuring interval of 3 seconds. For our test, we use 14 days of load profiles from house 1. We use the first 28,672 samples of the data set for each day, which corresponds to a measuring interval of 3.01 seconds. This sampling interval allows us a maximum wavelet decomposition depth of 12 without the need for border handling (because the number of samples equals $7 \cdot 2^{12}$), with the lowest resolution having a size of 7 samples (i.e., one aggregated value for every 3.4 hours). The sizes of the resolutions are given in Table I. Note that in real world setups, the number of measurements per day can be chosen by the smart meter, but will be affected by local legislation.

The following encryption scenarios were used: (1) wavelet transform only without any encryption, (2) Symmetric encryption: AES with 128-bit and 256-bit keys, and (3) Hybrid encryption: 128-bit and 256-bit AES resolution keys encrypted with 2048-bit RSA keys.

The following steps are executed in the scenarios: (i) Apply a Haar wavelet transform of depth 12 to the load profile (all scenarios), (ii) Generate 13 hierarchical AES resolution

TABLE I
RESOLUTION SIZES AND TEST RESULTS; X: ACTIVITY CAN BE INFERRED FROM DATA, C: COOKING, B: BATHROOM ACTIVITIES, H: HOUSEWORK, P: PRESENCE/ABSENCE

Resolution	Size	Approx. Interval	C	B	H	P
12	28672	3s	X	X	X	X
11	14336	6s	X	X	X	X
10	7168	12s	X	X	X	X
9	3584	24s	X	X	X	X
8	1792	48s		X	X	X
7	896	1.6m				X
6	448	3.2m				X
5	224	6.4m				X
4	112	12.9m				X
3	56	25.7m				X
2	28	51.4m				X
1	14	1.7h				
0	7	3.4h				

TABLE II
EXECUTION TIME FOR MULTI-RESOLUTION ENCRYPTION ON RASPBERRY PI 2, AVERAGE FOR 14 LOAD PROFILES WITH 500 EXECUTIONS EACH

Scenario	Avg. Exec. Time (ms)	Std. Deviation
(1) WAV only	9.1	1.2
(2) AES 128/256	46.3 / 52.6	2.3 / 0.6
(3) HYB 128-2048/256-2048	173.4 / 179.9	1.8 / 1.5
Key generation AES 128/256	1388 / 1390	17

keys (Scenarios 2 and 3 only), (iii) Encrypt $\mathbf{L}_0, \mathbf{H}_1, \ldots, \mathbf{H}_{12}$, each with a different key (Scenario 2 and 3 only), and (iv) Encrypt the 13 resolution keys with a 2048-bit RSA public key (Scenario 3 only).

The results are shown in Table II. The timing results are given in milliseconds comparing wavelet transform only (WAV) with AES and hybrid encryption (HYB) using an AES session key encrypted with an RSA public key. In each category, the 14 daily load profiles from the REDD data set were investigated, each of which was transformed and encrypted 500 times. The results present the average time needed for processing one load profile (i.e., one day).

It can be seen that compared to the computational demands of the encryption stage, the computational demands for the wavelet transform are almost negligible. On average, the transformation of a load profile takes 9 ms. This fact is a strong argument in favor of the proposed approach. Considering other tasks smart meters will need to be able to handle (such as key management), multi-resolution support comes at practically no additional cost.

Some overhead is incurred by the need to create the resolution keys. For creating 13 session keys with the Java standard pseudo-random number generator (SHA1PRNG), on a Raspberry Pi 2 and averaged over 500 executions our implementation took approx. 1390ms.

B. Privacy and Security Analysis

In the following, we first review the proposed protocol in more formal detail, to make clear which party generates which keys and which party performs the encryption. We then outline the basic assumptions regarding adversary behavior. The important aspect of information reduction through subsampling is discussed. Finally, based on this discussion, the used notion of privacy is defined in more detail.

1) Protocol Review: The proposed protocol is given in Figure 5. For the discussion, one exemplary use case of the proposed method was selected: A smart meter collects energy consumption data, performs multi-resolution analysis followed by encryption and sends the ciphertext to a DSO. We also include the possibility of an optional concentrator, which can adapt the resolution by applying $T(\cdot)$ to the ciphertext.

The DSO's public key $\mathsf{pk}_{\mathsf{DSO}}$ is made available to the smart meter via a public key infrastructure (assuming that the smart meter is initially provisioned with the public keys of the used certifying authorities). Note that, for the sake of simplicity, basic security measures (such as authentication and integrity checking) are not discussed here, but of course should be added in a real-world application.

A wavelet transformation is performed by the smart meter resulting in d resolutions (line 1). The individual resolution keys k_i are created by the smart meter with the hierarchical keying scheme discussed in Section III-E (line 2). The coefficients of each resolution are encrypted with the corresponding resolution key (line 3), using a symmetric cipher (in our tests we use AES).

The smart meter is configured to grant the DSO access to the consumption data up to resolution r (with $r \leq d$). The corresponding resolution key k_r is therefore encrypted with the DSO's public key (line 5), producing the "wrapped" key $\mathsf{wk}_{\mathsf{DSO}}$. Note that due to using a hierarchical keying scheme only this single key k_r needs to be transmitted to the DSO (the necessary keys for the lower resolution coefficients can be derived from k_r). The encrypted coefficients $\mathbf{C_d}$, together with the wrapped resolution key $\mathsf{wk}_{\mathsf{DSO}}$, are transmitted by the smart meter. Note that, additionally, the smart meter could add more wrapped resolution keys, encrypted for other recipients with the corresponding public keys.

In the topology, an optional data concentrator can be used: The data concentrator receives ciphertexts by various smart meters and passes them on to the DSO. The data concentrator can be configured to perform resolution adaption. In the protocol in Figure 5, the data concentrator can truncate the resolutions higher than r by simply discarding the corresponding encrypted coefficients (as discussed above, no decryption is necessary here).

The DSO decrypts the resolution key k_r from $\mathsf{wk}_{\mathsf{DSO}}$ with its private key $\mathsf{sk}_{\mathsf{DSO}}$ (line 6) and derives the resolution keys for the lower resolutions (line 7). It can then decrypt the encrypted coefficients for the resolutions up to r (lines 8-9). Finally, by applying the inverse wavelet transform, the load data $\mathbf{L}_r$ of resolution r is obtained (line 10).

2) Adversary Model Assumptions: The smart meter is assumed to be trusted: It will reliably realize wavelet transform, key generation and encryption. It is assumed that no active or passive adversary has access to the internal processing of the smart meter.

The DSO and potential other legitimate recipients of the load data (in a specific resolution) are assigned an honest-but-curios (semi-honest) role: They will reliably perform wavelet

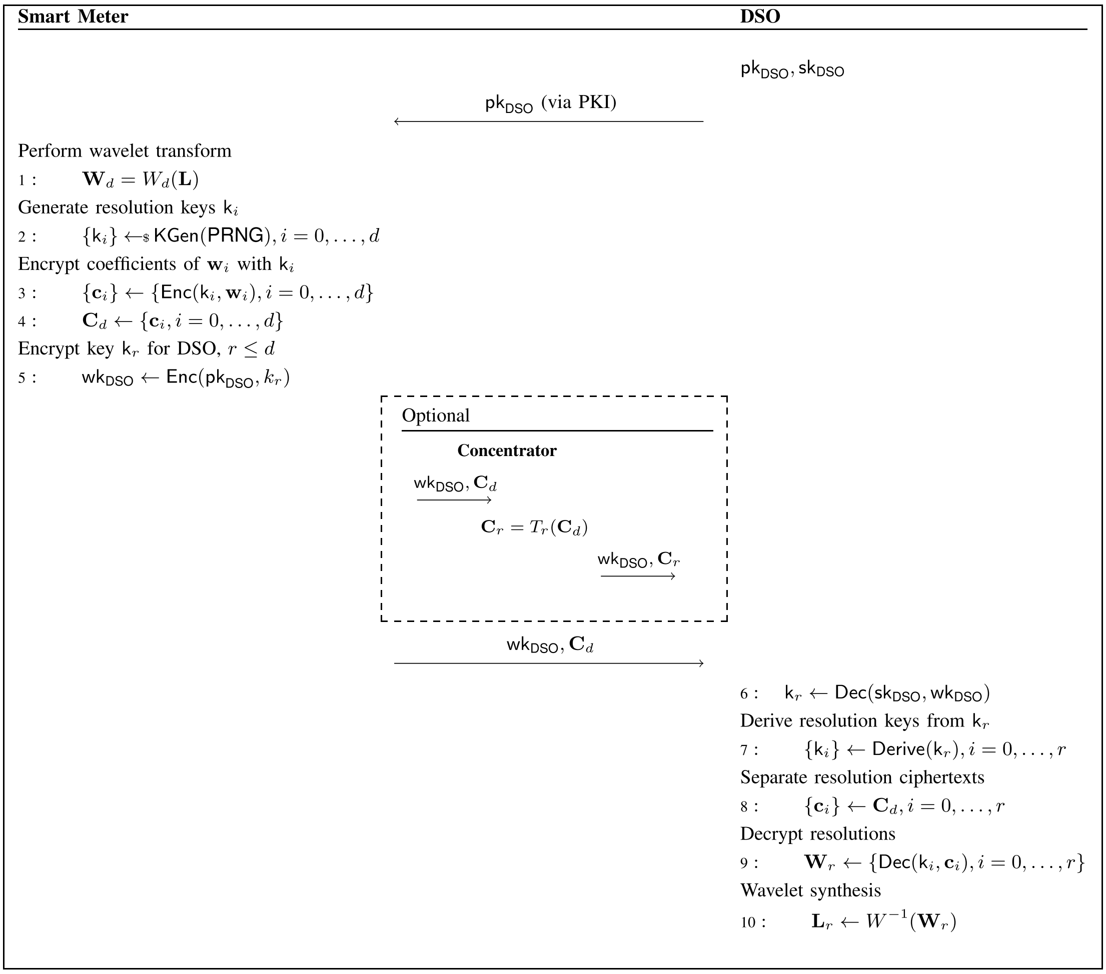

Smart Meter
DSO
$\mathsf{pk}_{\mathsf{DSO}}, \mathsf{sk}_{\mathsf{DSO}}$
$\mathsf{pk}_{\mathsf{DSO}}$ (via PKI)
Perform wavelet transform
1 : $\mathbf{W}_d = W_d(\mathbf{L})$
Generate resolution keys k_i
2 : $\{\mathsf{k}_i\} \leftarrow_{\$} \mathsf{KGen}(\mathsf{PRNG}), i = 0, \ldots, d$
Encrypt coefficients of $\mathbf{w}_i$ with k_i
3 : $\{\mathbf{c}_i\} \leftarrow \{\mathsf{Enc}(\mathsf{k}_i, \mathbf{w}_i), i = 0, \ldots, d\}$
4 : $\mathbf{C}_d \leftarrow \{\mathbf{c}_i, i = 0, \ldots, d\}$
Encrypt key k_r for DSO, $r \leq d$
5 : $\mathsf{wk}_{\mathsf{DSO}} \leftarrow \mathsf{Enc}(\mathsf{pk}_{\mathsf{DSO}}, k_r)$
Optional
Concentrator
$\mathsf{wk}_{\mathsf{DSO}}, \mathbf{C}_d$
$\mathbf{C}_r = T_r(\mathbf{C}_d)$
$\mathsf{wk}_{\mathsf{DSO}}, \mathbf{C}_r$
$\mathsf{wk}_{\mathsf{DSO}}, \mathbf{C}_d$
6 : $\mathsf{k}_r \leftarrow \mathsf{Dec}(\mathsf{sk}_{\mathsf{DSO}}, \mathsf{wk}_{\mathsf{DSO}})$
Derive resolution keys from k_r
7 : $\{\mathsf{k}_i\} \leftarrow \mathsf{Derive}(\mathsf{k}_r), i = 0, \ldots, r$
Separate resolution ciphertexts
8 : $\{\mathbf{c}_i\} \leftarrow \mathbf{C}_d, i = 0, \ldots, r$
Decrypt resolutions
9 : $\mathbf{W}_r \leftarrow \{\mathsf{Dec}(\mathsf{k}_i, \mathbf{c}_i), i = 0, \ldots, r\}$
Wavelet synthesis
10 : $\mathbf{L}_r \leftarrow W^{-1}(\mathbf{W}_r)$

Fig. 5. Protocol for Multi-resolution Privacy for transferring a load profile $\mathbf{L}$ from the smart meter to the DSO at reduced resolution r as $\mathbf{L}_r$, with optional additional rate adaption applied by a data concentrator. $\mathsf{pk}_{\mathsf{DSO}}$ denotes the DSO's public key, $\mathsf{sk}_{\mathsf{DSO}}$ denotes the DSO's private key. k_i denotes the resolution keys used for symmetric encryption of the wavelet coefficients in subband $\mathbf{w}_i$. $\mathsf{wk}_{\mathsf{DSO}}$ denotes the "wrapped key," i.e., the resolution key of the highest resolution intended for the DSO encrypted by $\mathsf{pk}_{\mathsf{DSO}}$.

transform and decryption, and will generally adhere to the protocol. However, if a recipient sees an opportunity to get data at a higher resolution than intended by the smart meter for this recipient, it will access this information.

The data concentrator needs not be a trusted entity. It only needs to be trusted to receive data and pass on this data, optionally with rate adaption. The data concentrator is not trusted with key material.

The communication links are assumed to be insecure and at least subject to eavesdropping, possibly also subject to active attacks, such as man-in-the-middle attacks.

3) Analysis of Potential Attacks: A **semi-honest recipient** (e.g., a DSO) could try to gain access to higher resolutions than the sender intended. There are different options to try: (i) derive a resolution key for a higher resolution than the wrapped key, (ii) derive higher resolutions from the lower resolutions, (iii) break the symmetric encryption of the coefficients for the higher resolutions, (iv) break the asymmetric encryption of other wrapped keys transmitted with the ciphertext (intended for other recipients). Option (i) is infeasible if a proper one-way hash function (e.g., SHA-3) is used to create the hierarchical keys. Option (ii) is infeasible given certain assumptions, which will be discussed in detail in Section *Privacy Analysis* below. Option (iii) and (iv) are infeasible if state-of-the-art cryptographic methods are used with suitable keys. In our tests we use AES and RSA. It can be concluded that even for a semi-honest recipient, the protocol achieves the intended effect of only granting access to a certain resolution.

A **collusion of multiple semi-honest recipients** will yield the plaintext data of the highest resolution granted to this group of recipients (by sharing the plaintext). However, the collusion will not yield the plaintext consumption data of higher resolutions.

A **semi-honest data concentrator** trying to gain access to the data has less options than the semi-hones recipient, as it does not have any access to the keys and therefore, other than the recipient, cannot decrypt even parts of the data. Options (i) and (ii) are therefore not applicable. Options (iii) and (iv) are infeasible, if proper ciphers are used. The only information that the data concentrator has, is the number of recipients (through the number of wrapped keys), the addresses of recipients for which the data concentrator acts as a direct relay, and the number of resolutions contained in the message.

A **malicious data concentrator** can refuse to pass on the message and can thus realize a denial of service attack. It can also send bogus messages: the data concentrator can make up consumption data, acquire the public key of the DSO (or any other recipient) and then run the protocol steps for encryption (lines 1-5 in Figure 5). This can be counteracted by integrity checks. A malicious data concentrator could also perform a man-in-the-middle attack. This can be counteracted by putting authentication into place. One attack by a malicious concentrator remains that cannot be counteracted: a malicious concentrator can always decrease the resolution of the ciphertext it passes on. Recipients behind the malicious concentrator would then receive lower resolutions than intended. This circumstance would be noted by the recipient (because the available resolutions would not match the resolution implied by the included wrapped key).

Eavesdropping attacks on the communication links are not feasible as only attack options (iii) and (iv) are available, which are both infeasible. An active **malicious attack on the communication links** (such as a man in the middle attack) can realize the same attacks as a malicious data concentrator.

4) Privacy Analysis: The effectiveness of reducing personal information through the reduction in resolution is one of the central questions in evaluating the usefulness of the proposed scheme. Furthermore, as discussed in the previous section, legitimate, but semi-honest, recipients could try to obtain higher resolution information from the lower resolutions.

a) Information reduction through subsampling: For a user-centric privacy approach, it is essential to answer the question, how much privacy is introduced by repeatedly halving the sampling rate – i.e., how does a decrease in resolution actually impact the degree of personal information contained in the underlying data? And, on a related note, how much useful information is contained in a certain resolution for realizing a use case desired by the consumer (such as energy usage optimization).

It is evident that the low-passing filtering achieved by the Wavelet transform will reduce information leakage for privacy-sensitive series of load measurements. As the approach proposed here supports a high number of different resolution, it is safe to state that it will effectively increase privacy. The question remains, what target resolution the end-consumer should aim for in a given use-case.

As an example, the study of Eibl and Engel [4] explicitly aimed at tackling the relation between NILM accuracy and data resolution (Table I). Empirical material to assess the utility of a reduction of resolution for increasing privacy is provided. The four columns on the right-hand side of Table I shows the effect of decreasing resolution with the wavelet-based approach on the detection accuracy for the activities “cooking”, “bathroom activities”, ”housework”, and “presence/absence”. It can be seen that decreasing the resolution is a measure to prevent detection of these activities (given the detection accuracy of current NILM methods). It should be noted that the transition from detectable to undetectable is not hard and slightly smoother than suggested by Table I.

In [4], an explicit algorithm is used for a privacy attack. Although it can be argued that typical NILM algorithms based on differences of power values should suffer from a decreased resolution this does not necessarily have to be true in general if other kinds of attacks are considered. For example the argument does not hold for approaches like the ones in [35] and [36] where the occupancy information is retrieved based on absolute values that can even be averaged over several hours.

As a candidate for a general privacy measure, Sankar, Rajagopalan *et al.* propose an information-theoretic approach that uses mutual information (MI) to evaluate privacy in [37] and [38]. From the theoretical side, there is also a relation between differential privacy and MI [39]. As a big drawback, MI has not been applied to real world data in these publications. One of the reasons for that lack may be the fact that MI is hard to estimate. There are approaches in the (Non-Intrusive Load Leveling) NILL community that use MI estimated by binning as a method to assess the similarity of the original and the changed signal [40]. The current method of choice for the estimation of MI is based on k nearest neighbors [41], [42] which has shown to be superior to the binning method which heavily depends on the bin size. In an attempt to evaluate privacy using MI, we programmed the algorithm in Matlab and successfully tested the correctness of our implementation for the correlated Gaussian example. However, the application to the real world data showed a big dependence on the number of nearest neighbors k making the results doubtful. In [41] it is mentioned that the estimation algorithm fails, if distributions are strongly peaked which is also the case here. Using ranks instead of the absolute values did not improve the estimation, so why this algorithm did not work stays an area for future research.

Instead, the regression approach of [43] and [44] is adopted. Instead of using Pearson’s coefficient of correlation the more robust Spearman correlation coefficient r_{Spearman} is used here. The correlation coefficient was computed between the original sample and a wavelet approximation for each scale. In [43] and [44] this is done for finite differences of the load profile corresponding to the privacy attacks on turn-on and turn-off events. In order to account for other attacks that are based on absolute values [35], [36] this is also done for absolute values. In order to get the sign right, i.e., a measure for privacy and not for correlation $1-r_{\text{Spearman}}$ is used as a privacy measure.

In Figure 6 the result is illustrated for the mains signal of house 1 of the REDD dataset [25]. As expected the privacy

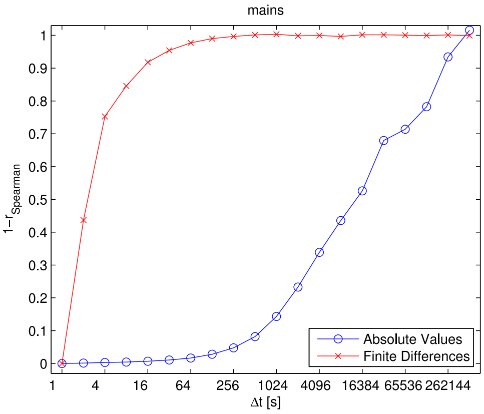


Fig. 6. Dependence of privacy, measured by 1 minus the Spearman correlation coefficient between the signal with highest resolution and lower resolutions.

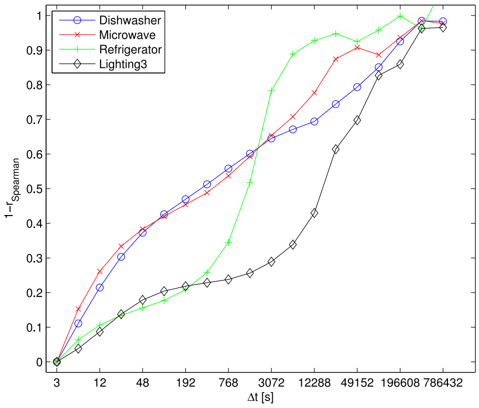


Fig. 7. Dependence of privacy, measured by 1 minus the Spearman correlation, on resolution based on absolute values for different appliances.

measure clearly increases when the resolution (here, given as a time scale) is decreased. This holds especially for the difference of subsequent values suggesting that methods based on absolute values like in [35] and [36] might be more robust with respect to a decrease of resolution.

To get further insights, the privacy measure is also computed for signals of individual appliances (Figure 7). Although absolute values are used here in contrast to [4], the results are qualitatively consistent with the results of [4]. At the same resolution, privacy of lights is estimated lower than privacy of other appliances. This is plausible due to the higher on-times of the lights [4]. Due to the more regular time-behavior of its load curve the refrigerator (typically considered as privacy-irrelevant) has a rather steep privacy increase in the middle.

b) Inferring higher resolutions from lower resolutions: Assume Eve has been granted access to resolution r of a bit-stream containing a maximum resolution of d. Eve could try to use the coefficients from the lower resolutions, for which she has access, to extrapolate the higher resolution. The feasibility of this attack depends on the characteristics of the original load profile. The question is directly related to the question, how much privacy is introduced by halving the sampling rate. If the original has a high number of high frequency components (which are also crucial in NILM accuracy), significant data loss occurs going from d to r and Eve will be unable to make any assumptions on resolution d from r.

V. Combination With Other Privacy-Enhancing Technologies

The multi-resolution approach to privacy presented above is compatible with many other privacy-enhancing technologies (PETs). By combining these PETs with multi-resolution analysis, additional degrees of freedom and a broader range of choices for the end-user can be realized. In the following, we review the compatibility of the multi-resolution approach proposed here with privacy-preserving protocols found in the literature.

A. Secure Aggregation With Homomorphic Encryption

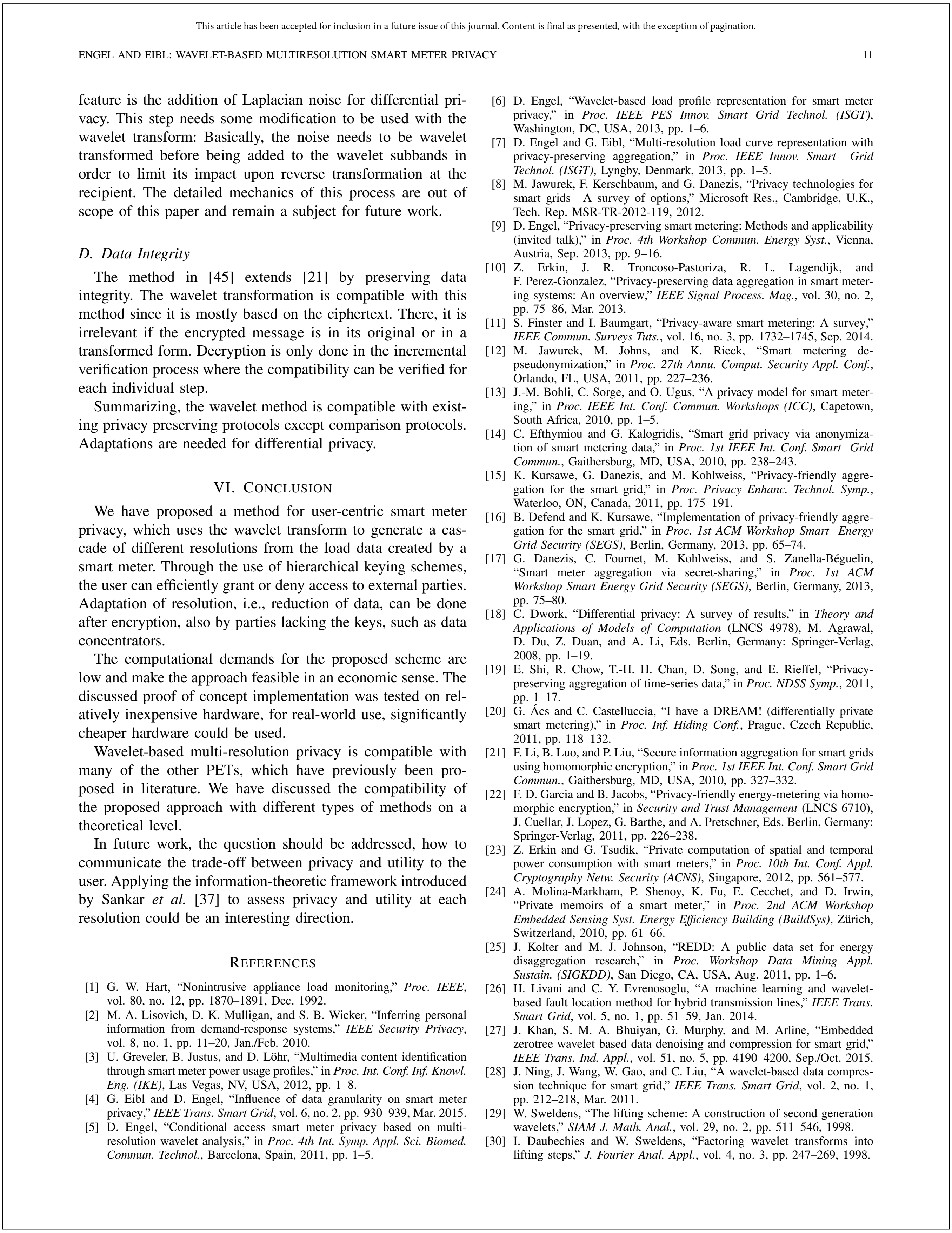
In [7], Engel and Eibl have shown that multi-resolution analysis can be used within privacy preserving protocols which directly rely on the homomorphic encryption property for secure aggregation, such as proposed by Li *et al.* [21] or Erkin and Tsudik [23]. In particular, it has been shown that when homomorphic encryption is applied to a signal represented in the wavelet domain, homomorphic additivity is not only preserved, but can be separately exploited for each resolution.

B. Additive Secret Sharing

The method proposed by Garcia and Jacobs [22] combines Paillier's homomorphic encryption with additive secret sharing. Generally, additive masking terms need no adjustment since they cancel out in the decryption step before the inverse transformation takes place. Thus, the method is compatible with the wavelet transformation. In [15], Kursawe *et al.* describe four different protocols which rely on masking. These protocols can be categorized into so-called aggregation and comparison protocols. All of the protocols are designed as simple as possible to be feasible for use in the field. All of the aggregation protocols are compatible with wavelets, and masking can be applied to each resolution separately. However, in the comparison protocols, the transformed sum of the values is in the exponent of the generating element of the Diffie-Hellman group. As the reverse transformation cannot be calculated for terms in the exponent, wavelets are not compatible with these comparison protocols.

C. Differential Privacy

In [20], Ács and Castelluccia use the modulo operation for homomorphic encryption instead of Paillier's homomorphic encryption scheme. Privacy and also confidentiality with the aggregator is achieved by masking. As stated above, the additive masking terms need no adjustment. The second main



feature is the addition of Laplacian noise for differential privacy. This step needs some modification to be used with the wavelet transform: Basically, the noise needs to be wavelet transformed before being added to the wavelet subbands in order to limit its impact upon reverse transformation at the recipient. The detailed mechanics of this process are out of scope of this paper and remain a subject for future work.

D. Data Integrity

The method in [45] extends [21] by preserving data integrity. The wavelet transformation is compatible with this method since it is mostly based on the ciphertext. There, it is irrelevant if the encrypted message is in its original or in a transformed form. Decryption is only done in the incremental verification process where the compatibility can be verified for each individual step.

Summarizing, the wavelet method is compatible with existing privacy preserving protocols except comparison protocols. Adaptations are needed for differential privacy.

VI. Conclusion

We have proposed a method for user-centric smart meter privacy, which uses the wavelet transform to generate a cascade of different resolutions from the load data created by a smart meter. Through the use of hierarchical keying schemes, the user can efficiently grant or deny access to external parties. Adaptation of resolution, i.e., reduction of data, can be done after encryption, also by parties lacking the keys, such as data concentrators.

The computational demands for the proposed scheme are low and make the approach feasible in an economic sense. The discussed proof of concept implementation was tested on relatively inexpensive hardware, for real-world use, significantly cheaper hardware could be used.

Wavelet-based multi-resolution privacy is compatible with many of the other PETs, which have previously been proposed in literature. We have discussed the compatibility of the proposed approach with different types of methods on a theoretical level.

In future work, the question should be addressed, how to communicate the trade-off between privacy and utility to the user. Applying the information-theoretic framework introduced by Sankar *et al.* [37] to assess privacy and utility at each resolution could be an interesting direction.

References

[1] G. W. Hart, "Nonintrusive appliance load monitoring," *Proc. IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992.

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

[3] U. Greveler, B. Justus, and D. Löhr, "Multimedia content identification through smart meter power usage profiles," in *Proc. Int. Conf. Inf. Knowl. Eng. (IKE)*, Las Vegas, NV, USA, 2012, pp. 1–8.

[4] G. Eibl and D. Engel, "Influence of data granularity on smart meter privacy," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 930–939, Mar. 2015.

[5] D. Engel, "Conditional access smart meter privacy based on multi-resolution wavelet analysis," in *Proc. 4th Int. Symp. Appl. Sci. Biomed. Commun. Technol.*, Barcelona, Spain, 2011, pp. 1–5.

[6] D. Engel, "Wavelet-based load profile representation for smart meter privacy," in *Proc. IEEE PES Innov. Smart Grid Technol. (ISGT)*, Washington, DC, USA, 2013, pp. 1–6.

[7] D. Engel and G. Eibl, "Multi-resolution load curve representation with privacy-preserving aggregation," in *Proc. IEEE Innov. Smart Grid Technol. (ISGT)*, Lyngby, Denmark, 2013, pp. 1–5.

[8] M. Jawurek, F. Kerschbaum, and G. Danezis, "Privacy technologies for smart grids—A survey of options," Microsoft Res., Cambridge, U.K., Tech. Rep. MSR-TR-2012-119, 2012.

[9] D. Engel, "Privacy-preserving smart metering: Methods and applicability (invited talk)," in *Proc. 4th Workshop Commun. Energy Syst.*, Vienna, Austria, Sep. 2013, pp. 9–16.

[10] Z. Erkin, J. R. Troncoso-Pastoriza, R. L. Lagendijk, and F. Perez-Gonzalez, "Privacy-preserving data aggregation in smart metering systems: An overview," *IEEE Signal Process. Mag.*, vol. 30, no. 2, pp. 75–86, Mar. 2013.

[11] S. Finster and I. Baumgart, "Privacy-aware smart metering: A survey," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1732–1745, Sep. 2014.

[12] M. Jawurek, M. Johns, and K. Rieck, "Smart metering de-pseudonymization," in *Proc. 27th Annu. Comput. Security Appl. Conf.*, Orlando, FL, USA, 2011, pp. 227–236.

[13] J.-M. Bohli, C. Sorge, and O. Ugus, "A privacy model for smart metering," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC)*, Capetown, South Africa, 2010, pp. 1–5.

[14] C. Efthymiou and G. Kalogridis, "Smart grid privacy via anonymization of smart metering data," in *Proc. 1st IEEE Int. Conf. Smart Grid Commun.*, Gaithersburg, MD, USA, 2010, pp. 238–243.

[15] K. Kursawe, G. Danezis, and M. Kohlweiss, "Privacy-friendly aggregation for the smart grid," in *Proc. Privacy Enhanc. Technol. Symp.*, Waterloo, ON, Canada, 2011, pp. 175–191.

[16] B. Defend and K. Kursawe, "Implementation of privacy-friendly aggregation for the smart grid," in *Proc. 1st ACM Workshop Smart Energy Grid Security (SEGS)*, Berlin, Germany, 2013, pp. 65–74.

[17] G. Danezis, C. Fournet, M. Kohlweiss, and S. Zanella-Béguelin, "Smart meter aggregation via secret-sharing," in *Proc. 1st ACM Workshop Smart Energy Grid Security (SEGS)*, Berlin, Germany, 2013, pp. 75–80.

[18] C. Dwork, "Differential privacy: A survey of results," in *Theory and Applications of Models of Computation* (LNCS 4978), M. Agrawal, D. Du, Z. Duan, and A. Li, Eds. Berlin, Germany: Springer-Verlag, 2008, pp. 1–19.

[19] E. Shi, R. Chow, T.-H. H. Chan, D. Song, and E. Rieffel, "Privacy-preserving aggregation of time-series data," in *Proc. NDSS Symp.*, 2011, pp. 1–17.

[20] G. Ács and C. Castelluccia, "I have a DREAM! (differentially private smart metering)," in *Proc. Inf. Hiding Conf.*, Prague, Czech Republic, 2011, pp. 118–132.

[21] F. Li, B. Luo, and P. Liu, "Secure information aggregation for smart grids using homomorphic encryption," in *Proc. 1st IEEE Int. Conf. Smart Grid Commun.*, Gaithersburg, MD, USA, 2010, pp. 327–332.

[22] F. D. Garcia and B. Jacobs, "Privacy-friendly energy-metering via homomorphic encryption," in *Security and Trust Management* (LNCS 6710), J. Cuellar, J. Lopez, G. Barthe, and A. Pretschner, Eds. Berlin, Germany: Springer-Verlag, 2011, pp. 226–238.

[23] Z. Erkin and G. Tsudik, "Private computation of spatial and temporal power consumption with smart meters," in *Proc. 10th Int. Conf. Appl. Cryptography Netw. Security (ACNS)*, Singapore, 2012, pp. 561–577.

[24] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin, "Private memoirs of a smart meter," in *Proc. 2nd ACM Workshop Embedded Sensing Syst. Energy Efficiency Building (BuildSys)*, Zürich, Switzerland, 2010, pp. 61–66.

[25] J. Kolter and M. J. Johnson, "REDD: A public data set for energy disaggregation research," in *Proc. Workshop Data Mining Appl. Sustain. (SIGKDD)*, San Diego, CA, USA, Aug. 2011, pp. 1–6.

[26] H. Livani and C. Y. Evrenosoglu, "A machine learning and wavelet-based fault location method for hybrid transmission lines," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 51–59, Jan. 2014.

[27] J. Khan, S. M. A. Bhuiyan, G. Murphy, and M. Arline, "Embedded zerotree wavelet based data denoising and compression for smart grid," *IEEE Trans. Ind. Appl.*, vol. 51, no. 5, pp. 4190–4200, Sep./Oct. 2015.

[28] J. Ning, J. Wang, W. Gao, and C. Liu, "A wavelet-based data compression technique for smart grid," *IEEE Trans. Smart Grid*, vol. 2, no. 1, pp. 212–218, Mar. 2011.

[29] W. Sweldens, "The lifting scheme: A construction of second generation wavelets," *SIAM J. Math. Anal.*, vol. 29, no. 2, pp. 511–546, 1998.

[30] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," *J. Fourier Anal. Appl.*, vol. 4, no. 3, pp. 247–269, 1998.

[31] F. Knirsch, D. Engel, M. Frincu, and V. Prasanna, "Model based assessment for balancing privacy requirements and operational capabilities in the smart grid," in *Proc. 6th Conf. Innov. Smart Grid Technol. (ISGT)*, Washington, DC, USA, 2015, pp. 1–5.

[32] C. D. Peer, D. Engel, and S. B. Wicker, "Hierarchical key management for multi-resolution load data representation," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Venice, Italy, Nov. 2014, pp. 926–932.

[33] S. Imaizumi, N. Aoki, H. Kobayashi, and H. Kiya, "Hierarchical key assignment scheme for multimedia access control with modified hash chain," in *Proc. 8th Int. Conf. Intell. Inf. Hiding Multimedia Signal Process. (IIH-MSP)*, Piraeus, Greece, 2012, pp. 293–296.

[34] L. Lamport, "Password authentication with insecure communication," *Commun. ACM*, vol. 24, no. 11, pp. 770–772, Nov. 1981.

[35] D. Chen, S. Barker, A. Subbaswamy, D. Irwin, and P. Shenoy, "Non-intrusive occupancy monitoring using smart meters," in *Proc. 5th ACM Workshop Embedded Syst. Energy Efficiency Build. (BuildSys)*, Rome, Italy, 2013, pp. 1–8.

[36] D. Chen, D. Irwin, P. Shenoy, and J. Albrecht, "Combined heat and privacy: Preventing occupancy detection from smart meters," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, Budapest, Hungary, 2014, pp. 208–215.

[37] L. Sankar, S. R. Rajagopalan, S. Mohajer, and H. V. Poor, "Smart meter privacy: A theoretical framework," *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 837–846, Jun. 2013.

[38] S. R. Rajagopalan, L. Sankar, S. Mohajer, and H. V. Poor, "Smart meter privacy: A utility-privacy framework," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Brussels, Belgium, 2011, pp. 190–195.

[39] W. Wang, L. Ying, and J. Zhang, "On the relation between identifiability, differential privacy and mutual-information privacy," in *Proc. 52nd Annu. Allerton Conf. Commun. Control Comput.*, Monticello, IL, USA, 2014, pp. 1086–1092.

[40] W. Yang, N. Li, and Y. Qi, "Minimizing private data disclosures in the smart grid," in *Proc. ACM Conf. Comput. Commun. Security (CCS)*, Raleigh, NC, USA, 2012, pp. 415–427.

[41] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Phys. Rev. E*, vol. 69, no. 6, 2004, Art. ID 066138.

[42] A. Papana and D. Kugiumtzis, "Evaluation of Mutual Information Estimators for Time Series", *Int. J. Bifurcation Chaos*, vol. 19, no. 12, pp. 4197–4215, 2009.

[43] G. Kalogridis, R. Cepeda, S. Z. Denic, T. Lewis, and C. Efthymiou, "ElecPrivacy: Evaluating the privacy protection of electricity management algorithms," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 750–758, Dec. 2011.

[44] G. Kalogridis and S. Z. Denic, "Data mining and privacy of personal behaviour types in smart grid," in *Proc. IEEE 11th Int. Conf. Data Min. Workshops (ICDMW)*, Vancouver, BC, Canada, 2011, pp. 636–642.

[45] F. Li and B. Luo, "Preserving data integrity for smart grid data aggregation," in *Proc. IEEE 3rd Int. Conf. Smart Grid Commun. (SmartGridComm)*, Tainan, Taiwan, 2012, pp. 366–371.

Dominik Engel (S'06–M'08) received the Ph.D. degree in computer science from the University of Salzburg, Salzburg, Austria, in 2008.

He was a Researcher with the University of Bremen, Bremen, Germany, and the University of Salzburg, and a Product Manager with Sony DADC, Anif, Austria, where he was responsible for video content security. He is a Professor with the Salzburg University of Applied Sciences, Austria, where he is the Head of the Josef Ressel Center for User-Centric Smart Grid Privacy, Security, and Control. His current research interests include smart grid security and privacy, and methods for enhancing trustworthiness of technical systems in general.

Günther Eibl (M'13) received the Ph.D. degree in mathematics and the M.Sc. degree in physics from the University of Innsbruck, Innsbruck, Austria, in 1997 and 2002, respectively.

He is a Research Associate with the Josef Ressel Center for User-Centric Smart Grid Privacy, Security, and Control, Salzburg University of Applied Sciences, Austria. He previously held research positions with the Institutes of Biostatistics, Innsbruck, and the Institute of Theoretical Physics, Innsbruck. In academic and nonacademic research, he worked in such fields as data mining and machine learning, particle and fluid simulations, computer vision, robot kinematics, control, and cryptography. His research interests include extraction of information from data with a focus on statistical modeling, data mining, and privacy preserving technologies.

3.2 EIBL15A

- G. Eibl and D. Engel. Influence of data granularity on smart meter privacy. *IEEE Transactions on Smart Grid*, 6(2):930–939, March 2015.

Influence of Data Granularity on Smart Meter Privacy

Günther Eibl, *Member, IEEE*, and Dominik Engel, *Member, IEEE*

***Abstract*—Through smart metering in the smart grid end-user domain, load profiles are measured per household. Personal data can be inferred from these load profiles by using nonintrusive appliance load monitoring methods, which has led to privacy concerns. Privacy is expected to increase with longer intervals between measurements of load curves. This paper studies the impact of data granularity on edge detection methods, which are the common first step in nonintrusive load monitoring algorithms. It is shown that when the time interval exceeds half the on-time of an appliance, the appliance use detection rate declines. Through a one-versus-rest classification modeling, the ability to detect an appliance's use is evaluated through F-scores. Representing these F-scores visually through a heatmap yields an easily understandable way of presenting potential privacy implications in smart metering to the end-user or other decision makers.**

***Index Terms*—Data granularity, privacy, smart metering.**

I. Introduction

THERE IS a lot of public concern and discussions on the privacy impact of smart metering. However, most discussions take place without knowing the extent of personal information that can be read out of smart meter load profiles. Even more so, there is nearly a complete lack of knowledge on how the amount of personal information relates to the measured time interval, i.e., the time granularity. For example, in many countries in Europe it is planned that smart meters will deliver load data in 15 min time intervals [1]. This has sparked a (sometimes emotional) debate on privacy (see [2]–[4]). However, to our knowledge, no one has tried to assess the amount of personal information that can be extracted on 15 min time interval load profiles, or how, in general, data granularity relates to the amount and nature of extractable personal data.

Although the decrease of the time granularity can be viewed as the most straightforward and simplest privacy enhancing technology—and this method has been suggested by a number of contributions in the past (see [5]–[7])—its impact on privacy has not yet been studied systematically, apart from an initial study we published in [8]. The goal of this paper is making the first step toward a systematic evaluation by studying the impact of time granularity on determination of appliance use. The main reasoning behind this approach is that activities of persons in the house trigger appliances, which in turn sum up to the total load. The activities themselves are influenced by various aspects of personal information such as presence, sleep-wake-cycles, and personal habits

$$\text{Personal Info} \Rightarrow \text{Activities} \Rightarrow \text{Appl. Use} \Rightarrow \text{Load Curve.} \quad (1)$$

This causal chain is the reason why the knowledge of activities leads to knowledge of personal information. As a first step toward a privacy assessment this paper focuses on the detection of appliance use with a short discussion on how activities could be assessed.

Information on appliances is usually extracted from the load data by means of so-called "nonintrusive appliance load monitoring analyzes" (NIALM). There is a lot of literature on NIALM algorithms ([9]–[16], to name a few). The primary goal of these algorithms is the disaggregation of the total load into the individual appliances loads for sake of providing an energy feedback to the end-user. Seen in a different perspective, such NIALM analyzes could also be used as the first step of methods attacking personal privacy by using NIALM as the basis for the extraction of personal information. Instead of using a whole NIALM algorithm as a method for gathering private information, in this paper, a simpler method is used which only uses the first part of typical low-frequency NIALM algorithms, namely edge detection ([2], [9], [11], [12], [14], [16]).

Compared to the large amount of literature aiming at providing energy feedback to the end-user, privacy implications are only rarely treated. In [2], load data were recorded with parallel video data which were processed into activity logs. A NIALM analysis was done yielding the input for subsequent behavior-extraction routines. Extracted behaviors include, e.g., presence, sleep cycles, or meal times. The amount of information disclosure is measured by an overall number called "degree of disclosure." In [4], the load profile is divided into so-called power segments using a density-based clustering technique. These power segments are described by features such as start time, average power, and duration. It is illustrated how such power events could be used for answering several privacy-sensitive questions. In [17], it is shown that under ideal conditions and using small measurement time intervals, even the consumed TV-program can be inferred from load curves.

Manuscript received March 4, 2014; revised July 16, 2014 and September 11, 2014; accepted November 20, 2014. Date of publication December 18, 2014; date of current version February 16, 2015. This work was supported in part by the Austrian Federal Ministry of Science, Research, and Economy, and in part by the Austrian National Foundation for Research, Technology, and Development. Paper No. TSG-00197-2014.

The authors are with the Salzburg University of Applied Sciences, Josef Ressel Center for User-Centric Smart Grid Privacy, Security, and Control, Puch/Salzburg A-5412, Austria (e-mail: guenther.eibl@en-trust.at).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSG.2014.2376613

TABLE I
TIME GRANULARITIES OF LOW-FREQUENCY NIALM-STUDIES

Time	1s	3s	15s	20s	1min
Paper	[4], [18], [11], [12]	[14], [15]	[2]	[13]	[16]

This paper is organized as follows. In Section II, NIALM and edge detection methods are reviewed. After the description of the experimental setup in Section III, event detection is applied on the load profiles in Section IV as a method for the extraction of personal information. After this attacking method has been developed, the decrease of time resolution is applied as a countermeasure in Section VI, where the influence of time granularity on the event detection performance is studied. By applying a classification setting, results are described by precision and recall rates which are used as inputs for a systematic privacy analysis in Section VII.

II. Background

A. NIALM Analyzes

NIALM analyzes can be broadly divided into two kinds of methods: 1) high frequency; and 2) low frequency methods. High frequency methods look at the waveform of appliances or study transients or higher order harmonics [10]. While the high-frequency methods need a sampling in the range of kHz, the low frequency methods typically analyze load profiles which are sampled using time intervals in the order of seconds to minutes (see Table I).

This paper focuses on low-frequency methods. Particularly, the methods developed here follow the class of supervised NIALM methods [9]. Supervised methods usually consist of several blocks: edge detection, cluster analysis, and finding pairs of on-and-off clusters for the determination of the duration of an appliance. Edges are sharp increases or decreases of the load signal due to turning on or off an appliance. More generally, edges arise due to the change from one state to another state of an appliance when modeled as a finite state machines (FSM). NIALM algorithms commonly use edges instead of the absolute values for two reasons.

1) First, if absolute values were used in the presence of unknown appliances, these appliances would have to be described as a combination of other known appliances.
2) Second, there are adverse cases, where a small change in the measured power would result in a big change in the configuration of used appliances which is not plausible [9].

Although the use of edges is most common other features can be used as well such as the shape features of [4]. A typical assumption in the disaggregation processes is the switch continuity principle which states that in a small time interval only a small number of appliances is expected to change the state [9]. Often, this assumption is tightened by requiring that in a time interval at most one appliance changes its state (one-at-a-time condition).

The usual performance measures of NIALM methods are the error in the total energy assigned to a given appliance or the error in the estimated on-time. Event-based methods state the performance in terms of precision p and recall r [2], [14] or the F-score [15]. Precision p is the proportion of events classified as stemming from appliance A which is really stemming from appliance A. Recall is the proportion of all events stemming from appliance A that is also classified as stemming from appliance A. Performance is either given by the pair (p, r), or if a single performance number is needed by the F-score F

$$F = 2\frac{p \cdot r}{p + r}. \tag{2}$$

B. Event Detection Methods

In this section, event detection methods are reviewed. The main assumption is the validity of modeling appliances as FSM having different power values for different states. An edge or event $e = (t_e, \Delta P_e)$ is a transition between two such states. It is represented by the onset time t_e and a transition value ΔP_e, which is the difference in power levels of the two states. Events with increasing signal ($\Delta P > 0$) are called on-events because they typically arise from turning on an on–off-appliance. Analogously, events with $\Delta P < 0$ are called off-events.

The most straightforward edge detection method, called difference method, detects an edge, if the difference $\Delta P_i = P_{i+1} - P_i$ between consecutive power values exceeds a threshold. Each detected edge is considered to be an event $e = (t_i, \Delta P_i)$. If the transition between two levels needs several time intervals, the method divides the transition between two levels in several edges having smaller values than the transition.

Due to this drawback, the edge merging method merges subsequently occurring edges into a single event [12]. The value of the event is the sum of the individual edge values, which can be both positive and negative. The time where the event occurs is defined as the onset time, i.e., the time of the first edge contributing to the event.

While the previous two methods focus on the transition between two levels of a signal, the next method focuses on the power levels of the two transition states. The method was proposed in [9], where it is called transient passing method for edge detection. A transition is inversely defined as being not a steady subsequence. In the first step the method finds the steady subsequences of the signal. This is done using a sliding window approach where a subsequence consisting of n points is considered as steady, if the range of its values does not exceed a given threshold. As a result, the whole signal is divided into consecutive steady parts st and unsteady transitions tr. For the description of the event e arising from transition tr_i the three subsequences (st_{i-1}, tr_i, st_i) are considered. The onset-time t_e for the description of the event is the last time point of the first steady part st_{i-1}. The transition value ΔP_e is the difference between the median of the values of the second steady part st_{i+1} and the median of the values of the first steady part st_{i-1}. Taking the median value over the whole steady subsequences increases the robustness of the event value ΔP_e.

In order to account for noise, for all methods, events e with a value ΔP smaller than a specified threshold are discarded.

III. Experimental Setup

In this section, the method that extracts personal information is described, decreasing granularity as a method for preserving privacy is briefly discussed and the used dataset is introduced.

A. Assessment of Appliance Use by Edge Detection

The goal of NIALM algorithms is energy disaggregation, which means that the interest lies in partitioning the consumed power into the portions used by individual appliances. In order to accurately measure the energy used by an appliance, the on-duration T_{on} of an appliance needs to be assessed precisely.

However, from a privacy viewpoint it is not necessarily important to assess the energy used by an appliance. In a privacy attack setting, the ultimate goal is the determination of private information like habits, personal properties or special circumstances. Since this information is typically not known in common data sets (including the REDD data set used here) this paper focuses on the determination of appliances together with a simple determination of activities within a household according to the causal chain (6). Regularly occurring activities could in turn provide information about e.g., habits, but such a study is out of scope of this paper. Here, the kind of an activity is inferred from the appliances that are used. For example, the activity cooking is inferred from the use of any one of the appliances stove, oven1, oven2, and microwave (compare also Table II).

The other important information about an activity is the usage time. For the description of an activity and possible inference of habits it is important when it takes place. Here, edge detection can provide the onset of an activity by providing the starting time of the corresponding appliance.

The second information is the duration of an activity or an appliance. The information about the duration could provide further information like e.g., the kind of meal that is cooked. Since no ground truth about activities is available and especially no details are known, it was decided not to assess the exact duration of an appliance. Moreover, initial trials showed that the matching of on- and off-events is far from being straightforward and would possibly limit the validity of results. Note that the matching applied by NIALM-algorithms for obtaining the on-durations needed for the assessment of the total energy used by an appliance is typically quite complicated. In order to keep the assessment clear and simple, it was decided to avoid the matching procedure. Instead, the on-duration of an appliance is simply measured as the time until the next off-event of this appliance occurs. Thus, typical on-durations of appliances are provided for the explanation of results in Section VI-B. However, the on-durations are never used for any other use including the determination of activities. Note that for FSMs the term "duration of stay in the present state" would be a more adequate name.

Since the signals of the available REDD-dataset [19] not only contain the mains but also the signals of the individual appliances, it is not necessary to compute the whole disaggregation. Instead, the following analysis focuses on the determination of events, which can directly be done using the edge detection methods of Section II-B.

B. Decreasing Time Granularity for Privacy Enhancement

Several possibilities for decreasing the time granularity exist. Considering a single time interval, different statistics could be computed. The most straightforward statistic is the average load value which should suffice for most practical solutions such as standard billing or time-of-use billing. For pricing based on the maximum load or for control reasons, the maximum load needed during the time interval could be useful. Additionally, (uniform) sampling could be done, i.e., taking the load value at (evenly) spaced points in time.

In the experiments presented in this paper, three variants are used: taking the average and maximum load in a time interval, respectively, and uniform sampling.

C. Dataset

All experiments were done using the so-called low-frequency dataset of the publicly available REDD-dataset [19]. The dataset contains measurements of the apparent power for six different houses. Measurements are available for mains1 and mains2, for some circuits, e.g., kitchen outlets and for individual appliances.

Although the analyzes were performed for all six houses, the evaluation is shown for house 1 only. House 1 has a relatively high number of measured appliances or circuits and includes labeled measurements both for high and low power appliances. The overlap of the power values of individual appliances is rather low, so that a possible increase in the overlap due to lower time resolutions could be detected.

One of the kitchen outlets, one of the washer dryers and the electric heat appliances showed less than three events at the highest time resolution and were excluded for further analyzes. Due to its automatic working mode, in the privacy attack setting the refrigerator is more a disturbing noise appliance than a privacy relevant appliance. A mains appliance was created as the sum of mains1 and mains2 by interpolating the values of mains2 to the values mains1 at the highest time granularity.

IV. Descriptive Event Detection Results

In this section, event detection is applied to the load curves of individual appliances. First, the quality of different event detection methods is assessed (Section IV-A), then it is shown how the overlap of events of different appliances affects the precision of subsequent classification algorithms (Section IV-B).

A. Comparison of Event Detection Methods

Since the results below are based on the events found, the performance of the event detection methods is assessed first. The evaluation is mainly done visually.

Generally, transient passing and edge merging yield good and very similar results (upper panel of Fig. 1). Note that the load curve is quite complex, especially power levels are not necessarily constant. As expected, the simple difference method yields more, but disturbing events and can therefore not be recommended as is (upper panel of Fig. 1).

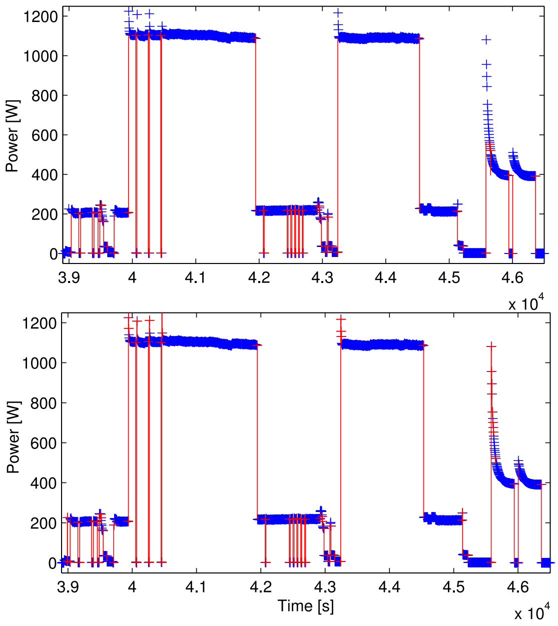


Fig. 1. Dishwasher events, marked as "+," detected using the transient passing method (upper panel) and the difference method (lower panel).

High-power devices such as heating are usually purely ohmic and consume high power values with a greater deviation of values. For low-power devices such as lighting the deviation of values is smaller. This leads to a tradeoff between noise removal and detection of events. If the noise threshold is set too low, the noise of high-power devices exceeds the threshold resulting in additional, unwanted edges. A noise threshold that is set too high in turn leads to a loss of events for low power devices. For all subsequent evaluations we used 20 W as noise threshold.

While for high time resolutions the edge merging and transient passing methods give very similar results, for lower time resolutions the transient passing method is more robust in determining the edge values. The results of the transient passing and edge merging methods turned out to be quite insensitive to the kind of statistic. For lower time resolutions the performance of the difference method is better with taking the max statistic or with sampling than with taking the average statistic. If not stated otherwise, the remaining analyzes will use the transient passing method and the average statistic.

B. Description of Events

This subsection contains a visual description of the events that occurred. Since the mains signal—which was generated by summing up the mains1 and mains2 signals—is supposed to contain the events of all appliances, the time between subsequent events is smaller than for the events of a single appliance. As a check that this property does not negatively influence the event detection of mains, the events of mains and the events of the individual appliances are compared in

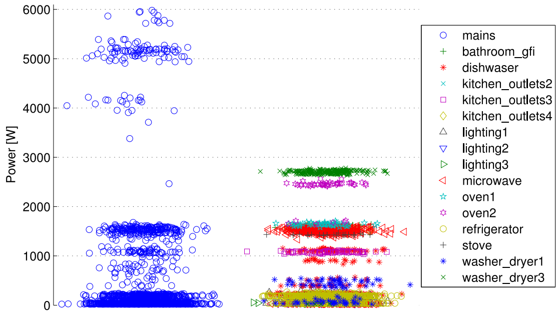


Fig. 2. Events of mains (left) compared with single appliances' events (right).

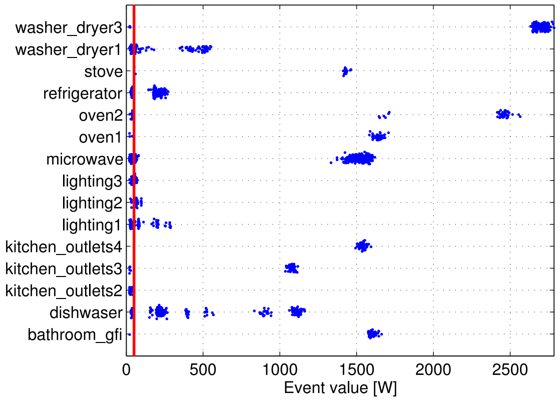


Fig. 3. Overlap of events for house 1.

Fig. 2. In fact, there is clear connection between events of the mains signal and the events of the individual appliances.

However, there are events that only occur for mains but not in any of the single appliances signals (Fig. 2). Additionally, there are events from appliances that do not occur in the mains signal (Fig. 2). This happens for all houses. To rule out that this could be an effect of bad event detection, both the absence of the appliance events and the presence of additional mains events was verified by visual inspection of the load curves. Due to this inconsistency of the mains signals and the signals of single appliances it was decided that all further analysis steps should be done with the load curves of the individual appliances only ignoring the mains signal.

Analyzing the quality of edge detection, for some high-power appliances unwanted noise events below 50 W are detected. Events below 50 W (left to the red, dashed line in Fig. 3) are considered as being hard to assign to appliances due to the high overlap of several appliances within this region. Therefore these events are discarded for further evaluation.

Even without performing a NIALM-analysis, the overlap of events stemming from different appliances can give valuable insights into the possibilities of disaggregation of the mains signal (Fig. 3). Appliances whose events have low overlap with other appliances' events, like e.g., washer_dryer3 will be easier to distinguish from them than appliances with high overlap

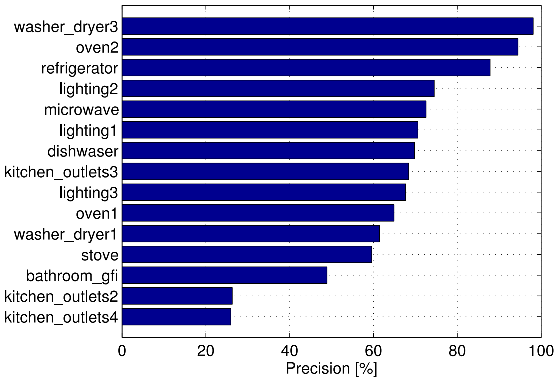


Fig. 4. Precision at highest resolution (3 s).

such as e.g., kitchen_outlets4 (compare also with Fig. 4). It should be noted that the negative events of an appliance typically have the same absolute values as the positive events, thus only the positive events are shown here.

V. Evaluation of Appliance Determination Ability

According to the causal chain (1), the first step in the determination of private information is considered, i.e., the ability to determine appliance use is evaluated

$$\text{Load Curve} \Rightarrow \text{Appliance Use.} \quad (3)$$

Note that the subsequent analysis models the detection of a given appliance. Due to the reasoning stated in Section III-A the assessment of the on-duration of appliances is not evaluated.

A. Classification Method

The chosen methodology can be identified more clearly, when the problem is stated in another form. Considering a detected event, one wishes to know, which appliance this event is stemming from. This is exactly a multiclass classification problem where the number of classes is the number of appliances. This multiclass classification problem is split into several one versus all two-class classification problems, one classification problem for each appliance. The input is the 1-D value of the event to be classified, the output is the information, if the event is stemming from this appliance or from one of the other appliances. Due to this setting, natural measures for appliance detection performance are precision and recall of classification. If a single performance value is required, the F-score (2) can be used. In contrast to a normal classification scenario where a good performance is requested, here small values are desirable with respect to privacy preservation.

It is expected that the overlap affects the precision of the classification task. Appliances with negligible overlap of their event values with event values of other appliances, such as washer_dryer3 and oven2, are expected to lead to simple classification problems with high precision. The precision of the corresponding classification problem is expected to decrease with increasing overlap.

Of course, more sophisticated analyzes could be done exploiting, e.g., the periodicity of the refrigerator or the typical duration between events of the appliance [15]. The information about the time of the day when the appliance was used could be taken into account [15], too. A dishwasher run consists of a series of events with different event values. The fact that different runs all look very similar to the time pattern of events shown in Fig. 1 could be exploited as follows. Event values of kitchen_outlets3 have similar values as one particular level of the dishwasher values (Fig. 3). Looking at the statistics of events over some past time window, if some other event values of the dishwasher do not occur, the dishwasher could be ruled out and thus kitchen_outlets3 could be distinguished from dishwasher. The same argument could be applied to washer_dryer1 and the dishwasher. However, such a detailed analysis is not the scope of this paper.

For the sake of simplicity, as classification algorithm the nearest neighbor method using three nearest neighbors is used. The resulting precision of the several two-class classification tasks for the highest time resolution is shown in Fig. 4. Precision is typically in the range between 60% and 80% with a maximum precision for washer_dryer3 of nearly 100%. By comparing Figs. 3 and 4, the negative influence of the overlap with events from other appliances on the precision is evident.

Note that here no direct NIALM analysis was done. Instead, only the event-values of the individual appliances (or circuits) are directly taken in order to analyze possible NIALM performance. The result can be used for an optimistic (in the sense of precision) estimate for the precision of a NIALM analysis, if several assumptions hold. The first assumption requires that the mains signal is the sum of the individual appliances loads plus a possible constant offset value which has no influence on events. Secondly, the noise must be of equal size both for all individual appliances and for the mains signal. Thirdly, and most importantly, the one-at-a-time condition which is a special form of the switch continuity principle [9] is assumed to be fulfilled. This condition states, that during each time interval at most one of the appliances changes its state.

B. Method Evaluating the One-at-a-Time Condition

The one-at-a-time condition is already known as a common necessary condition for some NIALM algorithms [9]. When more than one appliance change their state the edges of the aggregated signal are the sum of the individual edges. This leads to a much bigger search space of possible solutions which must be handled by the NIALM algorithm. Additionally, when more than one combination of appliances have the same aggregate edge value, ambiguities arise.

The classification method above looks at the signals of single appliances. Consequently, the one-at-a-time condition is ignored. The information about each appliance is obtained by separately applying the edge detection algorithm on the signal of each single appliance. However, in a usual setting, only the aggregate signal is given, thus hardening the disaggregation problem. The one-at-a-time condition suggests that a change of an appliances' state can only be detected, if only this single appliance changes its state during the measurement interval. For the assessment of the one-at-a-time condition, for each event found, it is checked, if this is the case or not.

First, the edges are computed from the individual signals of all appliances at the highest time resolution available and all event times are evaluated. An event is the only event within a measurement interval, if the duration to both the previous and the next event time exceeds the measurement interval. If the smaller duration is less than the measurement interval an event can be classified as single event, otherwise an event is classified as a coincidental event. As a performance measure now the proportion of single events for each appliance and measurement interval is calculated. Also here, small values are desirable with respect to privacy preservation.

VI. Influence of Time Granularity on Appliance Classification

In this section, the influence of time granularity Δt on precision and recall of the classification method shown above is studied.

A. Influence of Time Granularity on Recall

In a normal NIALM classification setting, the recall of a given appliance is defined as the proportion of events stemming from this appliance that can be found in the aggregate signal. However, due to the unknown differences between the mains signals and the signals of the individual appliances signals (Fig. 2), it was decided not to use the aggregate signal. As a consequence, the recall cannot be evaluated directly. In order to assess a quantity similar to the recall rate, the numbers of detected events of an appliance are compared for different time resolutions. Considering the events found at the highest resolution as ground truth, the number of events found at different time granularities can be normalized by this ground truth. Since the goal of this paper is studying the changes that arise due to changes in time resolution, this normalized number of events sufficiently serves as a measure of the recall rate. This measure for the recall is too optimistic because it is assumed that the recall at the highest resolution is 100% and the events of the appliances are found from the appliances signals instead of the mains signal. This overestimated recall measure goes down to near zero with decreasing granularity (Fig. 5) which is sufficient for a decrease of the exact recall rate.

In the privacy setting, the decrease of the recall to near zero means that with the time interval exceeding an appliance-specific threshold, a device will not be detected any more. Undetectability of devices in turn increases privacy.

B. Influence of On-Duration on the Recall

Fig. 5 shows that the recall of the appliances decreases with increasing measurement interval Δt. The measurement interval Δt_{drop} where this decrease takes place differs among appliances. This appliance-dependent quantity is denoted as drop-time. This subsection shows that the property of the appliances by which this critical duration is influenced is the on-duration T_{on}.

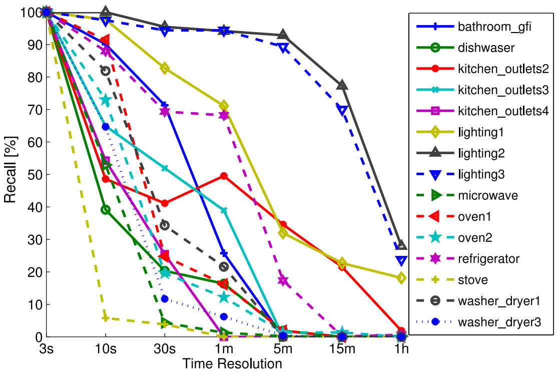


Fig. 5. Recall dependent on time granularity.

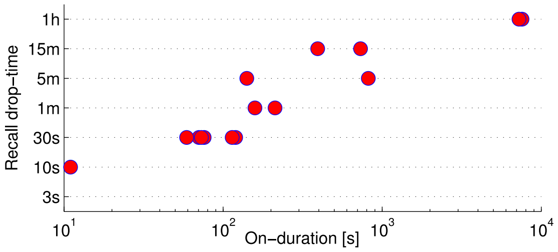


Fig. 6. Drop time Δt_{drop} and median on-durations of different appliances.

For an experimental assessment of this influence, for each appliance the drop-time Δt_{drop} of the recall is assessed as the time granularity where the recall in Fig. 5 is below 30% at first time. The value 30% was chosen for making Δt_{drop} robust against false positive events. A comparison of the obtained recall drop-time Δt_{drop} and the on-duration of the appliances in Fig. 6 shows a clear increase in drop time with increasing on-duration.

The connection between the on-duration and the drop of the recall can be explained by the mechanism of the transient passing method applied to a simple on–off-appliance with fixed on-duration T_{on}. For ease of explanation sampling of values is assumed. The transient passing method detects an on-state as a steady sequence of at least n values with higher energy consumption. As in [9], in this paper, n is set to 3 which is one of the smallest possible choices for n having thus a good detection property with reasonable robustness. If the on-duration T_{on} is too small, $T_{\text{on}} < (n-1)\Delta t = 2\Delta t$, at most two subsequent values can have higher loads which is just not enough to detect the on-state. Consequently, no change from or to the on-state can be detected. Rewriting this condition, the recall rate should drop to zero, if the time interval Δt exceeds a threshold which depends on the on-duration

$$\Delta t > \Delta t_{\text{drop,ideal}}(T_{\text{on}}) = \frac{T_{\text{on}}}{2}. \quad (4)$$

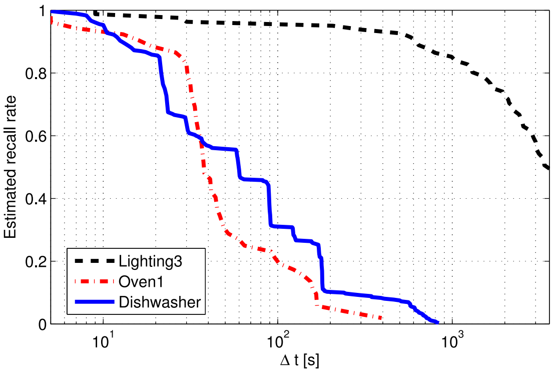


Fig. 7. Recall of 3 different appliances, estimated by the rule of thumb (5).

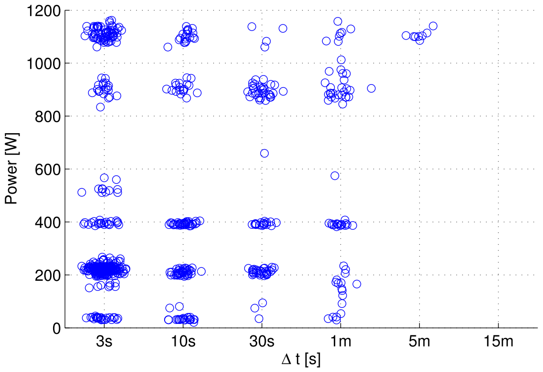


Fig. 8. Robustness of dishwasher event values when determined with transient passing edge detection.

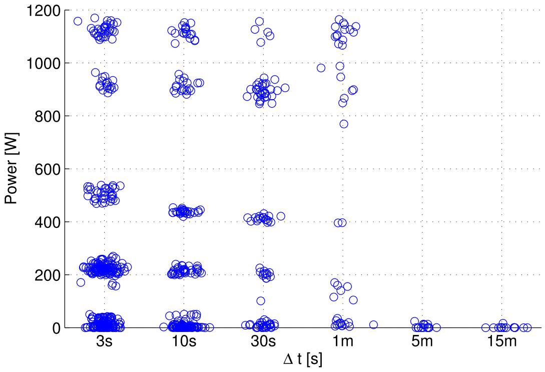


Fig. 9. Dishwasher event values determined with the edge merging method.

Using this connection, the knowledge about the on-duration of an appliance—which is often available as an initial guess without any NIALM-like analyzes—can be used for estimating the time interval Δt needed to significantly decrease the recall rate. If the time interval exceeds half of the typical on-duration of an appliance, a considerable proportion of events stemming from this appliance cannot be detected any more. Using the cumulative distribution of on-durations $F(T_{\text{on}})$, this rule can be formalized: dependent on the measurement interval Δt, an approximation for the recall rate $R(\Delta t)$ can be calculated as

$$R(\Delta t) = 1 - F(\Delta t/2). \tag{5}$$

This estimated recall rate of events is illustrated in Fig. 7 for lighting3, oven1 and the dishwasher. Despite the different choice of x-axes a strong similarity to Fig. 6 can be noticed. Due to the long on-durations, lighting3 exhibits high recall rates. The different on-durations of the dishwasher-states result in a staircase-like recall-curve.

C. Influence of Time Granularity on Precision

After studying the influence of the time resolution on the recall rate in Section VI-A, now the precision for the remaining events of the remaining appliances is investigated.

Interestingly, for increasing time interval Δt the precision for the classification of the remaining events keeps being high. This behavior is illustrated for house 1 and a time interval of 15 min. Due to the low recall, only four out of 15 appliances/circuits are still detectable. The precision of classification for these four remaining appliances is even higher than for the highest time resolution. One reason for this behavior is that a four-class classification problem is much simpler than a 15-class classification problem.

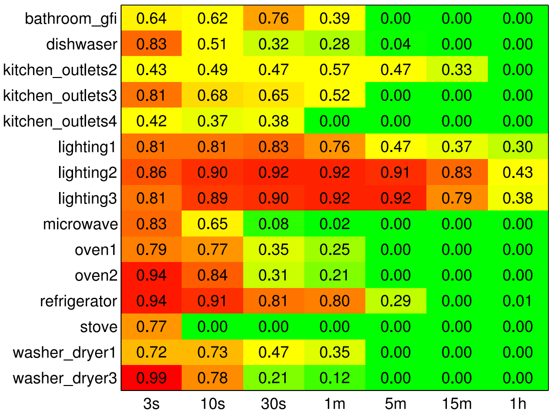
Another prerequisite for this behavior is the surprisingly robust estimation of the event values which is exemplarily shown for the dishwasher in Fig. 8. This stability property only holds for the transient passing method. For the edge merging method event values are relatively stable but show a slight decrease of event values (Fig. 9) while for the difference method event values get smeared for decreased time resolution (not shown). The amount of smearing for the difference method is most pronounced for the averaging-statistic.

VII. Understandable Privacy Analysis

This section aims at presenting the results about the influence of time granularity. As an important requirement, these results should be easily understandable and thus be suitable for unexperienced people like end-users or other decision makers. The influence of the time resolution is discussed in two parts: 1) the first part shows the influence on appliance use detection and 2) the second part shows the influence on higher-level personal information.

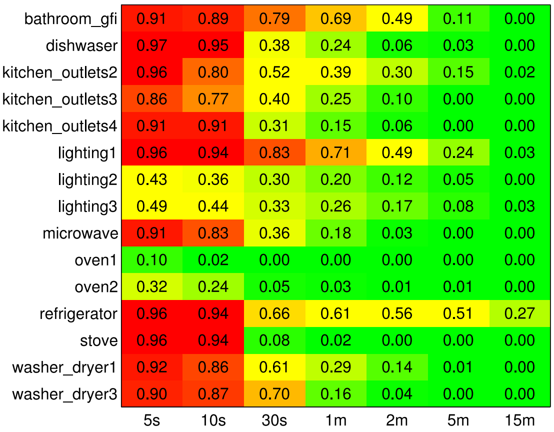
A. Detection of Appliance Use

An appliance can provide insights into personal information only if it can be detected and if the precision of detection is high. An appliance with these two properties will be called measurable. Measurability of an appliance itself does not necessarily imply danger for privacy, because appliances that are automatically controlled such as the refrigerator do not provide personal information even if their operational states are known. In contrast, nonmeasurability does imply privacy-safety which is the property that should be assessed here. Measurability



	3s	10s	30s	1m	5m	15m	1h
bathroom_gfi	0.64	0.62	0.76	0.39	0.00	0.00	0.00
dishwaser	0.83	0.51	0.32	0.28	0.04	0.00	0.00
kitchen_outlets2	0.43	0.49	0.47	0.57	0.47	0.33	0.00
kitchen_outlets3	0.81	0.68	0.65	0.52	0.00	0.00	0.00
kitchen_outlets4	0.42	0.37	0.38	0.00	0.00	0.00	0.00
lighting1	0.81	0.81	0.83	0.76	0.47	0.37	0.30
lighting2	0.86	0.90	0.92	0.92	0.91	0.83	0.43
lighting3	0.81	0.89	0.90	0.92	0.92	0.79	0.38
microwave	0.83	0.65	0.08	0.02	0.00	0.00	0.00
oven1	0.79	0.77	0.35	0.25	0.00	0.00	0.00
oven2	0.94	0.84	0.31	0.21	0.00	0.00	0.00
refrigerator	0.94	0.91	0.81	0.80	0.29	0.00	0.01
stove	0.77	0.00	0.00	0.00	0.00	0.00	0.00
washer_dryer1	0.72	0.73	0.47	0.35	0.00	0.00	0.00
washer_dryer3	0.99	0.78	0.21	0.12	0.00	0.00	0.00

Fig. 10. F-score matrix. Small values are desirable for privacy.



	5s	10s	30s	1m	2m	5m	15m
bathroom_gfi	0.91	0.89	0.79	0.69	0.49	0.11	0.00
dishwaser	0.97	0.95	0.38	0.24	0.06	0.03	0.00
kitchen_outlets2	0.96	0.80	0.52	0.39	0.30	0.15	0.02
kitchen_outlets3	0.86	0.77	0.40	0.25	0.10	0.00	0.00
kitchen_outlets4	0.91	0.91	0.31	0.15	0.06	0.00	0.00
lighting1	0.96	0.94	0.83	0.71	0.49	0.24	0.03
lighting2	0.43	0.36	0.30	0.20	0.12	0.05	0.00
lighting3	0.49	0.44	0.33	0.26	0.17	0.08	0.03
microwave	0.91	0.83	0.36	0.18	0.03	0.00	0.00
oven1	0.10	0.02	0.00	0.00	0.00	0.00	0.00
oven2	0.32	0.24	0.05	0.03	0.01	0.01	0.00
refrigerator	0.96	0.94	0.66	0.61	0.56	0.51	0.27
stove	0.96	0.94	0.08	0.02	0.00	0.00	0.00
washer_dryer1	0.92	0.86	0.61	0.29	0.14	0.01	0.00
washer_dryer3	0.90	0.87	0.70	0.16	0.04	0.00	0.00

Fig. 11. Proportion of single events. Small values are desirable for privacy.

can be assessed using the commonly used F-score which is computed from recall r and precision p by (2). Arranging the F-scores for all appliances and all time resolutions the resulting matrix can be visualized by a heatmap as shown in Fig. 10. There, the privacy-harmless appliances having low F-scores are colored green, while measurable and thus potentially privacy-decreasing appliances having high F-scores are colored red or orange.

The visualized F-score matrix (Fig. 10) clearly shows that measurability decreases and consequently privacy increases with increasing time interval.

Interestingly, measurability not necessarily decreases with increasing time interval. For example the F-score of appliance bathroom_gfi is maximal at a time interval of 30 s. This behavior can be explained by the high overlap of its event values with the events values of the appliances microwave, oven1, oven2, and kitchen-outlets4 (Fig. 3). This overlap leads to a rather small precision and consequently small F-scores of bathroom_gfi at high time resolution. However, the other appliances have shorter on-durations than bathroom_gfi. The short on-duration leads to a sharp drop of their recall at a time interval of 30 s. For bathroom_gfi the drop at the 30 s interval is relatively small, the sharp drop occurs later at a time interval of 1 min (Fig. 5). Thus, since the masking events of the other appliances are not present at a 30 s interval the precision of bathroom_gfi increases from 47% at 10 s intervals to 81% at 30 s intervals. This increase in precision overcompensates the drop in recall from 90% to 71% leading to an increased F-score (from 0.62 to 0.76) and thus explaining why bathroom activities are only measurable at 30 s time intervals.

Assessing appliance use with the one-at-a-time condition method shows that the proportion of single events decreases with increasing time interval (Fig. 11) which again implies an increase of privacy.

Comparing the results of the two evaluation methods shows a similar behavior. The only big differences can be seen for lighting1 and lighting2 which look much more privacy-safe when evaluated by the one-at-a-time condition method. This increase in privacy compared to the F-score assessment can be explained by the fact, that this method considers all appliances at once instead of just a single appliance. For the chosen house lighting1 and lighting2 are strongly co-occurring, therefore the proportion of single lighting events is small already at a very fine time resolution. This dependence of appliances can not be modeled with the classification method which looks only at the event values and not at the event time.

B. Detection of Activities

Now, according to the causal chain (1), higher level privacy implications of the resulting matrices are illustrated

$$\text{Appliance Use} \Rightarrow \text{Activities, Presence/Absence.} \quad (6)$$

For ease of explanation, a privacy-threshold of 0.7 is introduced. Entries with higher values are classified as measurable, entries with lower values as unmeasurable. Thus, red or orange entries are regarded as privacy-relevant while green or yellow entries are regarded as privacy-safe.

Looking at the F-score matrix, for 1 h time intervals all appliances are privacy-safe. For a 1 min time interval only the lights are privacy-relevant (because of its automatic operation mode the refrigerator is regarded as safe in this analysis). Interestingly, increasing the time interval from 1 to 5 or 15 min only negligibly increases privacy here. Bathroom activities (bathroom_gfi) are only measurable at exactly 30 s time intervals. Cooking (stove, oven1, oven2, and microwave) and housework (washer-dryer and dishwasher) are privacy-safe for time intervals of 30 s or more. It should be noted that the kitchen outlets were not considered for this analysis due to the unclear nature of the corresponding appliances. The result of this short discussion is shown in Table II.

Considering the one-at-a-time condition evaluation method, already at a measurement interval of 2 min, all appliances are privacy-safe. As before, the increase in privacy compared to the F-score assessment can be explained by the co-occurrence of lighting1 and lighting2.

The results of Tables II and III should be seen as a first evaluation of privacy that is likely to be too optimistic. On one

TABLE II
TIME INTERVAL Δt NEEDED TO INFER DIFFERENT KINDS OF PERSONAL INFORMATION FOR HOUSE 1 USING THE F-SCORE, THRESHOLD 0.7

Information	Inferred from	Safe for
Presence/Absence	*Lighting*	$\Delta t \geq 1h$
Bathroom Activities	*bathroom-gfi*	$\Delta t \neq 30s$
Cooking	*stove, oven1, oven2, microwave*	$\Delta t \geq 30s$
Housework	*washer-dryer, dishwasher*	$\Delta t \geq 30s$

TABLE III
TIME INTERVAL Δt NEEDED TO INFER DIFFERENT KINDS OF PERSONAL INFORMATION FOR HOUSE 1 USING THE ONE-AT-A-TIME CONDITION, THRESHOLD 0.7

Information	Inferred from	Safe for
Presence/Absence	*Lighting*	$\Delta t \geq 2m$
Bathroom Activities	*bathroom-gfi*	$\Delta t \geq 1m$
Cooking	*stove, oven1, oven2, microwave*	$\Delta t \geq 30s$
Housework	*washer-dryer, dishwasher*	$\Delta t \geq 1m$

hand, this privacy analysis is based on the effect of an increased measurement interval on event detection. While fine-grained personal information is likely to be based on appliance events, it seems plausible that coarse information such as presence or absence could easily be found using other methods. Such methods could for example examine the difference in average power consumption for times where the inhabitants are present or absent. For the detection of certain activities it could be sufficient to distinguish different groups of appliances such as appliances used for cooking.

On the other hand, the choice of the value 0.7 as the privacy-threshold is quite arbitrary and mainly intended for demonstrating the privacy evaluation. Choosing this value as a threshold for the F-score, an appliance is considered measurable, if nearly each single event can be detected and distinguished from other appliances events. However, for the detection of regular personal habits it is not necessary to detect each single event, it is rather necessary to detect enough events during the recording time. Having data for long durations such as years, a lower recall rate could be considered privacy-relevant leading in turn to a lower acceptable F-score privacy-threshold. Looking at a thought experiment of an appliance used twice a day and a measurement duration of three years leads to approximately 300 events. Even one-third of these events would be enough to estimate typical usage times.

The privacy-threshold should also be chosen separately for each appliance. For example, one run of a dishwasher leads to many events. Although for a time interval of 30 s the F-score goes down to 0.32 (Fig. 10), the main big events are still detectable at this time granularity (Fig. 12, upper panel) suggesting that a lower threshold is needed for the dishwasher. Averaging over a 5 min interval, only one edge is left (Fig. 12, lower panel), using a 15 min interval, also this last event can not be detected any more suggesting for the dishwasher an F-score threshold of 0.04 or less. Despite these open issues, the usefulness of the performed evaluations for a first assessment of the impact of time granularity on personal information could be shown.

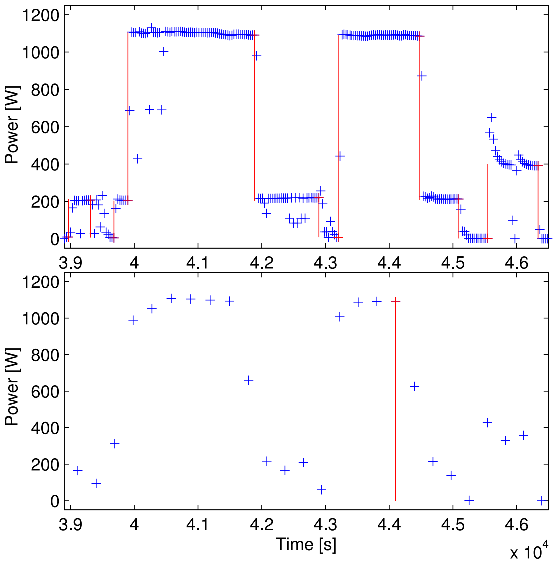


Fig. 12. Edges for dishwasher for Δt = 30 s (upper panel) and 300 s (lower panel).

VIII. CONCLUSION

Although being the simplest possible privacy enhancing technique, the impact of decreasing the time resolution on privacy analyzes of load signals obtained from smart metering to date has not been studied systematically. Since the first step in a privacy attack can consist of the assessment of appliance use which is in turn often based on edge detection methods, the influence of the time interval on edge detection methods applied on load signals is studied.

Using edge detection alone already leads to valuable insights about the disaggregation possibilities for different appliances, a full NIALM-analysis is not necessary. Appliances whose events have a small overlap with the events of the other appliances can more easily be disaggregated.

With increasing time interval, the recall, i.e., the proportion of detected edges stemming from a device decreases. This decrease is more pronounced for appliances with shorter on-durations. As a coarse rule of thumb, when the time interval exceeds half the typical on-duration of an appliance, the appliances event values cannot be reliably detected any more. For the house analyzed in detail, increasing the measurement interval to 15 min has the effect that only four out of 15 appliances/circuits remain detectable (three lighting circuits and the refrigerator). For these remaining appliances the disaggregation precision stays high, because even for high time intervals the transient passing edge detection method robustly determines edge values.

Privacy implications can be evaluated by F-score values or the proportion of single events of an appliance. Evaluating these values for different appliances and time granularities, the resulting matrices can be visualized. This visualization represents the impact of time granularity on privacy in an easily understandable way suited for nonexperts like the users themselves or other decision makers.

For the next natural steps toward privacy evaluation datasets that include personal information or activity logs are needed enabling a more direct assessment of personal information. Such data would be the basis for finding a well-founded way of choosing privacy-thresholds, an evaluation method that combine the two methods proposed here or other methods especially designed for low measurement intervals.

REFERENCES

[1] R. Segovia and M. Sánchez, "Set of common functional requirements of the smart meter," DG INFSO and DG ENER, European Commission, Brussels, Belgium, Tech. Rep. 73, Oct. 2011. [Online]. Available: http://ec.europa.eu/energy/gas_electricity/smartgrids/doc/2011_10_smart_meter_funtionalities_report_full.pdf

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

[3] A. Cavoukian, J. Polonetsky, and C. Wolf, "SmartPrivacy for the smart grid: Embedding privacy into the design of electricity conservation," *Identity Inf. Soc.*, vol. 3, no. 2, pp. 275–294, 2010. [Online]. Available: http://dx.doi.org/10.1007/s12394-010-0046-y

[4] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin, "Private memoirs of a smart meter," in *Proc. 2nd ACM Workshop Embedded Sens. Syst. Energy-Eff. Build. (BuildSys)*, New York, NY, USA, 2010, pp. 61–66. [Online]. Available: http://doi.acm.org/10.1145/1878431.1878446

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

[6] D. Engel, "Wavelet-based load profile representation for smart meter privacy," in *Proc. IEEE PES Innov. Smart Grid Technol. (ISGT)*, Washington, DC, USA, Feb. 2013, pp. 1–6. [Online]. Available: http://dx.doi.org/10.1109/ISGT.2013.6497835

[7] D. Engel and G. Eibl, "Multi-resolution load curve representation with privacy-preserving aggregation," in *Proc. IEEE Innov. Smart Grid Technol. (ISGT)*, Copenhagen, Denmark, Oct. 2013, pp. 1–5.

[8] G. Eibl and D. Engel, "Influence of data granularity on nonintrusive appliance load monitoring," in *Proc. 2nd ACM Workshop Inf. Hiding Multimedia Security*, Salzburg, Austria, 2014, pp. 147–151. [Online]. Available: http://doi.acm.org/10.1145/2600918.2600920

[9] G. Hart, "Nonintrusive appliance load monitoring," *Proc. IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992.

[10] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE Trans. Consum. Electron.*, vol. 57, no. 1, pp. 76–84, Feb. 2011.

[11] D. C. Bergman *et al.*, "Distributed non-intrusive load monitoring," in *Proc. IEEE/PES Conf. Innov. Smart Grid Technol. (ISGT)*, Anaheim, CA, USA, Jan. 2011, pp. 1–8.

[12] M. Baranski and J. Voss, "Genetic algorithm for pattern detection in NIALM systems," in *Proc. IEEE Int. Conf. Syst. Man Cybern.*, The Hague, The Netherlands, 2004, pp. 3462–3468.

[13] E. Vogiatzis, G. Kalogridis, and S. Z. Denic, "Real-time and low cost energy disaggregation of coarse meter data," in *Proc. 4th IEEE PES Innov. Smart Grid Technol. Europe (ISGT Europe)*, Lyngby, Denmark, 2013, pp. 1–5.

[14] J. Z. Kolter and T. Jaakkola, "Approximate inference in additive factorial HMMs with application to energy disaggregation," *J. Mach. Learn. Res. Proc. Track*, vol. 22, pp. 1472–1482, Apr. 2012.

[15] H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han, "Unsupervised disaggregation of low frequency power measurements," in *Proc. 11th SIAM Int. Conf. Data Min.*, Mesa, AZ, USA, 2011, pp. 747–758.

[16] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "Non-intrusive load monitoring using prior models of general appliance types," in *Proc. 26th Conf. Artif. Intell. (AAAI)*, Toronto, ON, Canada, 2012, pp. 356–362.

[17] U. Greveler, B. Justus, and D. Löhr, "Multimedia content identification through smart meter power usage profiles," in *Proc. Int. Conf. Inf. Knowl. Eng. (IKE)*, Las Vegas, NV, USA, 2012.

[18] A. Marchiori, D. Hakkarinen, Q. Han, and L. Earle, "Circuit-level load monitoring for household energy management," *Pervasive Comput.*, vol. 10, no. 1, pp. 40–48, Jan./Mar. 2011.

[19] J. Kolter and M. Johnson, "REDD: A public data set for energy disaggregation research," in *Proc. Workshop Data Min. Appl. Sustain. (SIGKDD)*, San Diego, CA, USA, 2011, pp. 1–6.

Günther Eibl (M'13) received the Ph.D. degree in mathematics and the M.Sc. degree in physics from the University of Innsbruck, Innsbruck, Austria, in 1997 and 2002, respectively.

He is a Research Associate with the Josef Ressel Center for User-Centric Smart Grid Privacy, Security, and Control, Salzburg University of Applied Sciences, Puch/Salzburg, Austria. He held research positions at the Institute of Biostatistics and the Theoretical Physics Institute, Innsbruck. His current research interests include extraction of information from data with a focus on statistical modeling, data mining, and privacy preserving technologies.

Dominik Engel (S'06–M'08) received the Ph.D. degree in computer science from the University of Salzburg, Salzburg, Austria, in 2008.

He is a Professor with the Salzburg University of Applied Sciences, Puch/Salzburg, Austria, where he heads the Josef Ressel Center for User-Centric Smart Grid Privacy, Security, and Control. He was a Researcher at the University of Bremen, Bremen, Germany, and the University of Salzburg, and the Product Manager at Sony Digital Audio Disc Corporation, Anif, Austria, where he was responsible for video content security. His current research interests include smart grid security and privacy, multimedia security, and technological methods for enhancing end-user trust.

3.3 UNTERWEGER15A

- A. Unterweger and D. Engel. Resumable load data compression in smart grids. *IEEE Transactions on Smart Grid*, 6(2):919–929, March 2015.

Resumable Load Data Compression in Smart Grids

Andreas Unterweger, *Student Member, IEEE,* and Dominik Engel, *Member, IEEE*

***Abstract*—We propose a compression approach for load profile data, which addresses practical requirements of smart metering. By providing linear time complexity with respect to the input data size, our compression approach is suitable for low-complexity encoding and decoding for storage and transmission of load profile data in smart grids. Furthermore, it allows for resumability with very low overhead on error-prone transmission lines, which is an important feature not available for standard time series compression schemes. In terms of compression efficiency, our approach outperforms transmission encodings that are currently used for electricity metering by an order of magnitude.**

***Index Terms*—Compression, evaluation, load data, resumability.**

I. Introduction

SMART GRIDS rely on information and communication technology to measure, transfer, and manage detailed data on grid status. Smart metering is an important component of this system, providing detailed data in the distribution network. This data forms one of the key components for use-cases in the smart grid, such as energy feedback [1], grid monitoring, and load forecasting [2].

The most important arguments for compressing smart meter data are discussed in detail below: 1) data volume; 2) communication bandwidth; and 3) energy efficiency. Each of these arguments is valid for almost all use-cases of smart metering. However, the degree to which compression is advantageous, depends on the specific requirements of the use-case, such as the volume of data produced in smart metering, the need for (near) real-time transmission, or the predominant direction of communication (while some use-cases, e.g., real-time pricing, push data to the meter, most use-cases involve the meter transmitting data to a data concentrator). Compression is needed most for use-cases which generate a high volume of data, such as monitoring of grid stability, which requires fine-grained data with low delay.

A. Data Volume

While in the traditional billing use-case, the data volume is very small and, therefore traditionally there was no need for data compression whatsoever (even for automated meter reading), it is evident that for the use-cases in the smart grid, data volume increases dramatically: for instance, the data volume of load profile data at a granularity of 1 s and double-precision floating point for the 40 million households in Germany amounts to 25 TB of raw data per day. With the method presented in this paper, this volume can be reduced by nearly 90% to approximately 2.6 TB (assuming data properties similar to the test data). This reduction is not only beneficial in terms of reducing the volume of transmitted data, but also positively impacts storage requirements at the Distribution System Operator.

B. Low Bandwidth

Many smart meters will be connected to low-bandwidth communication links, such as powerline communication (PLC) links. Compression is an important tool to make the best use of the available bandwidth. For example, PLC is more reliable for lower data rates. Through compression, reliability can therefore be increased. Another example for a benefit of compression is the number of communicating parties. In some scenarios a number of smart meters need to communicate within the same network segment, often using collision detection or avoidance. The probability of collision increases with the volume of data each smart meter tries to transmit in the same time interval, up to a point where communication becomes impossible. With compression, the data rate can be reduced and with it the probability of collision. Therefore, compression enables more smart meter to communicate in the same multiple-access segment.

C. Energy Efficiency

The case for data compression of load profiles is also supported from the perspective of energy consumption. The idea of smart grids is closely linked to increasing energy efficiency. Care should be taken for the components of the smart grid to also reflect this endeavor through economical use of energy. The power required for the transmission of bits significantly exceeds the power required for the computational complexity of compression algorithms (e.g., on the *Mica2dot* platform, for the power needed to transmit 1 bit, more than 2000 clock cycles can be executed [3]). Thus, the employment of compression methods saves energy and the effect is multiplied by the vast number of smart meters in the field. The computational capabilities of smart meters will definitely support compression methods such as suggested in this papers (smart meters will need to support at least basic cryptographic primitives [4], which are more demanding than the operations needed for compression as presented here).

Manuscript received March 1, 2014; revised July 2, 2014; accepted October 10, 2014. Date of publication November 6, 2014; date of current version February 16, 2015. This work was supported in part by the Austrian Federal Ministry of Science, Research, and Economy, and in part by the Austrian National Foundation for Research, Technology, and Development. Paper no. TSG-00181-2014.

A. Unterweger is with the University of Salzburg, Department of Computer Sciences, Salzburg A–5020, Austria; and also with the Salzburg University of Applied Sciences, Josef Ressel Center for User-Centric Smart Grid Privacy, Security, and Control, Urstein Süd A-5412, Austria.

D. Engel is with the Salzburg University of Applied Sciences, Josef Ressel Center for User-Centric Smart Grid Privacy, Security, and Control, Urstein Süd A-5412, Austria (e-mail: dominik.engel@en-trust.at).

Digital Object Identifier 10.1109/TSG.2014.2364686

Apart from good compression, a method for compressing load data should fulfill a number of other requirements.

1) Low to moderate computational complexity to keep power and processor requirements for smart meters low.
2) Low memory requirements (e.g., the use of large dictionaries makes the smart meter unnecessarily expensive).
3) Low overhead for initialization.
4) *Ability to Resume After Interruption:* If the communication link to a smart metering device is temporarily down, synchronizing the compression algorithm needs to be fast and efficient.

We propose an approach to compression of load profile data that fulfills all of the above criteria.

The rest of this paper is structured as follows. Section II gives an account of related work in the areas of load data representation and time-series compression. Section III describes the characteristics of load data, motivating the design of our proposed compression approach, which is presented in Section IV and analyzed in detail in Section V. In Section VI, our approach is evaluated and compared to existing representations with respect to transmission size and computation time. Finally, we provide an outlook in Section VII before we conclude our paper in Section VIII.

II. Related Work

In [5], standard general-purpose compression algorithms are applied to publicly available load data set. The evaluation shows that load data is well suited for compression. We use the same data set in this paper and conduct a more detailed analysis on compressibility of load data. Using the same data set, we can also show that the approach proposed here is comparable to standardized methods in terms of compression performance, while offering additional features such as resumability, which are important for real-world use.

Compression has been proposed in other areas of the smart grid. The compression of phasor measurement units (PMUs) data is the field most closely related to smart meter readings compression. In [6], different data compression techniques for PMU readings are discussed and evaluated. Ning *et al.* [7] proposed a wavelet-based compression technique for the readings of PMUs. In a similar vein, Khan *et al.* [8] proposed the use of the embedded zerotree method for PMU measurements. While approaches for compressing PMU data are relevant to the subject area considered here, the compression of load data from smart meters differs significantly from PMU readings, by: 1) the origin of the data and consequently the properties of the data; 2) the number of sensors in the field, which is vastly higher in smart metering; and 3) the requirements for practical operation, such as real-time transmission of values.

In the general area of time-series compression, there are a number of contributions, the most notable and active field being audio compression [9]. Another active research area in compression is, of course, centered on video (see [10]). Methods from both fields can be considered for adaptation for load data compression. In fact, some approaches from the area of motion data compression show potential to prove useful when adapted to load data, as will be discussed below.

In practical operation in energy grids, currently no compression is applied by any of the standardized data formats in smart metering. The "open smart grid protocol" [11], which is a protocol suite spearheaded by European Telecommunications Standards Institute (ETSI), defines a format for transferring metering data, using up to 16 channels in the same interval. "All channels are stored as total values (no differential values) [11, p. 34]," and no compression is applied.

The smart metering coordination group, working under EU standardization mandate M441, has defined a functional reference architecture for communications in smart metering systems in a CEN/CENELEC/ETSI technical report [12]. Data model standards and communications profile standards are considered in the report, but data compression is not addressed.

The "device language message specification" and the "companion specification for energy metering" provide data formats and communication standards for automatic meter reading. The relevant standards for the data format are IEC 62056-21 [13] and IEC 62056-53 [14]. A lower layer encoding for metering values, the A-XDR encoding rule, is specified by IEC 61334-6 [15].

III. Load Profile Data Characteristics

Load profiles are time series of electrical power consumption. While the measurement precision is configurable and use-case dependent to a large extent, all load profile data with similar sampling intervals exhibit certain characteristics, some of which will be described in this section.

Note that there may be several other data characteristics which depend on the use-case or are limited to a number of data sets. As we intend to describe general characteristics which apply to a large percentage or even all load profile data in a smart-meter scenario with second- to minute-granularity of sampling, we omit use-case- and data-set-specific characteristics.

On detailed examination of load profiles of consumer households, it can be noticed that, while most consecutive values within a load profile tend to exhibit small value differences between one another, few values exhibit large differences. Depending on the time difference between two consecutive values, this effect is more (e.g., when the sampling interval is in the range of seconds) or less dominant (e.g., when the sampling interval is in the range of minutes or even coarser).

In order to show that this is true for a large number of load profiles, we analyze the properties of consecutive values in a number of load profiles from different data sets. We use the low frequency Massachusetts Institute of Technology (MIT) Reference Energy Disaggregation Data Set (REDD) data set [16] (abbreviated as MIT data set henceforth) as well as the TU Darmstadt tracebase data set [17] (abbreviated as TUD data set henceforth).

The MIT data set consists of a total of 116 load profiles. Each load profile contains average power readings of one individual circuit from one of six houses. The data is sampled in intervals of 1 s with a precision of 0.01 watts, i.e., the

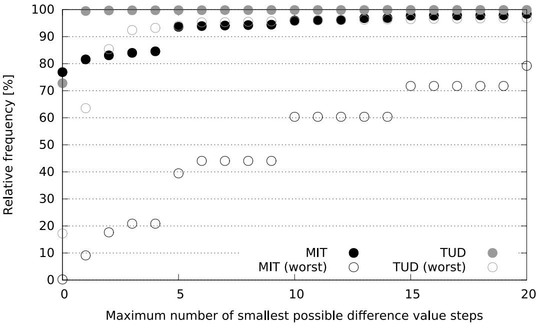


Fig. 1. Frequency of small consecutive value differences in the MIT (black) and TUD (gray) data sets. Filled circles denote the average relative frequency over all load profiles, while empty circles accentuate the load profile with the smallest relative frequency.

smallest nonzero difference between two consecutive values is 0.01 watts.

The TUD data set consists of a total of 1836 load profiles. Each load profile contains average power readings of one of 44 electric appliances. Like the MIT data, the data is sampled in intervals of 1 s with a precision of 1 watt, i.e., the smallest nonzero difference between two consecutive values is 1 watt.

Fig. 1 shows the relative frequency of the 20 smallest possible absolute value differences between consecutive values for both the MIT and the TUD data set. For example, 93% of all value differences in the load profiles of the MIT data set (illustrated by filled black circles) are between −0.05 and 0.05 watts (both inclusive), corresponding to the five smallest possible value differences. However, there is at least one data set (illustrated by empty black circles) in which only 40% of all value differences are within these limits.

The plots clearly show that the biggest part, i.e., more than 95%, of the consecutive value differences are between −0.1 and 0.1, corresponding to only ten of the smallest possible value steps. This is somewhat surprising considering that the maximum absolute value difference in the load profiles from MIT data set is as large as 6680.55 watts, corresponding to 6 68 055 value steps.

The distribution of the value differences in the TUD data set (gray filled circles) is even more surprising. Quasi all, i.e., more than 99.99%, of the consecutive value differences are −1, 0, or 1, although the absolute value differences are as large as 4879 watts maximum.

In both, the TUD and the MIT data set, less than the smallest 0.02% possible value differences (with respect to the maximum occurring value difference in all load profiles) make up more than 99% of all differences. Although, this is an average value summarizing all load profiles of the respective data set, it clearly shows that the load profile data tends to exhibit very small changes between two measurement points with respect to the corresponding data values.

As there is a high amount of load profiles per data set, the empty circles in Fig. 1 depict the characteristics of the load profiles with the lowest relative frequency of small value differences per data set for both, the MIT (black) and the TUD data set (gray).

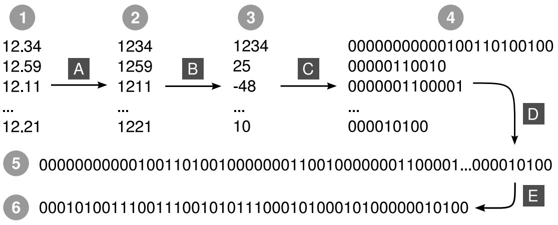


Fig. 2. Proposed compression approach: a list of values from a load profile (1) is transformed to a compressed bit string (6) through five successive encoding steps (A–E).

In the TUD data set, the worst case scenario in terms of the relative frequency of value differences of no more than 6 watts is one data set with 95% relative frequency. In other words, only 5% of all value differences in the worst case data set are larger than six. Similarly, in the worst case load profile in the MIT data set, there are only a little more than 20% of all value differences whose absolute value is greater than 20.

While both data sets contain load profiles where the number of small value differences is significantly smaller than on average, i.e., smaller than for all load profiles of the respective data set, the percentage of small value differences is very high and increases significantly with every additional value difference step.

Considering that the displayed value difference steps in Fig. 1 only cover about 0.003% (20 out of 6 68 055) and about 0.41% (20 out of 4879) of the actual value difference range of the MIT and TUD data set, respectively, this allows for the following conclusions on load profile data with second-granularity, thereby summarizing some of their compression-relevant characteristics.

1) Quasi all value differences of two consecutively sampled load profile data values are very small with respect to the possible range of value differences.
2) Large value differences are very rare, even in worst-case load profiles (in the analyzed data sets).

IV. Proposed Compression Approach

We propose a compression approach consisting of five steps (denoted as A–E) illustrated in Fig. 2, which exploits the load profile data characteristics described in Section III. Our algorithm takes a list of values from a load profile (1) as input and outputs a compressed binary representation of it (6). In the following, the five processing steps (A–E) are described.

A. Normalization

Floating point operations typically accumulate rounding errors due to their finite precision, which may lead to undesirable side effects. Furthermore, floating point operations are often more expensive in terms of computation time than their integer counterparts due to the lack of floating point units in most embedded systems [18]. Therefore, the first step of our approach (denoted as A in Fig. 2) is the conversion of floating point values to integer values, referred to as normalization.

As the precision of each value is bounded for both, technical and physical, reasons, so is the precision of a list of values. Let p_i denote the precision of the ith value v_i, expressed in the number of decimal places after the decimal point. Consequently, the precision of each value contained in the list of values is bounded by

$$p_{\max} = \max_i p_i. \tag{1}$$

Moving the decimal point of each value by $p_{\max}$ decimal places to the left, all values can be expressed as normalized integer values n_i with the same precision

$$n_i = v_i \cdot 10^{p_{\max}}. \tag{2}$$

This calculation is illustrated for $p_{\max} = 2$ as step A in Fig. 2. We observed that $p_{\max}$ is typically identical for all input values v_i for quasi all real-world load profile data since the measurement precision hardly ever changes.

Note that the normalization step may be omitted if the input values have no decimal places after the decimal point.

B. Differential Coding

As described in detail in Section III, the differences between consecutive values are mostly very small. This property can be exploited by differential coding, i.e., by storing only the differences between two consecutive values instead of the actual values. Note that this is closely related to differential pulse-code modulation [19].

The differential coding in our approach is illustrated as step B in Fig. 2. The differential values d_i are calculated from the normalized input values n_i from the previous step by a simple subtraction for all values but the very first

$$d_{i>0} = n_i - n_{i-1}. \tag{3}$$

The first value remains unchanged, since there is no reference value for it to be subtracted from

$$d_0 = n_0. \tag{4}$$

C. Variable Length Coding

In order to actually exploit the fact that a large number of difference values d_i are likely to be small (see Section III), a variable length code is needed. We use a zeroth order signed exponential-Golomb code as used in the H.264 standard [10] and described in detail in [20].

Although exponential-Golomb codes are optimal for geometrically distributed values [20] and the difference values d_i are unlikely to be exactly distributed in this way, we use this code as it is able to compactly represent small difference values, which are very likely to occur. Large difference values, which are not very likely to occur, may slightly affect coding efficiency.

Table I shows both signed and unsigned zeroth-order exponential-Golomb code words for the corresponding integer input values, i.e., possible d_i in our use-case. Note that it is necessary to use signed exponential-Golomb code words since the difference values d_i may be negative.

A value of zero can be encoded using just one bit. All other signed exponential-Golomb code words can be constructed by mapping the unsigned exponential-Golomb code words alternately to negative and positive values, respectively. Using this encoding, each difference value d_i is transformed into a corresponding variable length code word c_i as illustrated as in Fig. 2 (step C).

TABLE I
LIST OF EXEMPLARY VALUES AND THEIR RESPECTIVE SIGNED AND UNSIGNED ZEROTH ORDER EXPONENTIAL-GOLOMB CODE WORDS. HYPHENS DENOTE INVALID VALUES. ADOPTED FROM [21]

Value	Unsigned Exp.-G. code word	Signed Exp.-G. code word
...	–	...
-4	–	0001001
-3	–	00111
-2	–	00101
-1	–	011
0	1	1
1	010	010
2	011	00100
3	00100	00110
4	00101	0001000
...	...	...

D. Code Word Concatenation

To group the variable length code words c_i from the previous step for the subsequent step, they are concatenated to a bit string b as illustrated in Fig. 2 (step D). Note that no delimiters are required since the code words include implicit length information (the number of leading zero bits is equal to the number of value bits after the delimiting one bit).

E. Entropy Coding

As a final step, we perform entropy coding on the concatenated bit string b in order to get the final compressed bit representation e, as illustrated in step E in Fig. 2. By using an arithmetic coder, which theoretically allows perfect, i.e., zero-redundancy entropy coding under certain conditions [22], this aims at removing most of the remaining redundancy.

Since the code words c_i have variable length and are potentially large, we use binary arithmetic coding which operates on bit level and therefore only distinguishes two symbols—zero and one. Since the probabilities of these two symbols, which are required as an input for the arithmetic coder, may differ depending on the input values, we start with 50:50 probabilities and perform adaptive encoding, i.e., we modify the probabilities of the two symbols during the encoding process according to their actual occurrences.

In order to avoid floating point operations during arithmetic coding, an implementation relying only on integer operations is recommended. As described in Section VI, we use an implementation which is based on the algorithm proposed in [22]. Although, it is possible to faster implementations, see [20] or [23], the latter rely on approximations. Therefore, their final bit string length may in some cases be slightly different from our results.

F. Summary

As summarized in Fig. 2, the input values v_i (1) are normalized to integer values n_i (2), which are differentially coded as d_i (3). After each differentially coded value d_i is encoded as

TABLE II
WORST-CASE TIME COMPLEXITY OF OUR APPROACH AND ITS CORE STEPS WITH RESPECT TO THE NUMBER OF INPUT VALUES n AND THEIR MAXIMUM SIZE m IN BITS

Step	**Per value**	**Combined**
Differential coding	$O(m)$	$O(mn)$
Variable length coding	$O(m)$	$O(mn)$
Entropy coding	–	$O(mn)$
Combined	–	$\boldsymbol{O(mn)}$

TABLE III
WORST-CASE SPACE COMPLEXITY OF OUR APPROACH AND ITS CORE STEPS WITH RESPECT TO THE NUMBER OF INPUT VALUES n AND THEIR MAXIMUM SIZE m IN BITS

Step	**Per value**	**Combined**
Differential coding	$O(m)$	$O(m)$
Variable length coding	$O(m)$	$O(m)$
Entropy coding	–	$O(1)$
Combined	–	$\boldsymbol{O(m)}$

one variable length code word c_i (4), all code words are concatenated to a single bit string b (5), which is finally entropy coded as a compressed bit string e (6).

G. Decoding Process

Decoding a bit string encoded with our approach requires applying the inverse operations in the reverse order, i.e., the decoding equivalents of the encoding steps (E–A). These decoding steps are shortly described below.

First, entropy decoding (inverse of step E) is performed on the bit string, yielding a string of variable length code words, which can be split (inverse of step D) due to the implicit length information they contain (see Section IV-D). The exponential-Golomb code words are then decoded (inverse of step C) to yield difference values. Finally, these values are added to their respective predecessors (inverse of step B) in order to get the original absolute values, which can be denormalized through division (inverse of step A).

V. ALGORITHMIC PROPERTIES

The approach presented in Section IV exploits the characteristics of load profile data for compression. However, its applicability for a smart metering use-case is not obvious. Hence, this section analyzes its properties with a strong focus on practicality. It describes those features which are relevant when the proposed approach is used to process and transmit load profile data and provides a detailed overview of its time and space complexity.

A. Algorithmic Complexity

When encoding n values with a maximum size of i and m bits each before and after normalization, respectively, all subsequent steps of our proposed compression approach require processing time and memory depending on m, n or both. In this section, we derive the worst-case time and space complexity of our approach and the aforementioned steps. The results are summarized in Tables II and III.

We use asymptotic notation [24] and assume that all values are available in memory when they are needed by our algorithm, i.e., they are either all in memory the whole time, consuming $O(mn)$ bits, or loaded into memory one by one on demand, consuming $O(m)$ bits. Furthermore, we assume that the encoded data is either transmitted or stored immediately so that no temporary memory for the fully encoded bit representation has to be taken into account and the algorithm's space complexity analysis can focus on the overhead of the algorithm itself.

The first step, i.e., the normalization, largely depends on the format of the input values. If they are already integer, no operation needs to be performed. Otherwise, one multiplication for the i-bit input values, followed by an optional rounding operation, is required. This requires $O(i^2)$ time complexity when using a straight-forward implementation of an i-bit multiplication algorithm [24]. For the optional rounding operation, the same applies.

Since the normalization step itself is optional and, if performed, highly dependent on the input values and their format, its time and space complexity are not included in Tables II and III, respectively. It should be noted, however, that the space complexity of the multiplication is $O(i)$. If i is proportional to m, this is equal to $O(m)$.

The second step, i.e., the differential coding, can be performed value by value and requires only the last value to be stored in order to calculate the current value difference. This takes up m bits of space constantly, plus m bits for the calculated difference, being of a total space complexity of $O(m)$.

One m-bit subtraction per value has a time complexity of $O(m)$ [24]. Since $n-1$ values need to be processed, this step is performed $n-1$ times, corresponding to $O(n)$ time complexity times the complexity per value, totaling a time complexity of $O(mn)$ for all values.

The third step, i.e., the variable length coding, can also be performed value by value. Since the worst-case encoding, i.e., the variable length encoding of the largest possible value, is proportional to the input value bit size m [20], the total space complexity of this step is $O(m)$.

Calculating a variable code word of an i-bit value requires a constant number of additions and subtractions as well as i divisions or shift operations [20], followed by a maximum of $2i+1$ bit writing operations. With the worst-case input bit length being proportional to m, this requires a time complexity of $O(m)$, since m-bit additions, subtractions and shift operations are all of time complexity $O(m)$ [24]. In total, this yields a time complexity of $O(mn)$ for all values.

Note that the fourth step, i.e., the bit string concatenation, is not listed in Tables II and III. This is due to the fact that the output values of the preceding step can be passed directly to the next step, i.e., the entropy coding step, one by one so that no intermediate memory is required and no actual concatenation has to be performed. The bit string concatenation can therefore be regarded as a conceptual step rather than an actually necessary one in a straight-forward implementation.

The fifth and final step, i.e., entropy coding, is performed on the whole output of the third step, value by value, one bit at a time. This conceals the intermediate concatenation step as explained above. Since the output of the third step has a total length which is $O(mn)$, the per-bit time complexity of the entropy coder times mn yields the time complexity of the complete entropy coding step.

In total, the time complexity of an adaptive binary arithmetic encoder for each input bit is constant, i.e., of time complexity $O(1)$ with respect to m and n [22]. Hence, processing $O(mn)$ input bits is of time complexity $O(mn)$. The space complexity is constant, i.e., $O(1)$ [22].

Summing up the space and time requirements of all described steps, this yields a total time complexity of $O(mn)$, which is proportional to the number of input values and their maximum size in bits and thus equal to, e.g., the time complexity of performing n m-bit additions, i.e., relatively low. The space complexity is $O(m)$ and therefore not dependent on n, which enables encoding with very modest space requirements.

Since, decoding involves the inverse operations of the described encoding steps in opposite order (as explained in Section IV-G), the time and space complexity of a decoder are expected to be equal to the encoder's. In order to avoid a complete complexity analysis of the decoding process, we refer to the symmetry of the operations involved to claim this without explicit proof.

B. Resumability

As load profile data is transmitted extensively in smart grids, the adequacy of our compression approach for this use-case has to be assessed. Although effective compression reduces data size and thus transmission time, it may have undesirable side effects compared to uncompressed transmission, for example when data is lost.

Although our approach does not include native error detection capabilities, it allows for retransmissions of parts of the data with very low overhead. This subsection discusses the conditions under which a lost part of the transmitted data can be retransmitted so that the transmission process can be resumed. Furthermore, the retransmission procedure as well as its overhead are discussed.

The state of the decoder during the decoding process is limited to a small number of variables. First, the differential coding step stores one m-bit variable containing the last coded value, as described in Section V-A. Second, the arithmetic coder stores three machine-word-sized integers representing the probability of the symbol zero, the current interval and the number of bits to be output after the next one, respectively.

Since this information is sufficient to represent the entire state of the decoder, a decoding process can be resumed by sending the aforementioned variables. For example, if $m = 32$ and the machine word size is 16, the decoder state can be represented by $32 + 3 \cdot 16 = 32 + 48 = 80$ bits, or 10 bytes.

Since our compression approach is not able to detect errors, it relies on an encapsulation format or transmission protocol to do so. In case of an error, the decoder's state can be reset to the last known good state by keeping a copy of the decoder's state after each successfully decoded data packet. The retransmission overhead is then equal to the size of the decoder state, e.g., 10 bytes.

Alternatively, it is possible for the decoder to request the encoder's state in order to resynchronize. Since both perform symmetric operations, their states have to be equal. Note that they can only exchange their states if the used protocol and channel allow them to do so.

This allows for resumability, i.e., the possibility to resume the decoding process at a given point. However, it is not possible to reconstruct preceding values which have been omitted between the last known good value and the resumed one (if there are any), since prior decoder states cannot be reconstructed. However, it is possible to deliberately omit parts of the data as long as the decoder state required to start decoding after the omitted parts is available.

If our approach is used in combination with packet-based transmission, we suggest to add the decoder state to each packet. This way, the decoder can process each packet independently and does not rely on the retransmission of preceding packets in order to decode the current one. The overhead should be negligible for typically large packet sizes, e.g., 1500 bytes for IEEE 802.3 (Ethernet), as well as for practical values of m and typical machine word sizes.

VI. Performance Evaluation

In order to evaluate the performance of our proposed approach, we analyze its compression efficiency as well as its processing time for a number of data sets and compare the results to those of related standards. Before discussing the results, we describe our implementation and test environment as well as the used data sets and the related standards used for comparison.

A. Related Standards

As mentioned above, we use two common uncompressed data formats for comparison with the proposed method. We only consider the encoding of the actual payload, i.e., we omit any encapsulation and/or protocol overhead since this would bias the results.

The first uncompressed format is described in IEC 61334-6 [15], which is also referred to as A-XDR encoding. Although this standard describes a number of possible encodings for numerical values, we consider the fixed-length unsigned integer encoding [15, Section VI-A1a] to be the most practically relevant one due its low per-value overhead.

As explained above, we omit all encapsulating identifier and length fields to make the comparison between the our encoding and the one from A-XDR as fair as possible. This way, n encoded values with a maximum of m bits size each require a constant amount of $n \cdot \lceil m/8 \rceil$ bytes or $8n \cdot \lceil m/8 \rceil$ bits, as illustrated in Fig. 3(a).

Since A-XDR provides no explicit encoding for floating point values, all input values to the A-XDR encoding process are considered to be integer values. This can be assured

Byte offset	0	1
Payload	00000000	1111011

(a)

Byte offset	0	1	2	3
Payload **ASCII rep.**	00110001 ’1’	00110010 ’2’	00110011 ’3’	00000000 ’\0’

(b)

Fig. 3. Encodings of the integer value 123 from related standards. (a) 16-bit fixed-length unsigned integer A-XDR encoding. (b) ASCII-based IEC 62056-21 encoding with trailing zero (delimiter).

through a normalization preprocessing step, as described in more detail in Section VI-B. As this preprocessing step is part of our proposed compression algorithm anyway, it can be performed once on the input data instead of once within each algorithm. This also allows for a fairer comparison of all evaluated encodings since they are provided with the same input data.

The second uncompressed format is described in IEC 62056-21 [13]. It encodes values in data blocks [13, Section VI-C4] which consist of one or more data lines [13, Section VI-E1], which themselves contain one actual value and its corresponding unit each. All data lines consist of a limited set of printable characters expressed in the ASCII character set [25].

Again, we only consider the actual payload, i.e., the ASCII-encoded values and omit the units and other protocol overhead. Note, however, that one additional byte per value is required to separate consecutive values due to their variable length, as illustrated in Fig. 3(b). This is not necessary for our approach or the A-XDR encoding described above since the latter uses a fixed number of bytes and our approach implicitly encodes length information (see Section IV-D).

B. Implementation and Test Environment

We implemented our approach proposed in Section IV in Python. For the arithmetic coding step, we used D. MacKay’s implementation[1] which is practically identical to the implementation from [22]. Since MacKay’s implementation operates on a bit string, the preceding step of our approach, i.e., the exponential-Golomb coding, outputs such a bit string instead of the corresponding byte sequence, as opposed to all prior steps.

In order to make the comparison between our approach and the ones from related standards fair, we reimplemented the latter so that they output bit strings as well. Note that, although this alters the processing time slightly, it still allows comparing the algorithms with respect to the order of magnitude of processing time.

As described in Section VI-A, some of the algorithms from related standards are not capable of handling noninteger (i.e., in this case, floating point) values. To simplify processing and make the processing time comparison fairer, all data is preprocessed by applying the normalization step described in Section IV-A. Thus, all algorithms operate on integer input data. This reduces the processing time of those algorithms in need of preprocessing, but still allows for a comparison with respect to the order of magnitude of processing time.

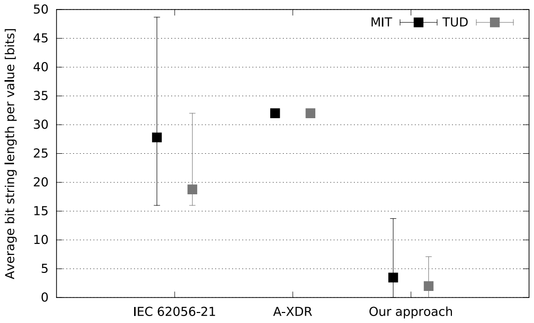


Fig. 4. Average value size in bits for the MIT (black) and the TUD data set (gray). The bars indicate the average number of bits per value for those load profiles which require the minimum and maximum number of bits in each data set, respectively.

Our test environment is a server hosting an Intel Xeon E5-2620 CPU with six physical cores running at 2 GHz each. The server runs Ubuntu 12.04.2 LTS on a 64-bit Linux 3.2.0-48 kernel. We use Python 2.7.3 and its built-in clock function from the time module to measure processing times.

C. Data Sets

For our evaluation, we use the MIT and TUD data sets described in Section III. As the values in the MIT data set are stored with a precision of 0.01 watts, we normalize them by multiplying them by 100 as described in Section IV-A. The values in the TUD data set are already stored as integers and therefore require no normalization.

For the fixed-length A-XDR encoding we used a bit length of 32 bits, since this is the smallest whole-byte size which is a power of two (which is typically used in practice) that allows covering the whole value range present in the input files. Similarly, our algorithm uses 32 bit variables for the difference calculation in the differential coding step (see Section IV-B).

D. Compression

For all load profiles of each data set, we evaluate the average number of bits required to represent one value. Fig. 4 illustrates this for all tested approaches. Obviously, fixed length A-XDR coding always requires 32 bits, while the IEC 62056-21 encoding and our approach requires a variable number of bits.

It is clear that our approach outperforms the other two by an order of magnitude. Furthermore, there is no load profile for which our approach is inferior to one of the other two, since the maximum number of bits per value required by our approach is always significantly smaller than the minimum amount of bits per value required by both, the fixed-length A-XDR and the IEC 62056-21 encoding.

As the value range of the TUD data set is smaller than the one of the MIT data set (see Section IV-B), the average number of bits required per value is significantly smaller for the TUD

[1] http://www.inference.phy.cam.ac.uk/mackay/python/compress/#AC

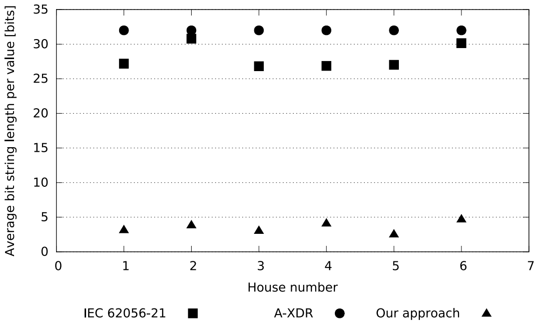


Fig. 5. Average value size in bits for each house from the MIT data set. The bars indicate the average number of bits per value for those load profiles which require the minimum and maximum number of bits in each house, respectively.

data set. This reflects in the results for both, our approach and the IEC 62056-21 encoding, which are variable-length codes.

Since the MIT data set is a collection of load profiles from multiple houses, it is possible to investigate the differences between them in terms of compression efficiency. Fig. 5 shows the average number of bits required per value and house for each encoding approach.

Again, our proposed compression approach outperforms the other two by an order of magnitude for each individual house, revealing that averaging the values per data set (as shown in Fig. 4) did not conceal any inefficiencies of our approach. Interestingly, the compression efficiency of our approach shows a slight correlation with the IEC 62056-21 encoding, albeit inherently not proportional.

As the load profiles of each house from the MIT data set correspond to one channel each, it is possible to investigate their individual compression performance. Fig. 6 shows the average bit length per value for each channel of each house. Note that the channel labels are taken directly from the MIT data set's companion files and have not been modified, i.e., corrected orthographically.

For almost all individual channels, i.e., load profiles, our compression approach is significantly superior to the other two. Although, it is not always more efficient by an order of magnitude, it is at least twice as efficient for all channels.

One advantage of our approach becomes clear for channels which are typically constant or quasi constant over long periods of time, e.g., the oven and stove channels in house 1 (top left). Since the load on such a channel typically changes rarely and rapidly, the number of zero and small differences (due to noise and measurement inaccuracies) is very high, allowing our approach to effectively compress the data.

However, channels with relatively unpredictable load behavior, like the mains of all houses, can still be compressed very efficiently as compared to the other two encodings. The same is true for channels with low variable load, like the kitchen outlets of all houses, showing that our proposed approach is able to compress these types of load profiles efficiently as well.

Although, our approach achieves better compression performance for channels with relatively small changes between values, the employed exponential-Golomb code and arithmetic coding still allow for sufficiently good compression in cases of larger changes between values. This is mainly due to the fact that the length of exponential-Golomb codes (in bits) only increases logarithmically with increasing input values, i.e., increasing value differences in our method. Thus, only very large value differences would generate notably longer codewords and thereby impact the compression performance significantly, which is a practical feature of our proposed approach.

E. Processing Time

Fig. 7 shows the average number of microseconds required to encode one value with each approach for the MIT (black) and the TUD data set (gray), respectively. As explained in Section VI-B, our implementation only allows assessing the order of magnitude of the processing times.

Thus, it cannot be asserted that our approach is significantly faster than the other two. However, it requires about the same order of magnitude in terms of processing time per encoded value, hence being comparable to both, fixed-length A-XDR and IEC 62056-21 encoding.

Since all variable-length coding approaches depend on the actual size of their input values, both, our approach and the IEC 62056-21 encoding, clearly require less time per value on average for the TUD data set than they do for the MIT data set. This is clear, since the latter's value range is larger. This confirms the linear dependency of both approaches to the input bit length, allowing for faster processing of small input values.

Note that further investigations, e.g., per house or per channel of the MIT data set, are not meaningful due to the limited measurement accuracy. This is supported by the difference between the average processing times of the MIT and the TUD data set for the A-XDR encoding, which should be zero in theory assuming a perfect implementation and execution environment, since A-XDR is a fixed-length code.

VII. Future Work

Our approach has been shown to perform very well in terms of compression efficiency, but there is still room for improvement. For example, signal characteristics like periodicity could be exploited as most load profiles tend to exhibit weekly, daily or even hourly patterns based on the types of appliances and users. By forming a prediction signal from the last week, day or hour, respectively, one could encode the difference between the current data values and the predicted ones instead of encoding consecutive value differences, thereby further increasing compression efficiency. It should also be possible to combine these two approaches.

Another issue to be addressed is the diversity of the evaluated load profiles. Although the amount and diversity of load profiles from the MIT and the TUD data set are already very high, it would be desirable to have a larger amount of real-world test data available in order to extend the results of this paper. This does not only include data from different sources,

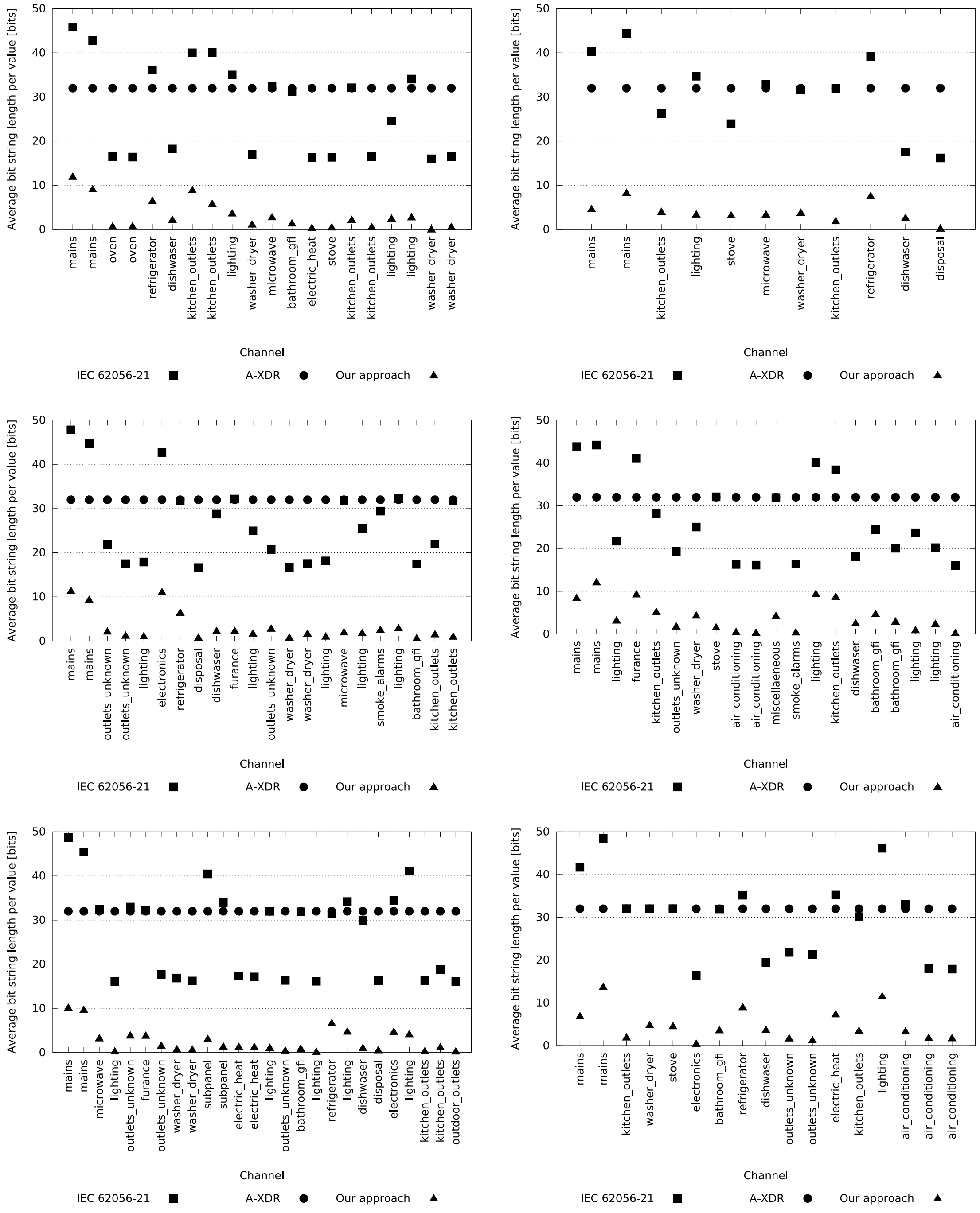


Fig. 6. Average value size in bits for each channel of each house from the MIT data set (top left: house 1; top right: house 2; etc.). Each channel corresponds to one load profile.

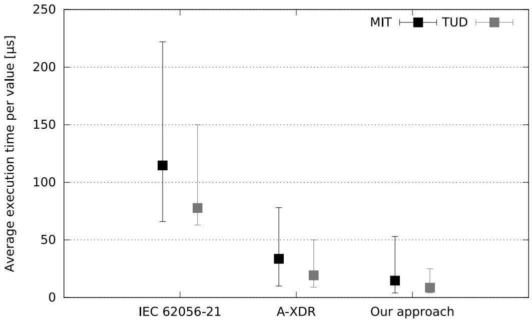


Fig. 7. Average encoding time per value for the MIT (black) and the TUD data set (gray). The bars indicate the average number of microseconds per value for those load profiles which require the minimum and maximum time in each data set, respectively.

but also different granularity, e.g., a sampling time of minutes or even hours. This way, the compression efficiency could be evaluated more thoroughly.

Finally, the overhead introduced by using our proposed resumability feature has to be evaluated thoroughly. This does not only include analyzing the overhead for different loss rates, but for different link types and protocols as well. Since such an analysis is beyond the scope of this paper, it remains future work.

VIII. Conclusion

We proposed a compression approach tailored for the requirements of load profile data transmission in smart metering. We showed that our approach allows for resumability with very low overhead, which enables it to operate in error-prone transmission lines in smart grids. Even with providing resumability, our approach has been shown to maintain the same compression results as standard compression algorithms, which do not provide this important feature. Currently employed state-of-the-art transmission encodings are outperformed by an order of magnitude in terms of compression performance without significantly impacting the processing time required for the encoding process. In summary, the proposed approach is ideally suited for compression of smart meter load data as it delivers competitive compression results with low computational complexity, low memory requirements, low overhead for initialization and the ability to resume after interruption.

References

[1] Z. Fan *et al.*, "Smart grid communications: Overview of research challenges, solutions, and standardization activities," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 21–38, Apr. 2013.

[2] M. Ghofrani, M. Hassanzadeh, M. Etezadi-Amoli, and M. Fadali, "Smart meter based short-term load forecasting for residential customers," in *Proc. North Amer. Power Symp. (NAPS)*, Boston, MA, USA, 2011, pp. 1–5.

[3] A. Wander, N. Gura, H. Eberle, V. Gupta, and S. Shantz, "Energy analysis of public-key cryptography for wireless sensor networks," in *Proc. 3rd IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, Kauai Island, HI, USA, 2005, pp. 324–328.

[4] P. McDaniel and S. McLaughlin, "Security and privacy challenges in the smart grid," *IEEE Security Privacy*, vol. 7, no. 3, pp. 75–77, May/Jun. 2009.

[5] M. Ringwelski, C. Renner, A. Reinhardt, A. Weigel, and V. Turau, "The hitchhiker's guide to choosing the compression algorithm for your smart meter data," in *Proc. 2012 IEEE Int. Energy Conf. Exhibit. (ENERGYCON)*, Florence, Italy, pp. 935–940.

[6] J. Kraus, P. Stepan, and L. Kukacka, "Optimal data compression techniques for smart grid and power quality trend data," in *Proc. 2012 IEEE 15th Int. Conf. Harmonics Qual. Power (ICHQP)*, Hong Kong, pp. 707–712.

[7] J. Ning, J. Wang, W. Gao, and C. Liu, "A wavelet-based data compression technique for smart grid," *IEEE Trans. Smart Grid*, vol. 2, no. 1, pp. 212–218, Mar. 2011.

[8] J. Khan, S. Bhuiyan, G. Murphy, and M. Arline, "Embedded zerotree wavelet based data compression for smart grid," in *Proc. 2013 IEEE Ind. Appl. Soc. Ann. Meeting*, Lake Buena Vista, FL, USA, pp. 1–8.

[9] N. Harada, Y. Kamamoto, T. Liebchen, T. Moriya, and Y. A. Reznik, "The MPEG-4 audio lossless coding (ALS) standard—Technology and applications," in *Proc. Audio Eng. Soc. Conv. 119*, New York, NY, USA, Oct 2005, Art. ID 6589.

[10] *Information Technology—Coding of Audio-Visual Objects—Part 10: Advanced Video Coding*, ISO/IEC Standard 14496-10:2009, 2005.

[11] *Open Smart Grid Protocol (OSGP)*, ETSI GS OSG 001 V1.1.1, 2012. [Online]. Available: http://www.etsi.org/deliver/etsi_gs/OSG/001_099/001/01.01.01_60/gs_osg001v010101p.pdf

[12] Smart Meters Coordination Group, "Functional reference architecture for communications in smart metering systems," CEN/CLC/ETSI/TR, Tech. Rep. 50572, Dec. 2011. [Online]. Available: ftp://ftp.cencenelec.eu/EN/EuropeanStandardization/HotTopics/SmartMeters/CEN-CLC-ETSI-TR50572%7B2011%7De.pdf

[13] *Electricity Metering—Data Exchange for Meter Reading, Tariff and Load Control—Part 21: Direct Local Data Exchange*, IEC Standard 62056-21, 2002.

[14] *Electricity Metering—Data Exchange for Meter Reading, Tariff and Load Control—Part 53: COSEM Application Layer*, IEC Standard 62056-53, 2002.

[15] *Distribution Automation Using Distribution Line Carrier Systems—Part 6: A-XDR Encoding Rule*, IEC Standard 61334-6, 2000.

[16] J. Kolter and M. Johnson, "REDD: A public data set for energy disaggregation research," in *Proc. Workshop Data Min. Appl. Sustain.*, San Diego, CA, USA, 2011, pp. 1–6.

[17] A. Reinhardt *et al.*, "On the accuracy of appliance identification based on distributed load metering data," in *Proc. 2nd IFIP Conf. Sustain. Internet ICT Sustain. (SustainIT)*, Pisa, Italy, 2012, pp. 1–9.

[18] J. Ganssle, *The Art of Designing Embedded Systems*. Amsterdam, The Netherlands: Elsevier Sci., 2008.

[19] C. C. Cutler, "Differential quantization of communication signals," U.S. Patent 2 605 361, July 29, 1950.

[20] D. Marpe, H. Schwarz, and T. Wiegand, "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 620–636, Jul. 2003.

[21] D. Engel, A. Uhl, and A. Unterweger, "Region of interest signalling for encrypted JPEG images," in *Proc. 1st ACM Workshop Inf. Hiding Multimedia Security (IHMMSEC)*, Montpellier, France, 2013, pp. 165–174.

[22] I. H. Witten, R. M. Neal, and J. G. Cleary, "Arithmetic coding for data compression," *Commun. ACM*, vol. 30, no. 6, pp. 520–540, Jun. 1987.

[23] W. B. Pennebaker, J. L. Mitchell, G. G. Langdon, Jr., and R. B. Arps, "An overview of the basic principles of the Q-Coder adaptive binary arithmetic coder," *IBM J. Res. Develop.*, vol. 32, no. 6, pp. 717–726, Nov. 1988.

[24] T. Cormen, C. Leiserson, R. Rivest, and C. Stein, *Introduction To Algorithms*, 2nd ed. Cambridge, MA, USA: MIT Press, 2001.

[25] *Coded Character Sets—7-Bit American National Standard Code for Information Interchange (7-Bit ASCII)*, ANSI Standard X3.4, 1986.

Andreas Unterweger (S'13) received the Diploma degree in information technology and systems management (with distinction) from the Salzburg University of Applied Sciences, Puch/Salzburg, Austria, in 2008, and the Master's degree in computer science (with distinction) from the University of Salzburg, Salzburg, Austria, in 2011, where he is currently pursuing the Ph.D. degree in selective video encryption.

He is an External Lecturer with the Salzburg University of Applied Sciences for the bachelor's degree program. His current research interests include real-time image and video compression, as well as selective video encryption.

Dominik Engel (S'06–M'08) received the Ph.D. degree in computer science from the University of Salzburg, Salzburg, Austria in 2008.

He was a Researcher with the University of Bremen, Bremen, Germany, and the University of Salzburg, and a Product Manager with Sony DADC, Anif, Austria, where he was responsible for video content security. Since 2010, he is a Professor with the Salzburg University of Applied Sciences, Puch/Salzburg, Austria, where he is the Head of the Josef Ressel Center for User-Centric Smart Grid Privacy, Security, and Control. His current research interests include smart grid security and privacy, multimedia security, and technological methods for enhancing end-user trust.

3.4 lausenhammer16a

- W. Lausenhammer, D. Engel, and R. Green. Utilizing capabilities of plug in electric vehicles with a new demand response optimization software framework: Okeanos. *International Journal of Electrical Power and Energy Systems*, 75:1–7, 2016.

Electrical Power and Energy Systems 75 (2016) 1–7

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Utilizing capabilities of plug in electric vehicles with a new demand response optimization software framework: Okeanos

Wolfgang Lausenhammer [a], Dominik Engel [a], Robert Green [b,*]

[a] Josef Ressel Center for User-Centric Smart Grid Privacy, Security and Control, Salzburg University of Applied Sciences, Salzburg, Austria
[b] Dept. of Computer Science, Bowling Green State University, Bowling Green, OH 43403, United States

ARTICLE INFO

Article history:
Received 29 September 2014
Received in revised form 14 August 2015
Accepted 21 August 2015

Keywords:
Demand response management
Multi-agent systems
Game theory
Plug in electric vehicles

ABSTRACT

Particularly with respect to coordinating power consumption and generation, demand response (DR) is a vital part of the future smart grid. Even though, there are some DR simulation platforms available, none makes use of game theory. This paper proposes Okeanos, a fundamental, game theoretic, Java-based, multi-agent software framework for DR simulation that allows an evaluation of real-world use cases. While initial use cases are based on game theoretic algorithms and focus on consumption devices only, further use cases evaluate the effects of plug in electric vehicles (PEVs). Results with consumers show that the number of involved households does not affect the costs per household. Further evaluation involving PEVs demonstrates that with an increasing penetration of PEVs and feed-in tariffs the costs per household per month decrease.

Introduction

Energy demand in the USA is expected to increase by at least 19%, the supply, in contrast, is only expected to rise by 6% [1]. Furthermore, this energy mismatch is not a US-specific problem [2,3]. While renewable energy could help relieve the load on the grid, it also poses a significant challenge to the grid in terms of keeping supply and demand in balance. With respect to coordination, demand response management (DRM) could pose an ideal solution to this problem [4,5]. DRM refers to “changes in electric usage by end-use customers from their normal consumption patterns in response to changes in the price of electricity over time, or to incentive payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardized” [6, 21].

Game theory, in its essence, aims to help understand situations in which several decision-makers interact. Being a mathematical framework and analytical tool, game theory helps study the relationships and actions among rational players. This characteristic renders it an ideal tool to model and understand the inherent complexity of demand response (DR) resulting from this interaction. Publications in this area range from load shifting approaches [7,8] to using storage devices such as PEVs in micro-grid storage games [9] to games that focus on utility companies [10,11]. One thing that these works have in common is a mathematical proof that by optimizing a utility function, a stable point called a Nash equilibrium will be reached [12,13].

This study proposes Okeanos, a novel, game theoretic, Java-based, multi-agent software framework for DR simulation that is capable of investigating the effect of optimizing multiple electric appliances using a game theoretic approach. It is fundamentally different from other DRM software approaches as it plans consumption and production ahead of time. By utilizing game theory, Okeanos benefits from mathematically sound solutions for finding the optimal schedule for household appliances. It supports the simulation of different types of loads and can be configured to work with different game theoretic DRM approaches. The current source has been released as open source[1] and can be used and extended to fit various needs.

While initial results show that savings of up to 6% can be achieved by changing the switch-on time of three household appliances, higher savings can be achieved either by adding more manageable devices to the simulation or by incorporating elective vehicles (EVs) of some sort. In this study, the focus will remain on plugin EVs (PEVs).

The remainder of this paper is structured as follows: The fundamental DR simulation platform, Okeanos, is introduced and key concepts are highlighted in Section “Okeanos”; Results of load

* Corresponding author.
E-mail addresses: wolfgang.lausenhammer@en-trust.at (W. Lausenhammer), dominik.engel@en-trust.at (D. Engel), greenr@bgsu.edu (R. Green).

[1] https://github.com/wolfgang-lausenhammer/Okeanos.

http://dx.doi.org/10.1016/j.ijepes.2015.08.014
0142-0615/© 2015 Elsevier Ltd. All rights reserved.

shifting are presented and described in Section ''Simulation of multiple households with load-shifting devices''; This is followed by simulations that incorporate PEVs in Section ''Evaluation of Okeanos with plug in electric vehicles''; and, finally, Section ''Conclusion'' concludes this work.

Okeanos

Okeanos is a novel DR simulation platform with a special focus on the inclusion of game theory. Unlike the software presented in [4,14,15], any coordination mechanism that complies with the defined interface is compatible with Okeanos.

Okeanos aims to be a holistic platform for DRM with support for a wide variety of appliances. Through the means of extensibility, new devices can be added by writing a driver for the specific appliance. With OSGi as the foundation, new features can be easily developed, deployed or replaced.

Independent smart household appliances

Similar to other approaches, Okeanos utilizes the multi-agent paradigm to represent household appliances. Thus, with a one-to-one matching between agents and household devices, every device can work towards and set goals or targets on its own. Appliances are proactive and make independent decisions according to the information available to them.

In order not to implement all multi-agent features from scratch, Okeanos builds on JIAC, a feature-rich, modularized and easy to use framework [16]. JIACs modern approach that uses the Spring framework as the basis for the whole system is unique throughout a comparison of multi-agent frameworks including JADE [17], Janus [18] and Jason [19]. Additional evaluation criteria included functionality, active development, ease of use and adoption throughout the software developer community.

In JIAC, the functionality of agents is defined by agent beans. Each bean is a small module with a well-defined responsibility, leading to improved reusability [16]. The energy consumption game described in Section ''Coordination mechanism in Okeanos'' is an ideal example for this. Its responsibility is to ensure the correct sequential execution of the algorithm. All agents taking part in the schedule optimization process use this bean. Due to the autonomy of agents, it is possible that agents use different games. The meaningfulness of such a mixture, however, is questionable, as no guarantee of the existence of a Nash equilibrium can be given under such circumstances.

The callback functions (cf. Fig. 2) allow for separation of concerns, as the agent itself is still responsible to forward requests to the corresponding components. Similarly, drivers and other services are agent beans as well, ready to be used by agents to support its goals.

Plug in support

OSGi and the Spring framework are two well-known Java frameworks that provide a solid foundation for Okeanos. While both are very powerful tools and offer many features for their respective fields, they share some key concepts, most notably loose coupling and separation of concerns. Naturally, it is beneficial to combine the two and have a module-based, service oriented system as the platform Okeanos runs on, using Spring for the wiring of the components. Eclipse Gemini Blueprint provides a clean and easy to use interface for integrating the two frameworks.

However, to be able to fully utilize benefits of loose coupling, thorough planning is required. Device drivers are the perfect example for the need for extensibility. A flexible and powerful interface eases the interaction with new implementations and the integration of new modules into the system. This is crucial to be able to keep the threshold for developing new modules as low as possible.

With Okeanos built on OSGi, it comprises a conglomerate of various bundles (see Fig. 1) rather than a monolithic core. To allow for optional bundles, the OSGi R5 specification [21] recommends separating interfaces from the implementation in a separate bundle. Consider, for example, a logging service: The application does not necessarily need an implementation for a correct execution, however, at least the interface needs to be present to allow for proper resolution.

As indicated in Fig. 1, every service in Okeanos could be represented in its own module. While, this is possible, it also implies an explosion of projects and, therefore, an increase in complexity. Therefore, layers serve as the boundaries for modules in Okeanos. As recommended, the interfaces of each layer are separated from the implementation and consolidated in different bundles.

Likewise, as it is possible to have no implementation in an OSGi container, it is possible to have multiple implementations present. This is especially true for device drivers, as they all implement the same interface. To be able to distinguish between drivers, additional properties, such as year and brand of a household device, can be specified.

Fig. 1 shows the logic separation between the supporting libraries in the infrastructure bundles area and the application bundles that provide the actual functionality. The Spring extender bundle that is part of the Eclipse Gemini Blueprint project is responsible for activating all Spring powered application bundles and starting up their Spring contexts. This is similar to a J2EE environment, where the Spring application context is started by the application server, whereas here, the extender bundle is responsible for starting all application contexts.

Every such bundle has its own independent context that can import and export services by using special tags[2] in the `context.xml` file. The exported services are regular Spring beans that are registered in the OSGi service registry and, thus, made available to other contexts. For imported services, respectively, Gemini Blueprint searches for a suitable match in the OSGi service registry, fetches it and makes it available to the context.

Coordination mechanism in Okeanos

The requirement of game theory that players have to act rationally is ensured by representing every player by its own agent. Players in this context are household appliances as described earlier.

While there are a number of published game theoretic approaches to DR management [7–11], the game proposed by Mohsenian-Rad and co-authors [7] was modeled with Okeanos as a first proof of concept. Reasons for this include that the algorithm was formulated in pseudo code, which allows for accurate adaptation. Further, potentially more devices can be integrated in the first place by utilizing load shifting as if it were possible with storage devices due to the lack of available data.

The decentralized objective function in [7] with $x_{n,a}^h$ as the energy consumption of a scheduled appliance a of user n at hour h is given by

$$\begin{aligned} \underset{x_n \in \mathscr{X}_n}{\text{minimize}} \quad & \sum_{h=1}^{H} C_h\left(\sum_{a \in \mathscr{A}_n} x_{n,a}^h + \sum_{m \in \mathscr{N} \setminus \{n\}} l_m^h\right) \\ \text{subject to} \quad & l_m^h = \sum_{a \in \mathscr{A}_n} x_{m,a}^h \qquad h \in \mathscr{H} \end{aligned} \tag{1}$$

[2] For detailed instructions on the exact syntax see [22].

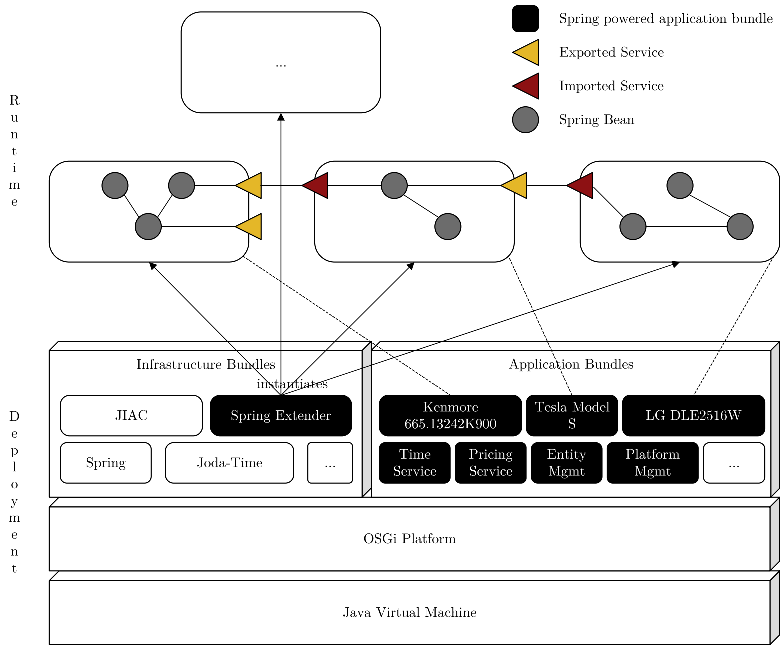


Fig. 1. Okeanos bundle structure with sample household devices and services. The indicated dependencies are only for illustrative purposes. Drivers do not depend on each other in Okeanos. Adapted from [20].

with a set of cost functions C_h that are increasing and strictly convex [7].

Every appliance only needs to optimize its own schedule $x^h_{n,a}$, because the consumption of all other players $l^h_m,\ m \in \mathcal{N} \setminus \{n\}$ is static. For debugging reasons and the sake of comprehensibility, Okeanos uses particle swarm optimization (PSO).

As denoted in Fig. 2, the algorithm proposed in [7] is started by the agent every time a new schedule is needed. Okeanos adopts the suggested 24 h planning horizon, which requires the agent to initiate it once a day.

The next step is to minimize the costs, i.e., solve the objective function (1). To be able to do that, the necessary information needs to be obtained first. The game has no knowledge, which device is used, therefore, it asks the agent. It knows about the configuration, obtains the information from the driver and returns it. Because the agent is the broker, it could also decide to alter this information. That is, stricter time frames could be set or it could remove itself completely from the schedule.

With that information in its memory, a configuration object of the local device and the most current information of all other devices is assembled. Subsequently, the agent is asked to optimize the configuration. Again, due to re-usability only the agent has knowledge about which optimization algorithm, game and drivers are used. The agent forwards the request to the optimization algorithm, e.g., PSO, which then returns the optimized schedule to the agent and, finally, to the energy consumption game.

The agent is then asked to approve the schedule before continuing. At that point, the algorithm proceeds by checking whether the optimized schedule has changed since the last announcement. If so, it broadcasts the new schedule to other agents. If not, Okeanos sets a timeout after which the agent assumes that no new schedules will be announced anymore.

If a new schedule is received within this time, the timeout is reset and the process starts again, as denoted by the loop-box in Fig. 2. This is repeated until no new schedules are received anymore and the timeout finally expires.

Once the timeout has expired, the equilibrium is reached and the agent informed about it by calling a callback function with the final schedule for the local device and the sum of the final schedules of all other devices.

Optimization algorithm

The optimization algorithm tries to find solutions to (1). Currently, only two different implementations of PSO are available. PSO belongs to the category of swarm algorithms and is loosely inspired by bird flocks or fish schools as first presented by Eberhart and Kennedy in 1995 [23].

The two implementations only differ in the solution space: The first implementation *PSORegulableLoadOptimizer* covers load shifting, while the second, *PSORegenerativeLoadOptimizer*, also handles charging and discharging of PEVs. Therefore, for load shifting, the velocity, as it represents a change relative to the current position, is represented as a vector of time differences in 15 min steps. This resolution tries to strike a balance between an optimal solution, which requires a higher resolution and a good solution, which can be calculated considerably faster. The position comprises the start times a device runs. That is, the position of a washing machine that has to run twice a day would be represented by a vector that comprises two values: the start time of the first run and the start time of the second run.

For charging and discharging of PEVs, Okeanos takes a different approach using *PSORegenerativeLoadOptimizer*. Not the start time is relevant here, but the amount of power charged or discharged at every 15 min interval is important. This is also exactly what a particle's position comprises. The velocity is a vector containing the change of charge for every interval. Optimizing regenerative loads like PEVs is more challenging, as a maximum capacity, minimum

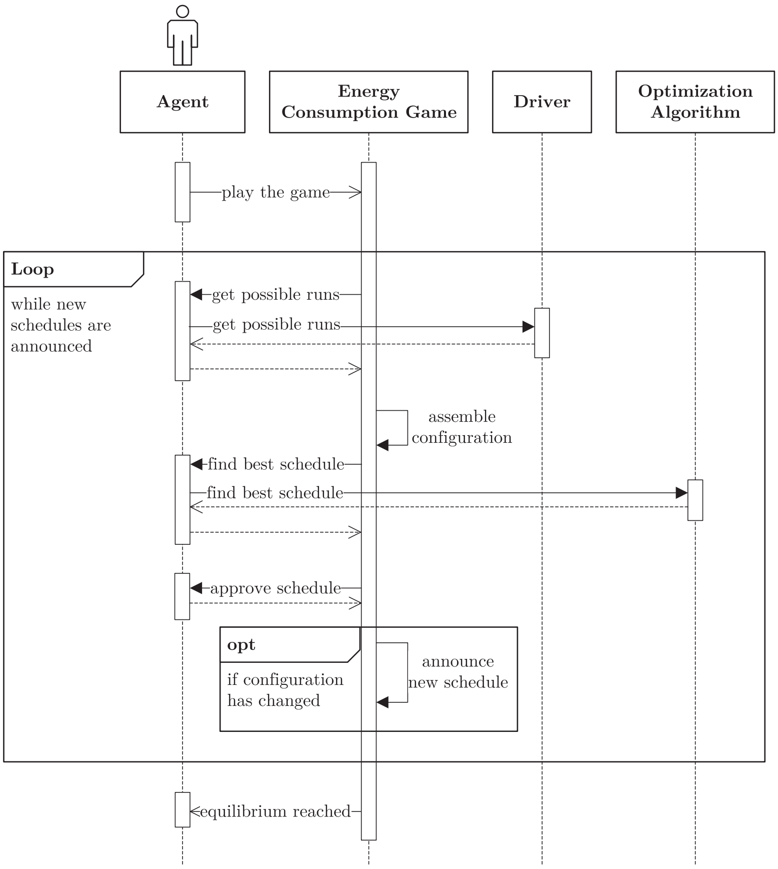


Fig. 2. Process flow of energy consumption game.

capacity and maximum charge per slot need to be taken into account too.

Simulation of multiple households with load-shifting devices

In a recent publication [24] Pipattanasomporn et al. collected the load profiles of selected major household appliances like dishwashers, AC units, refrigerators, washing machines and dryers. The data is available either in one second intervals or in one minute intervals that average the consumption over 60 s periods. Hence, due to the quality of the data, devices from this survey are modeled in Okeanos.

Here, the initial results presented in [25] are extended. These results show that by optimizing three household appliances of one household, Okeanos can save up to 5.9% of energy costs per month. The next logical step is to increase the number of households involved. That is, this section studies the impact of a rising number of households on the costs per household per month.

Not every household is alike, therefore, the load profile for every household is randomly scaled to either 25, 28, 30, 33 or 35 kWh per day. Additionally, it is randomly shifted between 1 h of its regular time. Finally, dishwashers, washing machines and clothes dryers run with a 33% chance. This configuration is chosen to account for different habits and usage patterns of customers.

Table 1
Comparison of costs per household per month with an increasing number of households.

	10 Households	20 Households	50 Households
Costs per month per household	$88.57	$86.84	$90.23

As illustrated in Table 1 and Fig. 3, altering the number of households does not change the outcome. It, however, can be seen that the peaks are getting more extreme the more households are involved.

At least two explanations should be considered when interpreting this results. On the one hand, there are too few devices that can be shifted. Also, because the load profiles of households have a minimum at the point in time when energy is cheapest, devices hardly have any other choice but to be switched on at that time. Further, because the average consumption of households is mostly the same, the energy consumption keeps stacking up and, as aforementioned, load shifting devices cannot smoothen the peaks.

On the other hand, the convex cost function at every point in time could need its parameters readjusted. This, however, is not very likely, as the devices that respond to costs, already run at the cheapest points in time. Households, however, do not react to different costs, which explains the peaks and the stacking of load profiles.

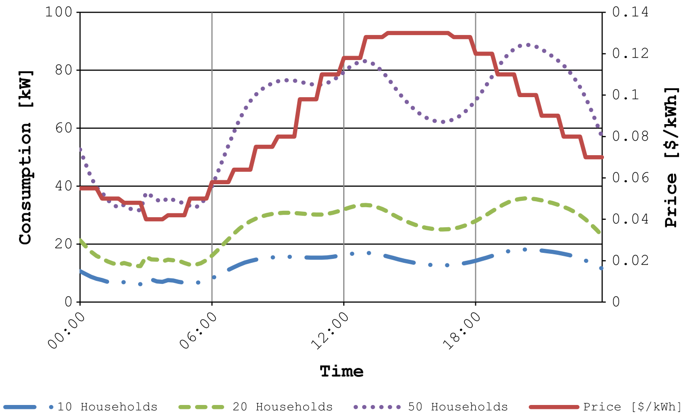


Fig. 3. Impact of the number of participating households on load profile.

According to Table 1, the costs per household per month do not show a significant difference when the number of participating households is increased. The reason for this is the same as described before: The load profiles are stacked.

Evaluation of Okeanos with plug in electric vehicles

This use case investigates the impact of integrating PEVs in the previous use case. As electric vehicles are all about storing energy, this is an extension to the implemented game theoretic algorithm [7], which proposes an energy consumption scheduling game. The original game was never designed for storage. The micro-storage management game proposed by [9] is contrary to that, it only proposes storage devices and does not do any load shifting. The proposed combination of both games is based on simulation only and there is no mathematical proof given unlike the individual games. Further, due to the use of PSO and the fact that it is a meta-heuristic, an optimal solution cannot be guaranteed. It should also be noted that all PEVs begin with an initial state-of-charge of zero.

Impact of penetration of plug in electric vehicles on costs per household

The first use case in the category of PEVs evaluates the impact of different penetrations of PEV on the total consumption. This simulation is based on 20 households, with either 0%, 25%, 50%, 75% or 100% of them owning one PEV. Owning really means having it stand around and not actively use it for transportation as for what it is made. In this configuration it acts like a rechargeable battery.

Furthermore, it uses a feed-in tariff of 50%. This means that if any device sells back energy to the grid, it will get 50% of the money it would cost the device to buy the same amount of energy. Additionally, as in the previous section, load shifting devices are switched on with a 33% chance.

As Fig. 4 shows, if only five of the 20 households, i.e., 25%, have a PEV, they completely change the load profile of households, overriding it with their own consumption pattern. This pattern, ultimately, is derived from the price function. As can be seen, PEVs charge themselves at the beginning of the day where the price for energy is cheap and use this energy later in the day to prevent the household from having to pay the peak price.

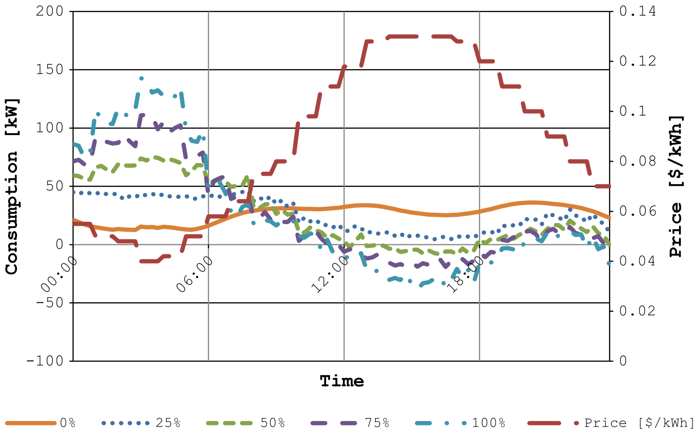


Fig. 4. Impact of penetration of PEVs on load profile. % PEV: variable, feed-in tariff: 50%.

An interesting phenomenon can be noticed at the end of the day at around 11 p.m. At this time devices start to discharge their remaining energy. This is due to the limited planing horizon, which is currently 24 h. Because devices cannot plan more than that, they want to sell the remaining energy to get the most out of the day.

The change of the load profile can be either wanted or unwanted. Even with a 25% penetration of PEVs, the peak consumption is nearly at 40 kW, compared to roughly 30 kW if there are no PEVs present. For higher penetrations, there is an even higher peak at the low-cost periods. This could be another unwanted peak as the grid needs to be prepared for that. If the grid is capable of transporting that amount of energy, this could be valuable to the utility company, because it sells cheap energy to customers and gets expensive energy for a cheap price, e.g., with a 50% feed-in tariff, which can be sold to other utility companies. Customers, despite the low feed-in tariff, still profit from selling energy back.

If the grid is not capable of handling that amount of energy, a possible countermeasure would be to adjust the cost function. The base price could either be changed or the factor, the costs per kWh at a point in time rise, could be adjusted as well. The latter countermeasure potentially has higher prospects of success, as it particularly penalizes high uses of energy, which, eventually, leads to a flatter load profile.

Table 2 compares the average costs per month for a household for a different penetration of PEVs with a 50% feed-in tariff. Most notably, the more households use PEVs the cheaper the average price for all households. Finally, when all households own a PEV and do not use it for anything else beside from participating in load scheduling, households can cut down electricity costs to approximately one fourth compared to not using PEVs at all.

This, however, is very unlikely to happen outside of simulation, as the simulation does not take a wide range of factors into account. Especially, (i) households own PEVs to use them and not let them stand in the garage at the charging station and (ii) the wear of batteries, etc. is not taken into account.

Table 2
Comparison of costs per household per month with an increasing number of households owning PEVs with a 50% feed-in tariff.

	Penetration of PEVs				
	0%	25%	50%	75%	100%
Costs per month per household	$88.20	$65.98	$52.27	$40.36	$27.50

The simulation, though, respects the maximum capacity, the minimum capacity, the maximum charge at a time and is also capable of “unplugging” a PEV, which means that the vehicle is currently in use and cannot be used for load scheduling. Furthermore, if a PEV is used, it also loses some charge, which can be expressed by the software as well.

Cross comparison of impact of feed-in tariff and penetration of plug in electric vehicles on costs per household

This use case is based on the previous use case, however, greatly expands the changed parameters. A parameter study of the feed-in tariff and the penetration with PEVs is done, unlike the previous use case that assumed a fixed feed-in tariff of 50%.

Fig. 5 illustrates the load profile when changing the feed-in tariffs. It clearly shows that the higher the incentive, i.e., the higher the feed-in tariff, the higher the likelihood that PEVs will charge during low-cost periods and discharge at high cost periods. Again, this is very similar to previous findings and is the result of trying to minimize the occurring costs for each device.

More interesting, however, is Table 3 and Fig. 6, which illustrates, respectively gives the exact numbers of the costs per household per month depending on the feed-in tariffs and the penetration with PEVs.

As previously pointed out, the costs per household per month decrease the more incentive is given (a higher feed-in tariff) or the more PEVs are available in the simulation. This effect results in households earning money at the end of the month when there are both, a high incentive and a high number of PEVs available.

The reason that the costs decrease with an increasing number of PEVs even with a 0% feed-in tariff is that the PEVs in that case are not actually selling the energy back to the grid, but provide it to other devices. Obviously, in total, this leads to a lower price, as PEVs provide energy during the high-cost periods.

Table 3
Comparison of costs per household per month with different feed-in tariffs and a different penetration of PEVs.

	Feed-in tariff				
	0%	25%	50%	75%	100%
Households with PEVs					
0%	$87.01	$86.41	$88.20	$86.78	$87.24
25%	$69.25	$68.70	$65.98	$63.77	$64.37
50%	$64.21	$61.55	$52.27	$38.81	$35.36
75%	$66.65	$58.73	$40.36	$17.47	$9.05
100%	$67.52	$56.01	$27.50	−$9.50	−$14.98

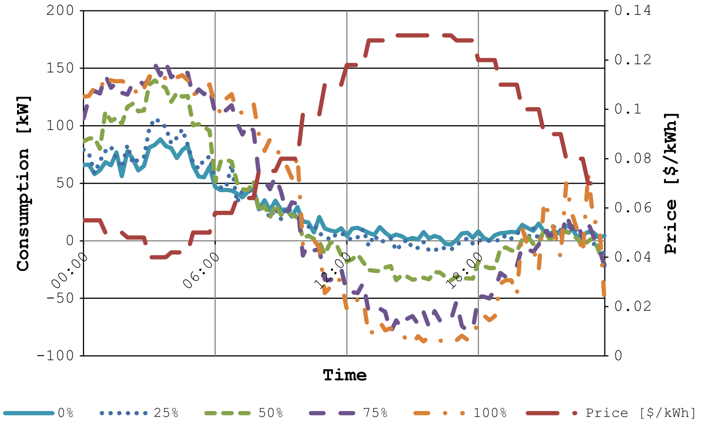


Fig. 5. Impact of feed-in tariffs on load profile. % PEV: 100%, feed-in tariff: variable.

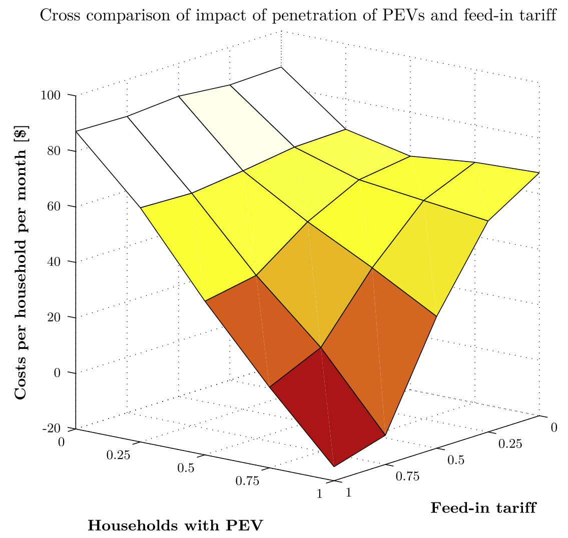


Fig. 6. Cross comparison on the impact of different feed-in tariffs and penetration of PEVs on the costs per household per month.

However, earning money through the use of PEVs seems unlikely as [9] simulated the impact of storage devices as well, with the result that in the UK 38% is ideal number of households owning a 4 kWh storage device, when the savings of up to 13% is at its maximum. These savings do not result in the households earning money at the end of the month. What can be done to make it more realistic is to adjust the aforementioned factor by which the costs per kWh rises.

Further, it can be noted that increasing the feed-in tariff from 75% to 100% has a significantly smaller impact than increasing it from 50% to 75%. One reason could be that the PEVs already use their whole available capacity when the 75% feed-in tariff is offered. Similarly, increasing the percentage of PEVs from 75% to 100% does only have a big impact with high feed-in tariffs.

There does not seem to be a particular parameter combination that is ideal for every case. The decision on the feed-in tariff has to be made by the utility company for every specific situation. Obviously, the number of PEVs in a grid need to be taken into account for that decision.

Conclusion

In this paper, Okeanos, a novel multi-agent demand response simulation platform focusing on the evaluation of game theoretic approaches was described. A major characteristic is its extensibility, which allows to support numerous household devices and enables the simulation of various games.

Simulation with three household devices shows that the costs per household are unaffected by number of involved households. Results involving PEVs demonstrate a decrease in the monthly utility bills per household with an increasing penetration of PEVs and feed-in tariffs. Future work will study the impact of using PEVs for commuting on the costs per household over longer periods (e.g. months, years, etc.).

Acknowledgments

The financial support by the Austrian Federal Ministry of Economy, Family and Youth and the Austrian National Foundation for Research, Technology and Development is gratefully acknowledged. Further, thanks belong to the Austrian Marshall Plan Foundation and The Chamber of Labour which have partly funded the research on this paper.

References

[1] Gudi N, Wang L, Devabhaktuni V, Depuru SSSR. Demand response simulation implementing heuristic optimization for home energy management. In: 2010 North American power symposium (NAPS 2010). Arlington (TX): IEEE; 2010. p. 1–6. http://dx.doi.org/10.1109/NAPS.2010.5619967.
[2] Han J-q, Piette MA. Solutions for summer electric power shortages: demand response and its application in air conditioning and refrigerating systems. Refrig Air Cond Electric Power Mach 2008;29(1):1–4.
[3] Palensky P, Dietrich D. Demand side management: demand response, intelligent energy systems, and smart loads. IEEE Trans Ind Inform 2011;7 (3):381–8. http://dx.doi.org/10.1109/TII.2011.2158841.
[4] Kok K. The PowerMatcher: smart coordination for the smart electricity grid. PhD thesis. Free University Amsterdam, Amsterdam, Netherlands; 7 2013.
[5] Fan Z, Kulkarni P, Gormus S, Efthymiou C, Kalogridis G, Sooriyabandara M, et al. Smart grid communications: overview of research challenges, solutions, and standardization activities. IEEE Commun Surv Tutor 2012;15(1):21–38. http://dx.doi.org/10.1109/SURV.2011.122211.00021.
[6] Federal Energy Regulatory Commission. Assessment of demand response and advanced metering; 12 2012.
[7] Mohsenian-Rad A, Wong VW, Jatskevich J, Schober R, Leon-Garcia A. Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid. IEEE Trans Smart Grid 2010;1(3):320–31. http://dx.doi.org/10.1109/TSG.2010.2089069.
[8] Ibars C, Navarro M, Giupponi L. Distributed demand management in smart grid with a congestion game. In: 2010 1st IEEE international conference on smart grid communication (SmartGridComm 2010). Gaithersburg (MD): IEEE; 2010. p. 495–500. http://dx.doi.org/10.1109/SMARTGRID.2010.5622091.
[9] Vytelingum P, Voice TD, Ramchurn SD, Rogers A, Jennings NR. Agent-based micro-storage management for the smart grid. In: Proceedings of the 9th international conference on autonomous agents and multiagent systems (AAMAS 2010). Toronto (Canada): ACM; 2010. p. 39–46.
[10] Maharjan S, Zhu Q, Zhang Y, Gjessing S, Başar T. Dependable demand response management in the smart grid: a stackelberg game approach. IEEE Trans Smart Grid 2013;4(1):120–32. http://dx.doi.org/10.1109/TSG.2012.2223766.
[11] Bu S, Yu FR, Liu PX. A game-theoretical decision-making scheme for electricity retailers in the smart grid with demand-side management. In: 2011 IEEE international conference on smart grid communication (SmartGridComm 2011). Brussels (Belgium): IEEE; 2011. p. 387–91. http://dx.doi.org/10.1109/SmartGridComm.2011.6102353.
[12] Rieck C. Spieltheorie: Eine Einführung. 11th ed. Christian Rieck Verlag; 2012 [in German].
[13] Saad W, Han Z, Poor HV, Başar T. Game-theoretic methods for the smart grid: an overview of microgrid systems, demand-side management, and smart grid communications. IEEE Signal Process Mag 2012;29(5):86–105. http://dx.doi.org/10.1109/MSP.2012.2186410.
[14] Struß O. Open-Source-Modellierung und auktionsorientierte Regulierung dezentraler Energienetze. Diploma thesis. University of Bremen, Department of Computer Science and Mathematics, Bremen, Germany, in German; 1 2012.
[15] Lehnhoff S, Krause O, Rehtanz C, Wedde H. Dezentrales autonomes Energiemanagement – Für einen zulässigen betrieb innerhalb verfügbarer kapazitätsgrenzen (distributed autonomous power management – for a reliable operation under feasibility constraints). Automatisierungstechnik 2011;59(3):167–79. http://dx.doi.org/10.1524/auto.2011.0906.
[16] JIAC Development Team. JIAC – Java Intelligent Agent Componentware, DAI-Labor, manual. Version 5.1.3; 10 2012.
[17] Bellifemine FL, Caire G, Greenwood D. Developing multi-agent systems with JADE. John Wiley & Sons; 2007. http://dx.doi.org/10.1002/9780470058411.
[18] Galland S, Gaud N, Rodriguez S, Hilaire V. Janus: another yet general-purpose multiagent platform. In: Proceedings of the 7th agent-oriented software engineering technical forum (TFGAOSE 2010), Agent Technical Fora, Paris, France; 2010.
[19] Bordini RH, Hübner JF, Wooldridge M. Programming multi-agent systems in AgentSpeak using Jason. John Wiley & Sons; 2007. http://dx.doi.org/10.1002/9780470061848.
[20] Spring Developers. Spring framework reference documentation, 4.0.6. RELEASE; 7 2014.
[21] The OSGi Alliance. OSGi core release 5; 3 2012.
[22] The OSGi Alliance. OSGi compendium; 5 2013.
[23] Kennedy J, Eberhart R. Particle swarm optimization. 1995 IEEE international conference on neural networks (ICNN 1995), vol. 4. Perth (Australia): IEEE; 1995. p. 1942–8. http://dx.doi.org/10.1109/ICNN.1995.488968.
[24] Pipattanasomporn M, Kuzlu M, Rahman S, Teklu Y. Load profiles of selected major household appliances and their demand response opportunities. IEEE Trans Smart Grid 2014;5(2):742–50. http://dx.doi.org/10.1109/TSG.2013.2268664.
[25] Lausenhammer W, Engel D, Green R. A game theoretic software framework for optimizing demand response. In: 6th European innovative smart grid technologies (ISGT 2015). Washington (DC): IEEE; 2015. submitted to IEEE ISGT 2015.

3.5 lueckenga16a

- J. Lückenga, D. Engel, and R. Green. Weighted vote algorithm combination technique for anomaly based smart grid intrusion detection systems. In *Proceedings of International Joint Conference on Neural Networks (IJCNN) 2016*, pages 2738–2742, Vancouver, Canada, July 2016.

Weighted Vote Algorithm Combination Technique for Anomaly Based Smart Grid Intrusion Detection Systems

Joris Lueckenga and Dominik Engel
Josef Ressel Center for User-Centric
Smart Grid Privacy, Security and Control
Salzburg University of Applied Sciences
Salzburg, Austria
Email: joris.lueckenga@en-trust.at
Email: dominik.engel@en-trust.at

Robert Green
Department of Computer Science
Bowling Green State University
Bowling Green, OH 43403
Email: greenr@bgsu.edu

***Abstract*—Intrusion Detection systems (IDS) are a crucial and necessary aspect of the smart grid, particularly when considering the possible attack vectors and their consequences. While there are many different approaches on IDS for Smart Grid, the benefits of an anomaly detection technique is still in discussion, due to its capability of detecting zero-day attacks and misuse. This paper proposes a weighted vote classification approach and a general weight calculation function to improve the detection performances of anomaly IDS systems. Initial results show that a combination technique is able to improve classifier performance by several percent.**

I. Introduction

Since the beginning of electrical power distribution, many ways have been found to compromise the metering system in order to gain financial benefits. A collection of attacks on the grid components and other issues regarding power grid security are given in [1]. These examples show that even with digital technology and securely acting implementations, attackers can and will try to manipulate the power grid infrastructure. One specific Smart Grid scenario, in which an attacker tries to lower the billing rates is discussed in [2]. This scenario describes the procedure of packet manipulation to inject false data into the Smart Grid network. As a result, billing rates are compromised and other clients have to pay more. Another problem is the intention to weaken a company or country by invading communication networks to limit or stop critical services. Recent events show that hacking activity is used for espionage and can be used as an instrument to threaten or damage countries. Viruses or attacks, like Stuxnet or countless examples of exploits available in the Internet are menacing threats for the Smart Grid security. In a survey of the anti-virus company Symantec, attacks on the energy sector were gathered and presented. It states that in the second half of 2012, the energy sector was the second most targeted asset, and attackers tended to go after valuable information. In addition to that, the sector is also a major target for sabotage attacks [3]. With the possibility of controlling electrical devices and cutting power supply, the Smart Grid functionality must be ensured and secured. Even though anomaly detection does have downsides (e.g.false positives, etc.), the flexibility of its implementation and the possibility to detect zero day attacks encourage the use of this technique. As it has previously been shown that the use of ensemble techniques may outperform single classifiers [13], this work focuses on extending the current research in the area of smart grid IDSs by proposing a novel, ensemble classification methodology using weighted voting to identify potential threats.

II. Literature Review

The area of intrusion detection, both in general and in the area of smart grid, is a well researched area [4], [5]. Some concepts have the potential to be adapted and integrated into a Smart Grid security concept including improvements in the Mobile WAN Connection, Rule-based IDS, Domain Knowledge for IDS, general Smart Grid Intrusion Detection systems, and ensemble classification.

A. Mobile WAN Connection Concept

In the work of [6], the Smart Meter is the core element of all functionality. Due to the mobile Internet connection, which is used to communicate with the power provider, a network infrastructure besides the Home Area Network (HAN) is not required. Also, the Smart Meter is considered a multi-functional device which contains a firewall and provides Power Line Connection, Zig-Bee or WLAN for HAN communication. To ensure the security and integrity of the connection, a tamper-resistant cryptoprocessor is used. The device will also have cryptographic algorithms and functions to support a public-private key infrastructure. The concept also suggests that, apart from the sender and receiver part of the transmitted data, packets should be fully encrypted. Using encryption would make this system safe, if the methods are applied and implemented correctly. It also would make IDS rather difficult, because the package content could not be read and checked against attacks. Another issue, which Anderson states in [7],

978-1-5090-0620-5/16/$31.00 ©2016 IEEE

is the problem of key compromising. If the private key of a provider is acquired, an attacker could remotely control a high number of meters and would be able to cause large electricity blackouts. Nevertheless, the concept provides good security measures and does not require the construction of a NAN and WAN infrastructure.

B. Rule-Based IDS

In the work of [8], a Behavior-Rule-Based Intrusion Detection System is suggested. The argumentation is that anomaly detection methods cannot avoid false classifications and still have too many false positives in the current research. Therefore it is argued, that they are not suitable for Smart Grid implementations. Instead, the work suggests a derivation of a specification-based IDS technique. In order to specify constraints for the possible actions of network nodes, a table of behavior rules is created. A rule defines valid behavior, for example the increase of billing rates on high demand. Any derivations from this rule would be considered as an attack. Every node has a monitor and a trustee to be able to control the behavior of each other. This approach can be very effective, since known as well as unknown attacks would trigger invalid behavior, like lowering the price, even if the demand is high. The downside of this is similar to the specification based concepts. If complex implementations already exist or if there is not enough effort done to be able to implement this rule based concept, the creation of this type of IDS might become too complex

C. Domain Knowledge for IDS

One method of enhancing anomaly-based intrusion detection is presented in the work of [8]. This concept suggests using Fuzzy Logic systems in order to represent human domain knowledge. To achieve this, several rules have to be defined. One given example is the increase of protocols during an attack. This can be the case if a system only uses a small variety of protocols and it is rather unlikely that a large protocol variety occurs. This knowledge can be mapped to a fuzzy rule, which controls the sensitivity of an anomaly detection algorithm. In case the amount of protocols increases, the detection sensitivity is high and packets are more likely to be considered as an attack. This triggers more findings. It will also increase the likelihood of false positives, but since an attack might occur, it would be useful to find every malicious packet. The other way round would happen when the number of protocols is low. In this case, less detections are triggered. This concept could be used very efficiently, when good human domain knowledge rules can be found for a system. It also could be combined with an already existing implementation, in order to enhance detection accuracy.

D. Smart Grid Intrusion Detection System

One promising and popular concept for detecting anomalous traffic in the Smart Grid is the Smart Grid Distributed Intrusion Detection System (SGDIDS) [9]. This concept uses interconnected, distributed IDS nodes in a network infrastructure. The concept was tested with a modified KDD-NSL dataset and three different anomaly detection methods were used. Also, clustering algorithms CLONALG and AIRS2Parallel were used in order test unsupervised classification efficiency. The basic principle is to have IDS systems in every network layer, which are able to communicate between the IDS nodes in the hierarchy. In case a node in the lower level of the hierarchy cannot classify the data, the packet is passed to the next instance, which has more powerful detection algorithms. Since this approach uses a popular infrastructure as well as a widely usable and retrofittable anomaly detection, this work aims to further improve the concept by providing a new approach for anomaly detection algorithms, which are used in the IDS nodes.

E. Ensemble Classification Concept

Ensemble classification scenarios already exist in different implementations [10], [13]. Ensemble techniques, as they are used in [11], [12], were successfully implemented in many other pattern recognition tasks. A common technique is to combine differently trained models in a voting system in order to improve the classification results. Usually, a large number of weak classifiers are used to achieve this. Different approaches exist to realize the combination, as stated in [13]. In general, two common methods show that models can be combined either with a majority vote or with a weighted technique. When a majority vote is used, the most votes collected for a single decision wins. This usually requires an uneven amount of votes or a decision rule, in case the votes are even. Another method is to use a weighted vote, which implies that for every classifier model, a certain weight is assigned. Each classifier votes with its weight and a decision will be made based on the highest assigned weight. The weight calculation can be done with different methods. In this approach, the weight is calculated based on the performance of a test set. Also, to achieve the best results, different weights will be used when a classifier either votes for an attack or against it. The difference in common methods is the use of a diverse set of algorithms in the combination, instead of differently trained models of the same algorithm. A similar concept has been used in [14]. To carry out this voting and weight checking, each classifier is first trained and the performance is determined by classifying a test set.

The combination is carried out by integrating true negative (TN) and true positive (TP) rates in a function and assign the calculated weight. In this case, it is intended that weights for normal and attack decisions differ. Each of the model weights are summed up and compared afterwards, in order to carry out the final decision. To assign and calculate weights for each model, a formula was developed to balance the given amount in an efficient way. Due to the problem that several classifiers might work better than others, it is crucial to assign more or less weight, based on the performance. To do this, the following function $f(x)$ was developed:

$$f(x) = \frac{1}{(1-x)*a+b} \tag{1}$$

The variables a and b are used for various adjustments. The variable x stands for the precision of TP or TN values. Those values are between 0 and 1 and represent the probability of a correct prediction of either normal traffic or an attack. The closer the value is to 1, the better is the prediction rate. This means that high values should produce high weights, which is enforced by the formula. Weight values increase strongly as they get closer to one. On the other hand, lower values in classification performance will be penalized with low weight. The a value is used to control the slope of the function. A small a value will create a slower rising slope, beginning with lower x values. This means that even low performance numbers get a higher base weight and the impact of the formula is decreased. On the other hand, when higher values are used, only very well performing classifiers will have higher weights assigned. The maximum achievable weight is also decreased. To be able to control the highest assigned value and to avoid an infinite number for the weight, the b value can be adjusted. Very small b values will allow the weight value to rise very high. When x is close to 1, smaller b values will decrease the slope and also decrease the impact of the formula. For example, if a 1 is chosen for b, the weight formula is practically non-existent and, therefore, the classification can be compared to a majority vote.

III. Test Setup

For testing and result analysis, a python environment with the Scikit-learn library [15] was used. The project was complemented with developed code and the open source library is available on Bitbucket[1]. The tested smart grid network is identical to that in [9] and the metholology for when and how to pass information between layers is also identical.

A. Dataset and Scaling

To evaluate the classification performance, the KDD-NSL dataset is used. Even though this work suggests a Smart Grid solution, an Internet traffic dataset was chosen for the evaluation. This is due to the lack of available Smart Grid communication data. A dataset with attacks and Smart Grid traffic is not yet available. The files can be downloaded on the website of the Information Security Center of Excellence [16]. This dataset is a modified version of the dataset KDD Cup 1999. For the classification purpose, 41 features with four different attack types are contained in the dataset. It is composed with normal traffic, U2R, R2L, Probing and DoS attacks. Contrary to the KDD Cup 1999, the NSL dataset comes down to a size of about 20 Megabytes, which makes it very applicable for experimenting with classification systems. The NSL training set consists of 125,973 instances, the testing set has 22,544. The Training dataset has 45,927 DoS and 11,656 Probing instances. The amount of R2L and U2R types is rather small, only 52 instances of U2R and 995 of R2L. In contrast to that, the amount of R2L and U2R attacks in the testing set is fairly large. There are 2,938 R2L instances and 781 U2R attacks. This means that an efficient training for those attacks is rather difficult with the provided data. This condition implies that an IDS needs high zero-day attack detection capability, in order to be efficient.

TABLE I
Model results for atan-scaled KDD-NSL test set.

	Decision Tree	AdaBoost	kNN	SVM
True Positive	85%	91%	97%	96%
True Negative	78%	76%	68%	70%
False Positive	15%	9%	3%	4%
False Negative	22%	24%	32%	30%
Accuracy	81.2%	83.1%	79%	80%

For the dataset preparation, several steps had to be completed. In order to use the set with Scikit-Learn modules, the string types had to be mapped to numeric values. This was carried out with a developed library provided within the project. In addition, scaling methods were used for more efficient classification and to produce a variety of different outputs. This helped to further identify the voting-classifier performance. After the KDD-NSL set had been mapped to only numeric values, either no scaling was applied, normalized scaling with the range between -1 and 1 or Arctangent scaling was used.

Since the dataset has a very specific test set, another modification was carried out to add three more testing scenarios. Therefore, the whole KDD-NSL set, including test and training data, was randomized. This balanced the intrusion and normal traffic instances and generated a dataset with increased training and testing performance. The newly generated sets were split into sets and scaling was applied. An issue that had to be addressed was the requirement of a weight-calibration set, in order to determine each classifier performance and calculate the weight. To do this, the test sets were bisected and one part was used for calibration. The other half was used for testing. For the modified dataset, a 5% split of the randomized data was used for the calibration task.

B. Classifiers

The classifier selection was composed of a K-Nearest Neighbour (kNN) , Decision Tree, AdaBoost and a Support Vector Machine (SVM) model. kNN was configured with $k = 1$ and “distance” as weight. The Decision Tree classifier used a minimum of 20 leaf samples and the AdaBoost classifier was configured to use a DecisionTree algorihm with a maximum depth of 3. The maximum allowed estimators were set to 70. For the SVM, a Radial Basis Function kernel was chosen, with $C = 1$ and a degree of 3. For evaluation purposes, a confusion plot was created and the overall TN,TP, false positive (FP) and false negative (FN) percentages were calculated. In addition to that, the accuracy for overall efficiency was given. The stated setup of classifiers uses different algorithms, where each produce a different output. Table I shows a test run with an arctagent scaled NSL-KDD dataset.

[1] https://bitbucket.org/bgsufhs/python-intrusion-detection

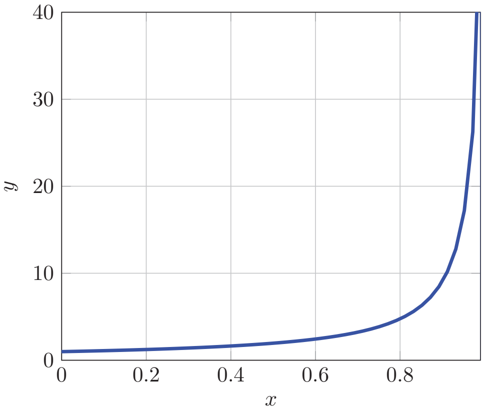


Fig. 1. Weight calculation function.

For the weight calibration, the constant a was set to 1, b was set to 0.01, resulting in (2) being used for weighting. For x, the specific TN or TP rate is used. TP rates are used to calculate the attack weight and TN is used for normal traffic weight. The chosen function values result in a graph as it is shown in Fig. 1.

$$f(x) = \frac{1}{(1-x)+0.01} \tag{2}$$

The graph rises when x is close to 0.85, which implies that 85% of the existing normal or attack instances in the dataset have been found. The intention with this formula is to reward good performances with a higher weight and increase the likelihood of a correct vote-prediction. To carry out the final classification, the assigned, summed up weights for attacks and normal predictions are compared with each other.

IV. Initial Results

For result analysis, the best accuracy, which was achieved by a single model, and the occurring false negative and positive values are compared with those of the vote-classifier. The results for all six scenarios are presented in Table II.

This data shows that voting was able to obtain a higher level of accuracy in most of the cases. To further illustrate the net performance improvements, Table 3 shows the subtracted values.

When the modified test set was used, each of the models started to become very efficient. In two cases, a large performance gap occurred between the classifiers, which caused the vote classification to be less accurate. In classification scenarios with lower performance, overall accuracy was increased significantly and either or both of the FP and FN rates were lowered. In the last case, the increase of performance is limited to very few false predictions, since the performance is already close to the 100%. Best possible results were achieved with the atan-scaled NSL set and the modified normalized set. The weighted vote was able to improve both the FP and FN rates and increased the accuracy. In two cases, individual models produced a lower FP rate than the vote classifier, but were not able to uncover many attacks. To further test, if this generalized approach of combination is exact enough, a script was programmed to iterate through the a and b values of the function, to find better working parameters. The results showed that only very little improvements on individual scenarios were achieved, which implies that the suggested function might be a good approximation to combine the models efficiently.

TABLE III
Net improvement values of vote classification.

Set/Scaling	Accuracy	False Positive	False Negative
NSL (Arc.)	+3.60%	-0.53%	-4.01%
NSL (Norm.)	+2.69%	+6.70%	-5.60%
NSL (None)	+2.31%	+4.26%	-4.06%
Mod. NSL (Arc.)	-0.02%	-0.09%	+0.10%
Mod. NSL (Norm.)	0.13%	-0.11%	-0.12%
Mod. NSL (None)	-0.03%	+0.00%	+0.05%

V. Conclusion

The reduction of FN or FP-rates and with it, the improvement of classification accuracy is a complicated task for even a single classifier. In the presented scenario, this is even more difficult as the use of an ensemble of techniques is being used, resulting in increased complexity. In contrast to this downside, several benefits have been discovered with the use of the voting-based ensemble classifier. With the experiments carried out on the classifier-voting system, it has been observed that a voting technique was able to show, in most cases, significant increase in prediction accuracy. When a combination of very efficient classifiers was used, the improvement of accuracy was either the same or only slightly significant. The positive aspects of this study included dropping FP- or FN-rates. Low FP rates are in many cases more favorable in machine-to-machine traffic, when the same accuracy can be achieved. Another aspect treated was the combination technique. Although the output is always dependent on the chosen classifiers and their performance, the weight balancing formula was able to produce favorable results in most of the test scenarios. This general approach with the stated formula proved to be successful in the different test scenarios. Based on these findings, it might be possible to apply a successful voting mechanism in a Smart Grid network. Especially since there are not any detailed attack scenarios available yet, clustering classifications or outlier detection might be used in the future. Those algorithms often have lesser performance than supervised classifiers and a voting scenario might be able to improve the resulting accuracy. Also, due to the machine-to-machine generated traffic of Smart Grid applications, a feature space for efficient classification might be developed more easily than in Internet applications. This will result in strong classification algorithms which can be improved afterwards with voting. Even if the enhancements are limited to several

TABLE II
MODEL RESULTS FOR ATAN-SCALED KDD-NSL TEST SET COMPARING SINGLE CLASSIFERS TO VOTE CLASSIFIERS.

	Single Classifier			**Weighted Voting Classifier**		
Dataset (Scaling Method)	**Accuracy**	**False Positive**	**False Negative**	**Accuracy**	**False Positive**	**False Negative**
NSL (Arctangent)	83.41%	8.75%	24.19%	87.01%	8.22%	20.18%
NSL (Normalized)	78.45%	2.57%	32.93%	81.14%	9.27%	27.33%
NSL (None)	78.03%	3.65%	33.04%	80.34%	7.91%	28.98%
Modifed NSL (Arctangent)	99.81%	0.20%	0.17%	99.79%	0.11%	0.27%
Modifed NSL (Normalized)	99.67%	0.23%	0.40%	99.80%	0.12%	0.28%
Modifed NSL (None)	99.86%	0.11%	0.15%	99.83%	0.11%	0.20%

percent, the output can avoid thousands of false or negative detections on the long run. Elaborated scenarios also showed the strong dependency on the chosen scaling method and training data. In terms of FP-Rates, one scenario showed a decrease of false attack predictions from 0.23% to 0.12%. Even though this does not seem much, it reduces the amount of FPs by roughly 50%. For Smart Grid implementations, this can be crucial to avoid false alarms and stabilize the system behavior. In addition to that, defective packets, which might be produced due to system errors, can very likely be detected by an anomaly IDS. In general, the scenarios showed that implementations may benefit from improved accuracy in Smart Grid applications. By combining this technique with other anomaly detection approaches, this technique could find its way in reliable IDS system for the Smart Grid.

VI. ACKNOWLEDGMENTS

The financial support of the Josef Ressel Center by the Austrian Federal Ministry of Science, Research and Economy and the Austrian National Foundation for Research, Technology and Development is gratefully acknowledged. Funding by the Austrian Marshall Plan Foundation is gratefully acknowledged.

REFERENCES

[1] R. Anderson and S. Fuloria, "Who controls the off switch?" in *IEEE International Conference on Smart Grid Communications (SmartGridComm)*, October 2010, pp. 96–101.

[2] C.-H. Lo and N. Ansari, "CONSUMER: A Novel Hybrid Intrusion Detection System for Distribution Networks in Smart Grid," *IEEE Transactions on Emerging Topics in Computing*, vol. 1, no. 1, pp. 33–44, June 2013.

[3] C. Wueest, "Targeted attacks against the energy sector: Security response," Symantec, Tech. Rep., 2014.

[4] M. Ahmed, A. N. Mahmood, and J. Hu, "A survey of network anomaly detection techniques," *Journal of Network and Computer Applications*, vol. 60, pp. 19–31, 2016.

[5] A. Murillo, "Review of Anomalies Detection Schemes in Smart Grids," Grupo de Teleinformatica e AutomacaoSymantec, Tech. Rep., 2013.

[6] J. Naruchitparames, M. Giine, and C. Evrenosoglu, "Secure communications in the smart grid," in *IEEE Consumer Communications and Networking Conference*, January 2011, pp. 1171–1175.

[7] R. Mitchell and I.-R. Chen, "Behavior-Rule Based Intrusion Detection Systems for Safety Critical Smart Grid Applications," *IEEE Transactions on Smart Grid*, vol. 4, no. 3, pp. 1254–1263, September 2013.

[8] O. Linda, M. Manic, and T. Vollmer, "Improving cyber-security of smart grid systems via anomaly detection and linguistic domain knowledge," in *International Symposium on Resilient Control Systems*, Aug 2012, pp. 48–54.

[9] Y. Zhang, L. Wang, W. Sun, R. Green, and M. Alam, "Distributed Intrusion Detection System in a Multi-Layer Network Architecture of Smart Grids," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 796–808, December 2011.

[10] Y. Ren, L. Zhang, and P. Suganthan, "Ensemble Classification and Regression-Recent Developments, Applications and Future Directions," *IEEEComputational Intelligence Magazine*, vol. 11, no. 1, pp. 41–53, February 2016.

[11] X. Mu, J. Lu, P. Watta, and M. Hassoun, "Weighted voting-based ensemble classifiers with application to human face recognition and voice recognition," in *International Joint Conference on Neural Networks*, June 2009, pp. 2168–2171.

[12] M. Panda and M. Patra, "Ensemble Voting System for Anomaly Based Network Intrusion Detection," *International Journal of Recent Trends in Engineering*, vol. 2, no. 5, 2009.

[13] T. G. Dietterich, "Ensemble Methods in Machine Learning," in *Proceedings of the First International Workshop on Multiple Classifier Systems*, ser. MCS '00. London, UK, UK: Springer-Verlag, 2000, pp. 1–15.

[14] A. Borji, *Advances in Computer Science – ASIAN 2007. Computer and Network Security: 12th Asian Computing Science Conference, Doha, Qatar, December 9-11, 2007*, Berlin, Heidelberg, 2007, ch. Combining Heterogeneous Classifiers for Network Intrusion Detection, pp. 254–260.

[15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[16] Information Security Center of Excellence, "Nsl-kdd data set for network-based intrusion detection systems," http://nsl.cs.unb.ca/NSL-KDD/, 2009.

3.6 KNIRSCH17A

- F. Knirsch, G. Eibl, and D. Engel. Multi-resolution privacy-enhancing technologies for smart metering. *EURASIP Journal on Information Security*, 2017(1):6, 2017.

RESEARCH

Multi-Resolution Privacy-Enhancing Technologies for Smart Metering

Fabian Knirsch[1,2*], Günther Eibl[1] and Dominik Engel[1]

[*]Correspondence:
[1]Salzburg University of Applied Sciences, Josef Ressel Center for User-Centric Smart Grid Privacy, Security and Control, Urstein Süd 1, 5412 Puch bei Hallein, Austria
Full list of author information is available at the end of the article

Abstract
The availability of individual load profiles per household in the smart grid end-user domain combined with non-intrusive load monitoring to infer personal data from these load curves has led to privacy concerns. Privacy-enhancing technologies have been proposed to address these concerns. In this paper the extension of privacy enhancing technologies by wavelet-based multi-resolution analysis (MRA) is proposed to enhance the options available on the user side. For three types of privacy methods (secure aggregation, masking and differential privacy) we show that MRA not only enhances privacy, but also adds additional flexibility and control for the end user. The combination of MRA and PETs is evaluated in terms of privacy, computational demands and real-world feasibility for each of the three method types.

Keywords: Smart meter; homomorphic encryption; masking; differential privacy; multiple resolutions; smart grid

Introduction

Intelligent energy systems, so-called smart grids, change the way electricity is generated, distributed and used. The widespread roll-out of smart meters is one of the consequences. Such smart meters record energy consumption in a specified granularity (usually the time between readings is between 1 and 15 minutes, cf. Table 10 in [1]) and have the ability to transmit these load curves in a specified interval (e.g., once a day). Therefore, this involves a considerable amount of information that needs to be processed and analyzed. Smart grids further demand accurate and fine-grained data on network status, as well as a detailed analysis of load profiles from customers [2]. This is crucial for applications such as billing with dynamic pricing, demand response and network monitoring.

However, it has been shown that personal information on the end-user can be inferred from fine-grained load curves [3, 4], and this has led to privacy concerns (e.g., [5]). This also implies some severe privacy threats such as the identification of customer presence at home, customer habits and even the customer position when using electric vehicles [6]. In [7] and [8], the authors show the impact of resolutions on privacy and that information can be deduced even at comparably low frequencies.

The accuracy of the inferred information is directly connected to the available resolution of the load data. A number of methods have been proposed to balance the need for privacy with the information needed for correct operation of smart grids. Two types of approaches show high potential to resolve this issue: (i) privacy-aware aggregation of encrypted load curves; and (ii) representation of load curves in multiple resolutions, each associated with different access levels.

Privacy-aware Aggregation

Approaches for privacy-aware aggregation can again be divided into three categories: protocols using **masking** [9, 10], protocols using secure aggregation by **homomorphic encryption** [11, 12] and protocols using **differential privacy** [13, 14]. In this paper, the focus is put on the application of multi-resolution load curve representation in combination with secure aggregation protocols.

Privacy-enabling encryption for smart meter data by the use of homomorphic encryption is suggested by, e.g., [11, 12, 15, 16], allowing the aggregation of encrypted signals, also termed "secure signal processing". A recent overview of secure signal processing, covering four proposals for privacy-preserving smart metering aggregation is given in [17]. Protocols that are using masking for aggregating data have been proposed by, e.g., [9, 10, 18]. Masking approaches aim to hide individual contributions by additive noise, but still produce a valid aggregate. Differential privacy follows a similar approach, where contributions are hidden in a noisy aggregate that fulfills some statistical properties. Differential privacy is adapted for applications in the smart grid by, e.g., [13, 19, 20, 21].

Multiple Resolutions

Approaches of this type suggest to represent load curve data in multiple resolutions, where each resolution can be used for a different purpose – e.g., low resolution for billing – and is therefore disclosed to selected parties only, e.g., [22]. Using the wavelet transform in order to produce an integrated bitstream supporting multiple resolutions has been proposed by [23]. Combined with conditional access, i.e., different encryption keys for each resolution [24], this wavelet-based representation allows user-centric privacy management: access can be granted or revoked for each resolution. Access to high resolutions, which are privacy-sensitive, may be reserved to a small number of trusted entities only, whereas resolutions of medium granularity may be provided more freely, e.g., to contribute to network stability (in exchange for lower energy prices or other incentives). An approach combining multiple resolutions and direct user control for smart metering is shown in [25]. The combination of MRA with homomorphic encryption, which is also one of the topics in this paper, has been discussed in [26].

Contribution

In this paper, a set of three privacy-preserving smart metering data aggregation methods that combine the two types of approaches, namely multi-resolution representation and (i) homomorphic encryption; (ii) masking; and (iii) differential privacy, is proposed. This improves the capabilities for managing privacy requirements, as the combination of "traditional" privacy enhancing methods with multi-resolution representation significantly increases the choices available for both, system operator, and end user. We further contribute the sketch of a protocol for distributing keys and providing distinct resolutions to different parties. Access control does not relate to the aggregated signal as a whole anymore, but access can be granted on the aggregate on each resolution *individually*. This is an important feature, as it allows to grant access to participants in the smart grid system, based on their roles and the functions they have to fulfill. Each role can be assigned access to the aggregate on the minimum resolution necessary to fulfill the functions associated with this role.

The combination of MRA with homomorphic encryption has previously been proposed in [26]. This paper extends the previous work by applying multi-resolution techniques to masking and differential privacy. A comprehensive presentation, discussion and evaulation of multi-resolution representation in combination with widely used PETs is given.

The rest of this paper is structured as follows: In Section Multi-resolution PETs the application scenario and common definitions are introduced. In Section Background, background is presented on wavelets for the multi-resolution representation of load curves as well as on the three privacy enhancing technologies (PETs) homomorphic encryption, masking and differential privacy. Sections Multi-resolution Secure Aggregation, Multi-resolution Masking and Multi-resolution Differential Privacy describe each of these PETs individually and propose the combination of these approaches with wavelets. In Section Evaluation the security features of the proposed protocols, as well as cost and complexity are discussed and further, the system is evaluated with respect to real-world applicability on the basis of a prototypical implementation. Section Conclusion and Outlook summarizes this paper and gives an outlook to future work.

Multi-resolution PETs

While homomorphic encryption, simple masking and differential privacy are efficient methods for the *spatial* reduction of resolution, *temporal* aggregation is not sufficiently covered with any of these approaches. Temporal resolution of time series can be reduced by subsequently applying a number of filters. When – for instance – applying an appropriate lowpass filter to a time series, all frequencies above the cutoff-frequency are omitted, which results in a signal with less information. This is effectively performed by applying the wavelet transform, in particular the Haar wavelet, to a series of values.

Application Scenario

Smart meter data has a wide range of applications, such as in-house monitoring, billing, network monitoring and demand response. As pointed out in [8, 2], data resolution depends on the use case and has an impact on the privacy, i.e., the information the recipient can gain from that data. In the following we introduce three typical application scenarios and motivate the need for multi-resolution PETs and their flexibility with respect to spatial and temporal resolution.

1 **Settlement and Profiling.** In the energy market electricity generators and electricity suppliers trade at a wholesale marketplace. The arrangement of payments among these parties is called settlement [9, 2]. Profiling is used for determining forecasts and training models, e.g., in the UK this is based on half-hourly meter data from a representative sample of households [2]. Both applications thus require data in a comparably low resolution, but spatially aggregated over a number of households.

2 **Network Monitoring.** Network monitoring is used for detecting outages and peaks, and thus maintaining the stability of the power grid. A detailed monitoring of power consumption, voltage levels and phase shifts is an important feature for network operators. For monitoring purposes, data at a high temporal resolution but with little spatial resolution is required.

3 **Billing.** Billing requires meter data in a low temporal resolution (e.g., one value per month or year), however on a per household or on a per meter basis, hence not spatially aggregated at all. In future applications, dynamic pricing might also require more fine grained data [27]. Multi-resolution PETs enable the provision of load profiles in certain resolutions depending on the particular use case.

Topology

For data aggregation in the smart grid, a number of different topologies are proposed, such as star topologies (e.g., [20], [28]), ring topologies (e.g., [10]) and tree topologies (e.g., [16]). In any case, the smart meters generate a time series of values that is either sent to a dedicated collector node or data concentrator, which is responsible for aggregating these measurements. Or the smart meters aggregate in a hop-by-hop manner, i.e., a smart meter sends its measurement to its successor or parent node where this measurement is combined with its own value. The data concentrator and the last smart meter, respectively, forward the aggregated measurements to one or more recipients (in the following referred to as *aggregators*). Each aggregator receives data in a different spatio-temporal resolution depending on the role of the recipient and the needed granularity. Fig. 1 shows examples for star and tree topologies. For the protocols presented in this paper, the aggregation method (direct or hop-by-hop) is not restricted and either of these approaches can be used.

Problem Statement and Definitions

Given a number of smart meters SM_i, for $i = 1 \dots N$, one or more aggregators A_k, for $k = 1 \dots M$, and a trusted third party (TTP), each meter i measures a time series of values, i.e., at time t it measures $m_{i,t}$. In this paper, a series of values measured by a meter i is denoted as m_i. In order to protect customer privacy, the sum of the energy consumption for all smart meters should be provided to the aggregator. The following restrictions and requirements apply (aggregator oblivious): (i) no aggregator can gain any information about individual contributions; (ii) each aggregator can only unmask a valid sum up to the time resolution $r \leq R$ (with R as the maximum resolution) that is intended to be revealed for this aggregator. Hence, the aggregator is considered to be untrusted. In practice, the smart meters can be considered to be physically arranged in either a tree or a ring topology. Logic topologies may defer and depend on the concrete protocol. For homomorphic encryption and masking, the TTP is needed to provide the keys (pk^r, sk^r) and the key shares (key^r), respectively, to the smart meters and aggregators.

For this paper we assume that there is a sufficient underlying secure communication infrastructure, i.e., the bidirectional and reliable exchange of information and the secure distribution of keys is given as well as authenticated communication among participants is guaranteed by, e.g., AES [29] and X.509 certificates [30]. We further assume all devices to be tamper-proof, i.e., the meter value itself cannot be manipulated.

Background

In this section, we briefly review the existing work on multi-resolution representation, homomorphic encryption, masking and differential privacy.

Wavelet-based Representation

A wavelet transform starts with the original load curve $m = (m_1, m_2, \dots, m_T)$, which denotes a series of values. Each step splits the original load curve into a highpass component h and a lowpass components l. If the wavelet transform is performed recursively in d steps, this is denoted as $W_d(m)$. In each step q, for $q = 1, \dots d$, half of the data (the highpass data) h_q are stored as the wavelet coefficients (subband) of scale q and the next step is performed for the lowpass data. At the end of the transformation the final subband h_d consists of a fraction of 2^{-d} samples compared to the original load curve. The higher the scale q, the lower the time resolution $r := d - q$. Reindexing, and introducing the notation $h^r = h_{d-q}$, at the end of the transformation one obtains a sequence $h = (l_0, h_1, \dots, h_d)$.

The synthesis step of the inverse wavelet transform W^{-1} starts with the lowest resolution $r = 0$. To get the next higher resolution of the signal the next higher resolution subband is needed, so that in a series of d steps one finally obtains the original load curve (since we only consider lossless transformations). In order to provide a signal m^r with maximum resolution r, only r synthesis steps must be performed and only the subbands with resolution $r \leq R$, i.e., $m^r = (l_0, h_1, \dots, h_r)$, are needed. Denoting the selection of the r highest resolutions as a function T_r, this can be written as

$$m^r = W^{-1} \left(T_r(W(m)) \right). \tag{1}$$

This selection can be realized in practice by replacing the highpass subbands with zeros, i.e., applying $T_r(\cdot)$ to a sequence $W(m) = (l_0, h_1, \dots, h_r, \dots, h_{d-1}, h_d)$ yields a sequence $T_r(W(m)) = (l_0, h_1, \dots, h_r, 0, \dots, 0)$. This limits, after applying the inverse wavelet transform, the resolution of the signal. Making the signal available at the needed resolution instead of the full resolution increases privacy because less (personal) information can be deduced [8].

In [23], a variety of wavelet filters regarding their utility for the multi-resolution representation of load curves was evaluated. Only lossless transformations are useful in the context of smart metering. The Haar wavelet filter preserves the average over all resolutions, which is an important property for many use cases. Using the lifting implementation of the Haar wavelet, the transformation can be realized efficiently.

The lifting steps for the forward transform with the Haar wavelet have been formulated by [31]. As the original Haar wavelet uses real coefficients, it is ill-suited for use with homomorphic encryption. Therefore, for the combination with PETs, a modified version of the Haar wavelet is used that only produces integer values for the transformed load curve. While this is generally not an issue for masking and differential privacy, we still use the modified version for all PETs. A detailed description of the Haar wavelet lifting scheme can be found in [23]. Note that the average of the original series is still preserved over all resolutions for the modified Haar filter:

$$\forall r : \sum_{t=0}^{T} m_t = 2^{-r} \sum_{t=0}^{T} m_t^r. \tag{2}$$

Additive Homomorphic Encryption

Following previous proposals [11, 12, 15], for this work the Paillier cryptosystem [32] is employed. This additive homomorphic cryptosystem has the following important property, which is called the *additive property*:

$$D\left(E(m_1)E(m_2) \bmod n^2\right) = (m_1 + m_2) \mod n. \tag{3}$$

This property means that the decryption of the product of the *ciphertexts* is the sum of the original plaintext messages.

In a practical setting, the network is assumed to have tree-like connections. Each smart meter sends its measured load in encrypted form to its parent node. The parent smart meter multiplies the obtained encrypted signals with its own encrypted signal and in turn sends this product to its parent node. Finally, the aggregator multiplies the obtained signals and decrypts the product. Due to the additive homomorphic property, the result is the sum of the measurements. With E and D denoting Pailler encryption and decryption this can be stated as

$$D\left(\prod_i E(m_i) \bmod n^2\right) = \sum_i m_i \mod n. \tag{4}$$

Privacy is preserved because of the distributed way of processing. Smart meters only have the plaintext information of their own messages, because they cannot decrypt the messages they get. The aggregator can decrypt messages, but, as it receives the product of the individual ciphertexts, can only decrypt the sum of the load curves.

Masking

Masking refers to the obfuscation of individual contributions, such that the summation of load profiles over a number of households yields the correct sum, but no individual contribution is traceable. This is achieved by adding for each SM_i at time t a random share s_i in the range $1, \ldots, \kappa - 1$ to the meter value m_i. This results in a masked meter value $\tilde{m}_i = m_i + s_i \mod \kappa$. The set of random shares is constructed in such a way that

$$\sum_i \tilde{m}_i = \sum_i (m_i + s_i) = \sum_i m_i \text{ (all mod } \kappa), \tag{5}$$

hence the shares cancel each other out upon summation.

Principally, smart meters calculate the masked value $\tilde{m}_i$ and submit this value to an aggregator. Once the aggregator has received all masked values, it can calculate the unmasked sum. If a single value is missing, the secret shares will not cancel each other out, and neither the aggregate, nor any individual contribution can be reconstructed.

Kursawe et al. [9] present a number of methods for constructing such shares that meet the requirement for untraceability of individual contributions: (i) aggregation protocols for determining the sum as described above; and (ii) comparison protocols that require the aggregator already knows an (at least) approximate sum. For our purpose we focus on the low-overhead protocol from the first group which has already been used in practical implementations [33]. For the low-overhead protocol all smart meters hold a public key $\mathrm{pk}_i = g^{X_i}$ with X_i as a secret key and $g \in \mathbb{G}$ as a generator of a group satisfying the computational Diffie-Hellman assumption [34]. Each SM_i is given the set of all public keys and computes a set of $N-1$ shared keys by $K_{i,j} = H(\mathrm{pk}_j^{X_i})$ with $j = 1 \ldots N$.

As described in [9], for each meter value at time t each SM_i creates a random share by

$$s_i = \sum_{k \neq i} (-1)^{b(i,j)} H(K_{i,j}||t), \tag{6}$$

where $b(i,j)$ returns 1 if $j < i$ and 0 otherwise, and $H : \{0,1\}^* \to \mathbb{G}$ is a hash function mapping its input to an element of $\mathbb{G}$. This term in Equation 6 results in $+H(K_{i,j}||t)$ for $b(j,i) = 0$ and $-H(K_{i,j}||t)$ for $b(j,i) = 1$. Summing up this values assures that all s_i cancel each other out pairwise since $K_{i,j} = K_{j,i}$ because of $g^{X_i X_j} = g^{X_j X_i}$. This is shown for N smart meters for one point in time t in Table 1, where the rows represent k and the columns represent i for values from 1 to N. Summing up the resulting terms in each row yields the random share $s_{i,t}$.

Differential Privacy

Differential privacy is a privacy definition that defines privacy of a function f by an indistinguishability property of the function result. In this paper the function is the time series of the sum of different smart meter measurements $f(t) = \sum_{i=1}^{N} T_r\left(W(m_i)\right)$. However, note that here the noise is added to the selected resolutions (operator T_r) in the wavelet domain and not in the original domain. The aim is that by examining a perturbed result $\tilde{f}(t)$, one cannot distinguish whether a single person's entry is contained or not. Since the noise is only added to the needed resolutions $\leq r$, only a small amount of noise is added. More formally, two neighboring datasets $\mathcal{D}$ and $\mathcal{D}'$ that differ in the entries of a single person/household only are considered. The function mechanism $\tilde{f}$ is then ϵ-differentially private, if for a small privacy parameter $\epsilon > 0$

$$\Pr[\tilde{f}(\mathcal{D}) = y] \leq \exp(\epsilon) \Pr[\tilde{f}(\mathcal{D}') = y]. \tag{7}$$

While differential privacy is a theoretically appealing definition with nice properties (e.g., a function is differentially private under postprocessing), it is achieved by perturbing the function result with Laplacian noise $\tilde{f}(t) = f(t) + n_t$ [19], where each noise value n_t is independently and identically sampled from a Laplacian distribution $n_t \sim \mathrm{Lap}_\lambda$ (the parameter λ must be set using the sensitivity of the function

f [19]). As a drawback, the function result is not exact and can be useless if the number of entries in the dataset is too small.

More specifically, according to the Theorem of Dwork [19], the L_p sensitivity of a function $f : D^n \to \mathbb{R}^d$ is the smallest number $S_p(f)$ such that for two neighboring datasets x and x'

$$S_p(f) = \underset{x,x'}{\operatorname{argmax}} \|f(x) - f(x')\|_p. \tag{8}$$

The most common mechanism that achieves differential privacy is the Laplace mechanism $\mathcal{M}_L$ that perturbs the output of f by adding noise from a Laplace distribution having the density

$$\mathrm{Lap}_\lambda(x) = \frac{1}{2\lambda} \exp\left(-\frac{|x|}{\lambda}\right), \tag{9}$$

in a non-interactive way, yielding

$$\mathcal{M}_L(x, f(\cdot), \epsilon) = f(x) + (Y_1, \ldots, Y_k), \quad \text{with } Y_l \overset{i.i.d.}{\sim} \mathrm{Lap}_\lambda. \tag{10}$$

An important theorem states that the Laplace mechanism is ϵ-differentially private, if the parameter λ is chosen by

$$\lambda = \frac{S_1(f)}{\epsilon}. \tag{11}$$

The resulting noise does not need to be added directly to the function result. It can also be added in a distributed manner [20, 35] when each contributing party i adds i.i.d. noise $G_{\lambda,N}$ defined by

$$\Pr[G_{\lambda,N} = x] = \mathrm{G}^1_{1/N,\lambda}(x) - \mathrm{G}^2_{1/N,\lambda}(x), \tag{12}$$

where G^1 and G^2 are two i.i.d. gamma distributions with identical shape parameter $1/N$ and scale parameter λ. Then

$$\Pr[n_t = x] = \sum_{i=1}^{N} G_{\lambda,N}(x) = \mathrm{Lap}_\lambda(x). \tag{13}$$

Multi-resolution Secure Aggregation

In this section the combination of the wavelet transform with homomorphic encryption is presented. The principal scheme is shown in Fig. 2. First, the basic approach with only one aggregator is presented and second, it is shown that this approach can easily be extended to multiple aggregators.

Principal Secure Aggregation Scheme

Homomorphic encryption is applied to each resolution separately with a different pair of keys $(\mathrm{pk}_r, \mathrm{sk}_r)$ for each resolution r. The resulting signal m is the sum of all signals m_i (each of which has a maximum resolution of R) at resolution $r \leq R$, whereby $W(\cdot)$ denotes a wavelet transformation. The collector node can perform aggregation (i.e., multiply) in the encrypted domain, i.e., it does not have any keys. This ensures that the aggregating node cannot get information about the loads of its children, e.g., by divisions.

Basic Approach

The basic approach covers a number of smart meters and a single aggregator. Writing the principal scheme mathematically yields the following calculation of the ciphertext c

$$c = \prod_i E\left(T_r\left(W\left(m_i\right)\right)\right) \mod n^2. \tag{14}$$

The ciphertext c is decrypted by the aggregator by

$$m = W^{-1}\left(D\left(c\right) \mod n\right). \tag{15}$$

Using this procedure, the wavelet transformation is compatible with homomorphic encryption, i.e., the property that the message m equals the sum of the messages is preserved (choosing $r = R$). Even more, choosing $r \leq R$, the decrypted message m equals the sum of the messages of resolution r:

$$m = W^{-1}\left(D\left(\prod_i E\left(T_r\left(W\left(m_i\right)\right)\right) \mod n\right)\right) = \sum_i m_i^r \mod n. \tag{16}$$

The aggregator gets the product of the encrypted messages and can therefore not extract any information about the individual messages. However, it can calculate the sum of the messages which is the information needed, e.g., for load forecasting. Note again that the product of the ciphertexts is calculated in either a distributed way by the smart meters or by a data concentrator and not by the aggregator (see Section Topology). The number n must be chosen depending on the desired security level. It further determines the aggregation group size, since $\prod_i E\left(T_r\left(W\left(m_i\right)\right)\right) < n^2$ and $D\left(\prod_i E\left(T_r\left(W\left(m_i\right)\right)\right)\right) < n$. In Section Space considerations the issue of aggregation group sizes is discussed in detail. For the sake of readability the modulus parts of the calculations are omitted in the following proof.

Proof Without loss of generality two messages are considered. To simplify the analysis the notation $y_i := T_r\left(W\left(m_i\right)\right)$ is used, so $E\left(T_r\left(W\left(m_i\right)\right)\right) = E(y_i)$. The aggregator calculates the signal $W^{-1}(D(c))$. Using the fact that the ciphertext c is the product of the individual ciphertexts and the homomorphic encryption property leads to

$$\begin{aligned} W^{-1}\left(D\left(c\right)\right) &= W^{-1}\left(D\left(c_1 c_2\right)\right) \\ &= W^{-1}\left(D\left(E\left(y_1\right)E\left(y_1\right)\right)\right) \\ &= W^{-1}(y_1 + y_2) \end{aligned} \tag{17}$$

Substituting the y_i, using the linearity of the wavelet transform and the definition of m^r yields

$$\begin{aligned} W^{-1}\left(D\left(c\right)\right) &= W^{-1}\left(T_r\left(W\left(m_1\right)\right) + T_r\left(W\left(m_2\right)\right)\right) \\ &= W^{-1}\left(T_r\left(W\left(m_1\right)\right)\right) + W^{-1}\left(T_r\left(W\left(m_2\right)\right)\right) \\ &= m_1^r + m_2^r \end{aligned} \tag{18}$$

So in general for N different messages and ciphertext $c = \prod_i c_i$, the desired property (16)

$$W^{-1}\left(D\left(c\right)\right) = \sum_{i=1}^{N} m_i^r. \tag{19}$$

is obtained. □

Multiple Aggregators

An example use-case scenario is the use of aggregated load information for energy monitoring by the network operator, as, e.g., suggested by [17]. The approach proposed here adds an additional layer of flexibility by making the aggregates available at different resolutions and only grant access to parties on the resolutions they need to fulfill a specific task. In combination with suitable key management, this approach implements the "need-to-know" principle of access for aggregated signals. The secure aggregation scheme presented above can be extended to support multiple aggregators. Each aggregator receives data in a certain resolution. This is easily achieved by encrypting with different keys at the collector node.

Multi-resolution Masking

In this section the multi-resolution masking approach is presented. The principal scheme is shown in Fig. 3. After briefly recapitulating the principal masking scheme, first, the basic approach for one aggregator is presented and second, this approach is extended to multiple aggregators receiving data in different resolutions. The latter is especially useful for application scenarios such as settlement and profiling, where different parties should be provided information in different resolutions.

Principal Masking Scheme

Each smart meter SM_i calculates at each time $t = 0 \ldots T$ a masked value $\tilde{m}_{i,t}$ by adding a random share $s_{i,t}$ to its measured value $m_{i,t}$. Upon spatial aggregation the shares s_i cancel each other out and the aggregator receives an unmasked sum. Note that in the following, operations involving masking of type $a + b \mod \kappa$ are written as $a + b$, i.e., the modulo parts are omitted for the sake of brevity and readability.

This approach can be enhanced by allowing to reduce the temporal resolution of the signal. Even more, a number of different resolutions can be provided within the same bitstream and the key for a certain resolution is only given to the aggregator. This is achieved by applying a wavelet transform to the signal. Hence, even if the aggregator is given the full load curve data, it can only unmask the bitstream up to the resolution for which it holds the key share.

Basic Approach

The basic approach describes spatio-temporal masking with one aggregator.

Initialization. TTP agrees with all smart meters in the group $G = \{SM_1, \ldots, SM_N\}$ on providing a resolution r of a total of T values to an aggregator A.

Masking. Simultaneously, all SM_i and TTP calculate a random share $s_{i,t}$ for $t = 0 \ldots T$, as described for the principal masking above. Each smart meter now holds a set of shares s_i and TTP holds a key share key.

All SM_i now calculate a series of masked values $\tilde{m}_i = W(m_i) + s_i$ and submit this series to A. TTP calculates the key share key^r for the resolution r of its key share by $T_r(\text{key})$ and submits this to A. Note that the wavelet transform is only applied to the metered value, and before adding the random share.

Aggregation. After receiving both, the shares from all smart meters and the key share, A can calculate the aggregated sum over all smart meters at a time resolution r by

$$\sum_i m_i^r = W^{-1}\left(T_r\left(\sum_i \tilde{m}_i\right) + \text{key}^r\right). \tag{20}$$

If the aggregator attempts to retrieve any resolution $r^+ > r$, the result will be noisy and useless. However, the aggregator may reconstruct arbitrary resolutions $r^- \leq r$ from the data.

Proof Proof that reconstructing a resolution r^+ for a key with resolution r will be noisy. The aggregator receives an aggregation of the masked meter values

$$\sum_i \tilde{m}_i = \sum_i \left(W\left(m_i\right) + s_i\right), \tag{21}$$

and a key share $\text{key}^r = T_r(\text{key})$ for some resolution r. Applying this function $T_r(\cdot)$ to a series of values replaces the highpass components by zeros. The key share and the random shares for masking have the property that

$$\sum_i s_i + \text{key} = 0, \tag{22}$$

but that the key share for a particular resolution r yields

$$\sum_i s_i + \text{key}^r \neq 0, \tag{23}$$

since the highpass components are set to zero in the key share and do not cancel out the corresponding components in the sum of the shares. Therefore,

$$W^{-1}\left(\sum_i \left(W\left(m_i\right) + s_i\right) + \text{key}\right) = \sum_i m_i, \tag{24}$$

and

$$W^{-1}\left(\sum_{i}\left(W\left(m_{i}\right)+s_{i}\right)+\text{key}^{r}\right)\neq\sum_{i}m_{i}. \tag{25}$$

However, after applying the function $T_r(\cdot)$ with the same parameter r to the equation, this yields Equation 20 which is the correct result for this particular resolution r. Note that the wavelet transform is recursively applied to the resulting lowpass band, i.e., any resolution $r^{-} < r$ can be retrieved, since applying $T_r(\cdot)$ to the key share only replaces the highpass components by zero and only the lowpass components remain for reconstructing the signal.

□

Note that this scheme fulfills both of our initial requirements: (i) individual contributions are masked and the aggregator cannot gain any information without having all the values from all $SM_i \in G$; and (ii) the highest resolution that is accessible for the aggregator is determined by the resolution of the key share.

Multiple Aggregators

The scheme we present in the following extends the basic approach with multiple aggregators that receive data in different resolutions. Extending the scheme requires more overhead and communication than for the secure aggregation. A simple approach would be to have multiple bitstreams in multiple resolutions for each aggregator. The advantage of the MRA approach is, however, to have all the information for different resolutions in a single bitstream where no data expansion occurs. Therefore, a different key share for every recipient is created with the trade-off of distributing an aggregate of $M-1$ key shares in addition to the actual key share.

Initialization. For the enhanced scheme supporting multiple aggregators $L = \{A_1, \dots, A_M\}$, a TTP agrees with all smart meters in the group $G = \{SM_1, \dots, SM_N\}$ on providing a resolution r_k of a total of T values to each aggregator $A_k \in L$.

Masking. As in the basic scheme, all smart meters SM_i calculate a random share $s_{i,t}$ for $t = 0 \dots T$. Again, each smart meter now holds a set of shares s_i, calculates the series of masked values $\tilde{m}_i = W(m_i) + s_i$ and submits this series to all aggregators $A_k \in L$. TTP calculates a total of M (number of aggregators) key shares $\text{key}_1, \dots, \text{key}_M$. For each key share $k = 1 \dots M$, TTP further calculates the resolution r_k by $\text{key}_k^{r_k} = T_{r_k}(\text{key}_k)$ and submits this to A_k. It further submits the sum of all other key shares $\sum_{i \neq k} \text{key}_i$ to A_k.

Aggregation. After receiving both, the shares from all smart meters and the set of key shares, each $A_k \in L$ can calculate the aggregated sum over all smart meters at a time resolution r_k by

$$\sum_{i} m_i^{r_k} = W^{-1}\left(T_r\left(\sum_{i}\tilde{m}_i\right) + \text{key}_k^{r_k} + T_r\left(\sum_{i\neq k}\text{key}_i\right)\right). \tag{26}$$

As for the basic approach, both of our initial requirements are fulfilled: (i) individual contributions are masked and none of the aggregators can gain any information without having all the values from all $SM_i \in G$ and the sum of all other key shares $\sum_{i \neq k} \text{key}_i$; and (ii) the highest resolution that is accessible for each aggregator is determined by the resolution of the individual key share. These requirements are fulfilled due to the properties of the masking approach as introduced in Section Masking and formally shown in Section Basic Approach.

Proof of Correctness

In the following it is shown that applying the wavelet transform to a meter value and masking can be combined in order to provide a certain resolution only. This proof is – for simplicity and without loss of generality – for a single smart meter and a single aggregator. The proof also applies to multiple smart meters and multiple aggregators. The only difference is that instead of a single meter value, share and key, respectively, a (spatially) aggregated sum of values is used. For multiple aggregators the sum of all other key shares is also required as shown in the previous section.

Proof Proof for a single aggregator that it is receiving m_i^r at the end of the above masking scheme. Starting from

$$\underbrace{W\left(m_i\right) + s_i}_{SM_i} + \underbrace{\text{key}}_{TTP} = \underbrace{W\left(\hat{m}_i\right)}_{A}, \tag{27}$$

where the braces indicate what the smart meter and the TTP calculate, respectively, and what the aggregator receives at the end of the protocol, $T_r \circ W^{-1}$ is applied on both sides of the equation:

$$W^{-1}\left(T_r\left(W\left(m_i\right) + s_i + \text{key}\right)\right) = W^{-1}\left(T_r\left(W\left(\hat{m}_i\right)\right)\right). \tag{28}$$

Due to the linearity of both, the wavelet transform and the function $T_r(\cdot)$ this is equivalent to

$$W^{-1}\left(T_r\left(W\left(m_i\right)\right)\right) + W^{-1}\left(T_r\left(s_i\right) + T_r\left(\text{key}\right)\right) = \hat{m}_i^r. \tag{29}$$

Substituting $m_i^r = W^{-1}\left(T_r\left(W\left(m_i\right)\right)\right)$, $s_i^r = T_r\left(s_i\right)$ and $\text{key}_i^r = T_r\left(\text{key}\right)$ results in

$$m_i^r + W^{-1}\left(s_i^r + \text{key}_t^r\right) = \hat{m}_i^r. \tag{30}$$

Given the property of masking, shares cancel each other out by $s_i^r + \text{key}^r = 0$, and therefore $m_i^r + W^{-1}\left(0\right) = \hat{m}_i^r$. This is obviously equivalent to $m_i^r = \hat{m}_i^r$, i.e., the aggregator only receives a certain resolution m_i^r of the original meter value m_i. □

Multi-resolution Differential Privacy

In this section, it is shown that the wavelet approach can be combined with an additional differential privacy method. The benefit of this approach is an additional ϵ-differential privacy guarantee (Equation 7) for the resulting *aggregated* signal.

Combining Wavelets and Differential Privacy

Combining differential privacy in a *distributed* way with wavelets, only guarantees differential privacy for the sum, but not for the individual signals. Therefore, combining differential privacy with wavelets alone, would not enhance privacy so that the combination with homomorphic secure aggregation is needed. Similar to the masking approach, the combination with the differential privacy method requires a distributed addition and later summation of random values. The scheme is described in Fig. 2 and leads, using the additive homomorphic property of the encryption, to the following intermediate result.

$$\tilde{f} = W^{-1}\left(D\left(\prod_{i=1}^{N} c_i^{r_i}\right)\right) = W^{-1}\left(\sum_{i=1}^{N} w_i^{r_i}\right) = W^{-1}\left(\sum_{i=1}^{N}\left(T_{r_i}\left(W\left(m_i\right)\right) + G_{\lambda,N}\right)\right). \tag{31}$$

The additional use of homomorphic encryption is not the only difference to masking. In contrast to masking the random values are drawn *independently* from each other from a non-uniform probability distribution $G_{\lambda,N}$, denoted as block DP in Fig. 2. Due to Equation 13 these distributedly generated probability distributions sum up to the Laplacian distribution which is needed for differential privacy of the *aggregate* profile

$$\tilde{f} = W^{-1}\left(\sum_{i=1}^{N} T_{r_i}\left(W\left(m_i\right)\right) + \mathrm{Lap}_\lambda\right). \tag{32}$$

Thus, if the noise parameter λ is chosen such that $\sum_{i=1}^{N} T_{r_i}\left(W\left(m_i\right)\right)$ is ϵ-differentially private, due to the postprocessing property of differential privacy, also $\tilde{f} = W^{-1}\left(\sum_{i=1}^{N} T_{r_i}\left(W\left(m_i\right)\right) + \mathrm{Lap}_\lambda\right)$ is ϵ-differentially private. Finally, using the linearity of W,

$$\tilde{f} = \sum_{i=1}^{N} m_i^{r_i} + W^{-1}(\mathrm{Lap}_\lambda), \tag{33}$$

is shown to be a perturbed function of the smoothed consumption sum. This smoothed consumption sum is ϵ-differentially private, if the Laplacian noise is set in the right manner. Therefore, in principle the wavelet decomposition is compatible with differential privacy.

Another difference to the presented masking scheme is that the noise is added to the *restricted* wavelet values instead of the unrestricted values $W(m_i)$ (Equation 32). However, since several different resolutions occur, setting the right amount of noise λ is not trivial and remains a task for future research. First preliminary steps in that direction show that it is possible to derive a choice for λ which, however, only provides differential privacy for a single resolution r. With such a noise differential privacy can only be provided for a single resolution r and, due to the post-processing property, all coarser solutions.

Choice of Parameter λ

In this subsection we show how the parameter λ must be chosen by proving the following theorem.

Theorem (choice of λ): The presented algorithm is ϵ-differentially private, if (i) W is a tight frame; and (ii) parameter λ is chosen as

$$\lambda = \frac{\sqrt{\mathcal{R}}}{\epsilon} \underset{m_{i,\cdot}}{\operatorname{argmax}} \left\| m_{i,\cdot} \right\|_2 , \tag{34}$$

where $\mathcal{R}$ denotes the number of coefficients up to resolution $r = d - q$.

Note that $\mathcal{R}$ consists of a fraction of 2^{-q} samples compared to the original load curve. The smaller the resolution r, the smaller λ and therefore the added noise is chosen.

Proof First, the situation of this algorithm must be properly mapped into the differential privacy setting. Note that the term $\tilde{w}^r$ of the algorithm can be rewritten as

$$\tilde{w}^r = \sum_{i=1}^{N} \left(T_r \left(W(m_i) \right) + \mathrm{G}^1_{1/n,\lambda_\epsilon}(x) - \mathrm{G}^2_{1/n,\lambda_\epsilon} \right) . \tag{35}$$

Due to the divisibility property, the sum of the Gamma-distributions yield the Laplace distribution. Thus, we have

$$\tilde{m}^r = W^{-1}(\tilde{w}^r) = W^{-1}\left(\mathcal{M}_L(m, f(\cdot), \epsilon)\right) = W^{-1}\left(f(m) + \mathrm{Lap}_{\lambda_\epsilon}\right) . \tag{36}$$

If we manage to prove differential privacy for our choice of f, the proof is finished since a function applied to a differentially private mechanism can not destroy the differential privacy property (closure under post-processing property of differential privacy). Therefore, if $\tilde{w}^r$ is ϵ-differentially private this also holds for $\tilde{m}^r = W^{-1}(\tilde{w}^r)$.

In order to prove differential privacy for our choice of f, we will show that the choice of λ ensures that it is at least as big as the one of theorem Differential Privacy, whose application then proves differential privacy. Since m and m' differ in a single household's entry we can write without loss of generality that $m = (m_{1,\cdot}, \ldots, m_{N,\cdot}, m_{N+1,\cdot}) = (m', m_{N+1,\cdot})$. Since in Theorem Differential Privacy a 1-norm is needed instead of a 2-norm, first the transition is done using the inequality

$$\left\| x \right\|_2 \geq \left\| x \right\|_1 / \sqrt{\mathcal{R}}.$$

Note that this inequality can itself be proven by applying the Cauchy-Schwarz inequality to $\langle \nVdash, |x| \rangle$. Together with the linearity of T_r and W this yields

$$\begin{aligned}
\|f(m) - f(m')\|_1 &\leq \sqrt{\mathcal{R}}\, \|f(m) - f(m')\|_2 && (37)\\
&= \sqrt{\mathcal{R}} \left\| \sum_{i=1}^{N+1} T_r\left(W(m_{i,\cdot})\right) - \sum_{i=1}^{N} T_r\left(W(m_{i,\cdot})\right) \right\|_2 && (38)\\
&= \sqrt{\mathcal{R}} \left\| T_r\left(W \left(\sum_{i=1}^{N+1} m_{i,\cdot} - \sum_{i=1}^{N} m_{i,\cdot} \right)\right)\right\|_2 && (39)\\
&= \sqrt{\mathcal{R}}\, \left\|T_r\left(W\left(m_{N+1,\cdot}\right)\right)\right\|_2 && (40)
\end{aligned}$$

The restriction to a smaller resolution is equivalent to setting the higher resolutions to zero. Therefore the restriction T_r decreases the norm while the wavelet transformation does not change it due to our restriction of using only transformations with the tightness property

$$\begin{aligned}
\|f(m) - f(m')\|_1 &\leq \sqrt{\mathcal{R}}\, \left\|W\left(m_{N+1,\cdot}\right)\right\|_2 && (41)\\
&= \sqrt{\mathcal{R}}\, \left\|m_{N+1,\cdot}\right\|_2 . && (42)
\end{aligned}$$

Finally, this equation directly yields

$$\begin{aligned}
\lambda &= \frac{\sqrt{\mathcal{R}}}{\epsilon} \operatorname*{argmax}_{m_{N+1,\cdot}} \left\|m_{N+1,\cdot}\right\|_2 && (43)\\
&\geq \frac{1}{\epsilon} \operatorname*{argmax}_{m,m'} \|f(m) - f(m')\|_1 && (44)\\
&= \frac{S_1(f)}{\epsilon}. && (45)
\end{aligned}$$

Thus, theorem Differential Privacy can be applied and proves differential privacy for f. □

Evaluation

In this section we evaluate the proposed PETs in combination with MRA with respect to the security features, cost and complexity and real-world applicability.

Applications

In this paper, multi-resolution secure aggregation has been introduced for both, a single aggregator and multiple aggregators. Given the building blocks of the additive homomorphic Paillier cryptosystem, masking and differential privacy in combination with the wavelet transform, a scheme can be constructed that allows to encrypt different resolutions with different keys while maintaining a single bitstream. In Section Application Scenario, three typical application scenarios for smart grid have been introduced: (i) Settlement and Profiling; (ii) Network Monitoring; and (iii) Billing.

Settlement and profiling require data in a comparably low resolution, but spatially aggregated over a number of households for determining forecasts and training models. Network monitoring, by contrast, still works with the aggregate, but requires a much higher temporal resolution. Both, homomorphic encryption and masking can be used for aggregating over a number of smart meters, e.g., from households connected to the same substation or participants belonging to the same consumption group (residential/industrial). By adding the ability to selectively decrypt a subset of multiple resolutions, the same aggregated bitstream, but with different keys, can be provided to both, the utility provider for forecasts and model training and the network operator for network monitoring. This reduces the overhead for managing and transferring various bitstreams simultaneously to distinct recipients. While network monitoring might require very high accuracy (e.g., voltage levels must remain in a narrow band), for settlement and profiling, customer privacy can be even enhanced by adding differential privacy in order to prevent the detection of the presence of a single household in the aggregate, while at the same time providing a certain guaranteed ϵ-differential privacy-level. While differential privacy is a compelling approach due to this property, it is not suitable for applications that require the exact aggregate.

Billing and dynamic pricing will require data at high resolutions and generally not aggregated. Further, differential privacy is not a desired property for billing. However, if in a dynamic pricing scenario, data in different granularity is needed over the day (e.g., a stable night tariff and more dynamic tariffs at noon), the multi-resolution approach allows to dynamically adjust the level of granularity of the provided meter data.

Security Analysis

In this section a security analysis of the proposed PETs is conducted. We consider a honest-but-curious adversarial model, meaning the adversary follows the protocols but tries to gain additional information.

Secure Signal Processing: For MRA with secure signal processing, an honest-but-curious aggregator will not learn any information. Due to the additive homomorphic property of the cryptosystem, even at collector nodes all operations are performed in the encrypted domain and the aggregator can only decrypt the sum.

Masking with Single Aggregator: For a total number of smart meters $N > 1$ and a single aggregator $M = 1$ the masking scheme preserves full privacy in terms of spatial resolution and it preserves full privacy with respect to temporal resolution. Given exactly one aggregator $M = 1$, the *basic approach* for multi-resolution masking is applied. A receives a set of N masked values and a single key share. By combining both, A can calculate the sum at a particular resolution. For spatial aggregation privacy is preserved by the scheme proposed by Kursawe et al. [9], i.e., the individual measurements are masked and the random shares cancel each other out upon summation. The temporal resolution is limited by the resolution of the key share. The privacy preserving feature of this approach has been discussed in detail in Section Basic Approach. Section Proof of Correctness includes the proof of correctness for the masking approach in combination with wavelets.

Masking with Multiple Aggregators: For a total number of smart meters $N > 1$ and a total number of aggregators $M = 2$, the *multiple aggregators approach* for multi-resolution masking is applied. Each aggregator $A_k, k = \{1, 2\}$ receives a set of N masked values, an individual key share key_1^r and key_2^r, respectively and the sum of the keys of all other aggregators $\sum_{i \neq 1} \text{key}_i^r = \text{key}_2^r$ and $\sum_{i \neq 2} \text{key}_i^r = \text{key}_1^r$, respectively. Therefore, each aggregator additionally holds the other aggregators share in full resolution and thus privacy in terms of temporal aggregation is not given anymore.

For a setting with $M > 2$ privacy is preserved, as the key shares of all other aggregators are hidden in the sum. Therefore, the above limitation for $M = 2$ does not apply, since $\sum_{i \neq k} \text{key}_i^r \neq \text{key}_i^r$ for any $i, k \in \{1 \dots M\}$. This means that holding all keys except for one does not yield a valid key. This assures that the aggregator cannot learn anything beyond the resolution of the key, which is formally shown in Section Multiple Aggregators.

Differential Privacy: It is a proven property of differential privacy, that the aggregator has no means to decrease privacy of the *aggregated* signal by any kind of postprocessing. If differential privacy would be combined with wavelets only, the aggregator could, however, inspect a *single* smart meter's consumption profile. A single profile is only protected by Gamma-distributed noise which does not provide differential privacy. Therefore, the mechanism achieving differential privacy must include a way to protect the summation operation by using a secure aggregation scheme, e.g., as described in Section Multi-resolution Secure Aggregation.

Space considerations

When using homomorphic encryption for aggregation, the modulus n determines the amount of data that can be stored within one encrypted packet. Let's denote the number of bits needed to represent a wavelet coefficient by $\bar{m}$ and the number of values used for the wavelet transform by T. The sum of two coefficients will take up $\bar{m}+1$ bits of space. More generally, the sum of u coefficients requires $\lceil \log_2(u) + \bar{m} \rceil$ bits. If encrypting each wavelet coefficient individually, i.e., using T encryptions, the modulus of n bits allows to sum up a total of $u \leq 2^{n-\bar{m}}$ wavelet coefficients, since $n = \lceil \log_2(u) + \bar{m} \rceil$, i.e., u represents the total number of wavelet coefficients from household measurement values that can be aggregated for a given modulus.

Setting $T = 256$, $n = 1024$ and $\bar{m} = 16$, this allows for the aggregation of more than $2 \cdot 10^{303}$ households but requires 256 encryptions. However, in practice such large aggregation groups are not needed. Instead of encrypting each coefficient individually, the available space of n bits can be exploited better when using data packing [36]. Values are shifted to a certain bit range, such that a number of values can be packed within a single encryption. The available space is therefore split into p packets of fixed size n', i.e., $p = \frac{n}{n'}$. This allows for $n' = \lceil \log_2(u') + \bar{m} \rceil$ a number of $u' \leq 2^{n'-\bar{m}}$ wavelet coefficients per packet and a total of $u' \cdot p$ wavelet coefficients of household measurement values per encryption. This results in only $T' = \frac{T}{p}$ encryptions.

Setting $n = 1024$, $\bar{m} = 16$ and $n' = 32$, this results in $p = 32$ packets and still allows to aggregate up to 65536 households, but with only a fraction ($T' = 8$) of the number of encryption operations compared to the above approach where each coefficient is encrypted separately. In practice, these values have to be chosen with respect to the number of households that will be aggregated.

Cost and complexity

Both methods, secure signal processing with the Paillier cryptosystem[1] and masking have been implemented together with the wavelet transform. The proof of concept implementation is built on Oracle Java 1.8 and tested on a HP Z230 workstation with 8 GB RAM and an Intel Xeon CPU (3.4 GHz). Results are shown in Table 2: Each value represents the execution time for a single load curve consisting of 96 values for the wavelet transform combined with different encryption settings and masking, averaged over 400 load curves with 100 encryptions/additions for masking each (acquisition of the load curve and key generation as well as precalculating the masking shares are not considered in the timing results). WAV denotes the wavelet transform only, without any encryption or masking applied. AES denotes the wavelet transform followed by encryption with the symmetric AES cipher with a 256 bit key for each subband. HYB denotes hybrid encryption, which adds RSA 2048 bit public key encryption of the AES keys with a different public key for each subband. PAI-n denotes Pailler encryption with a module of n bits and a different key for each subband. For practical applications and according to [37] a module of at least 2048 bits should be chosen. Finally, MA denotes the masking of values.

It can be seen that by using a lifting implementation the computational overhead of the wavelet transformation is negligible compared to the encryption step. Homomorphic encryption comes at the cost of a significant increase in computational overhead compared to conventional encryption. The results show that the computational demands grow exponentially with the module size. Although the used implementation of Paillier is not optimized and could be improved considerably in terms of efficiency, it is clear that running homomorphic encryption on smart meter hardware will provide a challenge: While AES encryption only takes 1.25 ms, for the used (non-optimized) implementation, Paillier encryption with a 2048 bit module of a load curve with 96 values takes approximately 52 seconds. Further, it can be seen that masking is highly efficient in terms of computation time when compared to encryption, however, at the cost of losing the entire aggregate when a single smart meter fails.

Conclusion and Outlook

The approaches proposed in this paper allow to get both, temporal and spatial aggregation by combining the wavelet transform with homomorphic encryption, masking and differential privacy. In this paper it has been shown, that it is possible to combine homomorphic encryption, masking and differential privacy with the Haar wavelet transform. Furthermore, a protocol has been sketched for addressing different aggregators with different resolutions of the measured time series while still maintaining a certain level of privacy. For masking, future work will focus on a scheme that is more error-resilient and still yields the correct result even if a subset of smart meters fail.

Competing interests

The authors declare that they have no competing interests.

[1] Building on the implementation by Kun Liu http://www.csee.umbc.edu/~kunliu1/research/Paillier.html

Author's contributions

This paper was written by Fabian Knirsch (40%), Günther Eibl (30%) and Dominik Engel (30%). The detailed contributions are as follows: The Abstract was written by Fabian Knirsch (100%) The Introduction was written by Fabian Knirsch (80%) and Dominik Engel (20%). The Background section was written by Fabian Knirsch (40%), Dominik Engel (40%) and Günther Eibl (20%). The Multi-resolution Secure Aggregation Section was written by Dominik Engel (100%). The Multi-resolution Masking Section was written by Fabian Knirsch (100%). The Multi-resolution Differential Privacy Section was written by Günther Eibl (100%). The Evaluation Section was written by Fabian Knirsch (60%), Dominik Engel (30%) and Günther Eibl (10%). Conclusion and Outlook were written by Fabian Knirsch (100%). The figures were created by Fabian Knirsch (100%). Measurements for homomorphic encryption were performed by Dominik Engel (100%).

Acknowledgments

The financial support by the Austrian Federal Ministry of Science, Research and Economy and the Austrian National Foundation for Research, Technology and Development is gratefully acknowledged. Funding by the Federal State of Salzburg is gratefully acknowledged.

Author details

[1]Salzburg University of Applied Sciences, Josef Ressel Center for User-Centric Smart Grid Privacy, Security and Control, Urstein Süd 1, 5412 Puch bei Hallein, Austria. [2]University of Salzburg, Department of Computer Sciences, Jakob-Haringer-Str. 2, 5020 Salzburg, Austria.

References

1. European Commission: Cost-benefit analyses & state of play of smart metering deployment in the EU-27. Technical report (2014). http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52014SC0189&from=EN
2. McKenna, E., Richardson, I., Thomson, M.: Smart meter data: Balancing consumer privacy concerns with legitimate applications. Energy Policy **41**, 807–814 (2012)
3. Hart, G.W.: Nonintrusive appliance load monitoring. Proceedings of the IEEE **80**(12), 1870–1891 (1992)
4. Molina-Markham, A., Shenoy, P., Fu, K., Cecchet, E., Irwin, D.: Private memoirs of a smart meter. In: Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building. BuildSys '10, pp. 61–66. ACM, New York, NY, USA (2010)
5. Lisovich, M., Mulligan, D., Wicker, S.: Inferring Personal Information from Demand-Response Systems. IEEE Security & Privacy **8**(1), 11–20 (2010)
6. Knirsch, F., Engel, D., Frincu, M., Prasanna, V.: Model Based Assessment for Balancing Privacy Requirements and Operational Capabilities in the Smart Grid. In: Proceedings of the 6th Conference on Innovative Smart Grid Technologies (ISGT2015), pp. 1–5. Innovative Smart Grid Technologies Conference (ISGT), 2015 IEEE Power & Energy Society, Washington, D.C., USA (2015)
7. Eibl, G., Engel, D.: Influence of Data Granularity on Nonintrusive Appliance Load Monitoring. In: Proceedings of the Second ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec '14), pp. 147–151. ACM, Salzburg, Austria (2014)
8. Eibl, G., Engel, D.: Influence of Data Granularity on Smart Meter Privacy. IEEE Transactions on Smart Grid **6**(2), 930–939 (2015)
9. Kursawe, K., Danezis, G., Kohlweiss, M.: Privacy-friendly aggregation for the smart grid. In: Privacy Enhanced Technology Symposium, pp. 175–191 (2011)
10. Gomez Marmol, F., Sorge, C., Petrlic, R., Ugus, O., Westhoff, D., Martinez Perez, G.: Privacy-enhanced architecture for smart metering. International Journal of Information Security **12**(2), 67–82 (2013)
11. Li, F., Luo, B., Liu, P.: Secure Information Aggregation for Smart Grids Using Homomorphic Encryption. In: Proceedings of First IEEE International Conference on Smart Grid Communications, Gaithersburg, Maryland, USA, pp. 327–332 (2010)
12. Erkin, Z., Tsudik, G.: Private computation of spatial and temporal power consumption with smart meters. In: Proceedings of the 10th International Conference on Applied Cryptography and Network Security. ACNS'12, pp. 561–577. Springer, Berlin, Heidelberg (2012)
13. Rastogi, V., Suman, N.: Differentially private aggregation of distributed time-series with transformation and encryption. In: Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data. (2010)
14. Danezis, G., Kohlweiss, M., Rial, A.: Differentially Private Billing with Rebates vol. 6958 LNCS, pp. 148–162. Springer, Berlin, Heidelberg (2011)
15. Garcia, F., Jacobs, B.: Privacy-Friendly Energy-Metering via Homomorphic Encryption. In: Cuellar, J., Lopez, J., Barthe, G., Pretschner, A. (eds.) Security and Trust Management. Lecture Notes in Computer Science, vol. 6710, pp. 226–238. Springer, Berlin Heidelberg (2011)
16. Li, F., Luo, B.: Preserving data integrity for smart grid data aggregation. In: Third International Conference on Smart Grid Communications (SmartGridComm) 2012, pp. 366–371. IEEE, Tainan (2012)
17. Erkin, Z., Troncoso-pastoriza, J.R., Lagendijk, R.L., Perez-Gonzalez, F.: Privacy-preserving data aggregation in smart metering systems: an overview. IEEE Signal Processing Magazine **30**(2), 75–86 (2013)
18. Biselli, A., Franz, E., Coutinho, M.P.: Protection of Consumer Data in the Smart Grid Compliant with the German Smart Metering Guideline. In: Proceedings of the First ACM Workshop on Smart Energy Grid Security. SEGS '13, pp. 41–52. ACM, New York, NY, USA (2013)
19. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: Theory of Cryptography, pp. 265–284. Springer, Berlin Heidelberg (2006)
20. Acs, G., Castelluccia, C.: I have a DREAM! (DiffeRentially privatE smArt Metering). In: Proc. Information Hiding Conference, pp. 118–132 (2011)

21. Shi, E., Chow, R., Chan, T.-h.H., Song, D., Rieffel, E.: Privacy-preserving aggregation of time-series data. In: Proc. NDSS Symposium 2011 (2011)
22. Efthymiou, C., Kalogridis, G.: Smart Grid Privacy via Anonymization of Smart Metering Data. In: Proceedings of First IEEE International Conference on Smart Grid Communications, Gaithersburg, Maryland, USA, pp. 238–243 (2010)
23. Engel, D.: Wavelet-based Load Profile Representation for Smart Meter Privacy. In: Proc. IEEE PES Innovative Smart Grid Technologies (ISGT'13), Washington, D.C., USA, pp. 1–6 (2013)
24. Peer, C.D., Engel, D., Wicker, S.B.: Hierarchical key management for multi-resolution load data representation. In: 2014 IEEE International Conference on Smart Grid Communications (SmartGridComm), pp. 926–932. IEEE, Venice, Italy (2014)
25. Engel, D., Eibl, G.: Wavelet-Based Multiresolution Smart Meter Privacy. IEEE Transactions on Smart Grid **PP**(99), 1–12 (2016)
26. Engel, D., Eibl, G.: Multi-Resolution Load Curve Representation with Privacy-preserving Aggregation. In: Proceedings of IEEE Innovative Smart Grid Technologies (ISGT) 2013, pp. 1–5. IEEE, Copenhagen, Denmark (2013)
27. Jawurek, M., Johns, M., Kerschbaum, F.: Plug-in privacy for smart metering billing. In: Privacy Enhancing Technologies (PETS), pp. 192–210 (2011)
28. Erkin, Z.: Private Data Aggregation with Groups for Smart Grids in a Dynamic Setting using CRT. In: 2015 IEEE International Workshop on Information Forensics and Security (WIFS). IEEE, Rome, Italy (2015)
29. National Institute of Standards and Technology (NIST): Specification for the Advanced Encryption Standard (AES) (2001)
30. ITU-T: Recommendation ITU-T X.509 – Information technology – Open Systems Interconnection – The Directory: Public-key and attribute certificate frameworks (2012)
31. Daubechies, I., Sweldens, W.: Factoring Wavelet Transforms into Lifting Steps. J. Fourier Anal. Appl. **4**(3), 247–269 (1998)
32. Paillier, P.: Public-Key Cryptosystems Based on Composite Degree Residuosity Classes. In: Stern, J. (ed.) Advances in Cryptology — EUROCRYPT '99: International Conference on the Theory and Application of Cryptographic Techniques Prague, Czech Republic, May 2–6, 1999 Proceedings. Lecture Notes in Computer Science, vol. 1592, pp. 223–238. Springer, Berlin, Heidelberg (1999)
33. Defend, B., Kursawe, K.: Implementation of privacy-friendly aggregation for the smart grid. In: Proceedings of the First ACM Workshop on Smart Energy Grid Security - SEGS '13, pp. 65–74 (2013)
34. Diffie, W., Hellman, M.: New Directions in Cryptography. IEEE Transactions on Information Theory **22**(6), 644–654 (1976)
35. Kotz, S., Kozubowski, T.J., Krzysztof, P.: The Laplace Distribution and Generalizations. Birkhäuser Basel, Basel (2001)
36. Erkin, Z., Veugen, T., Toft, T., Lagendijk, R.L.: Generating Private Recommendations Efficiently Using Homomorphic Encryption and Data Packing. IEEE Transactions on Information Forensics and Security **7**(3), 1053–1066 (2012)
37. Barker, E., Barker, W., Burr, W., Polk, W., Smid, M., Division, C.S.: NIST 800-57: Computer Security. NIST (2012)

Figures

Figure 1 Examples of two different topologies (star, left and tree, right). In the star topology the data concentrator (DC) collects measurement values from the smart meters (SM) and forwards the aggregated values to the aggregators (A). In the tree topology measuring values are aggregated in a hop-by-hop manner by the smart meters.

Figure 2 MRA aggregation scheme for secure aggregation with homomorphic encryption and multiple aggregators. The component "DP" representing the addition of Gamma distributed noise is only needed if additionally differential privacy wants to be achieved. In this figure $w_i := W(m_i)$ denotes the wavelet transform of m_i.

Figure 3 MRA aggregation scheme for masking and multiple aggregators. In this figure $w_i := W(m_i)$ denotes the wavelet transform of m_i and $\tilde{w}_i := W(m_i) + s_i$ denotes the masked values of w_i.

Tables

j/i	1	2	3	$\dots$	N
1		$-H(K_{2,1}\|\|t)$	$-H(K_{3,1}\|\|t)$	$\dots$	$-H(K_{N,1}\|\|t)$
2	$+H(K_{1,2}\|\|t)$		$-H(K_{3,2}\|\|t)$	$\dots$	$-H(K_{N,2}\|\|t)$
3	$+H(K_{1,3}\|\|t)$	$+H(K_{2,3}\|\|t)$		$\dots$	$-H(K_{N,3}\|\|t)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
N	$+H(K_{1,N}\|\|t)$	$+H(K_{2,N}\|\|t)$	$+H(K_{3,N}\|\|t)$	$\dots$	
$\sum_j$	$s_{1,t}$	$s_{2,t}$	$s_{3,t}$	$\dots$	$s_{N,t}$

Table 1 In the method proposed by [9], shares cancel each other out pairwise, since $s_i + s_{N-i} = 0$. The columns correspond to N smart meters, the last row is the share created for each smart meter for one point in time t.

	WAV	AES	HYB	PAI-2048	PAI-4096	MA
t	< 0.001	0.07	0.7	$5,219$	$38,700$	< 0.001
σ	< 0.001	0.02	0.01	25.4	51	< 0.001

Table 2 Execution time t in milliseconds and standard deviation σ for transforming/encrypting/masking a single load curve (average over 400 load curves with 100 encryptions each)

3.7 KNIRSCH15B

- F. Knirsch, D. Engel, C. Neureiter, M. Frincu, and V. Prasanna. Model-driven privacy assessment in the smart grid. In *Proceedings of the 1st International Conference on Information Systems Security and Privacy (ICISSP)*, pages 173–181, Feb 2015. Best Paper Award.

Model-driven Privacy Assessment in the Smart Grid

Fabian Knirsch[1], Dominik Engel[1], Cristian Neureiter[1], Marc Frincu[2] and Viktor Prasanna[2]
[1]*Josef Ressel Center for User-Centric Smart Grid Privacy, Security and Control,*
Salzburg University of Applied Sciences, Urstein Sued 1, A–5412 Puch/Salzburg, Austria
[2]*Ming-Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, U.S.A.*
{*fabian.knirsch, dominik.engel, christian.neureiter*}*@en-trust.at,* {*frincu, prasanna*}*@usc.edu*

Keywords: Smart Grid, Privacy, Model-based, Assessment.

Abstract: In a smart grid, data and information are transported, transmitted, stored, and processed with various stakeholders having to cooperate effectively. Furthermore, personal data is the key to many smart grid applications and therefore privacy impacts have to be taken into account. For an effective smart grid, well integrated solutions are crucial and for achieving a high degree of customer acceptance, privacy should already be considered at design time of the system. To assist system engineers in early design phase, frameworks for the automated privacy evaluation of use cases are important. For evaluation, use cases for services and software architectures need to be formally captured in a standardized and commonly understood manner. In order to ensure this common understanding for all kinds of stakeholders, reference models have recently been developed. In this paper we present a model-driven approach for the automated assessment of such services and software architectures in the smart grid that builds on the standardized reference models. The focus of qualitative and quantitative evaluation is on privacy. For evaluation, the framework draws on use cases from the University of Southern California microgrid.

1 INTRODUCTION

In a smart grid a number of stakeholders (actors) have to cooperate effectively. Interoperability has to be assured on many layers, ranging from high level business cases to low level network communication. Data and information is sent from one actor to another in order to ensure effective communication. Furthermore, the exchange of vast amounts of data is crucial for many smart grid applications, such as demand response (DR) or electric vehicle charging (Cavoukian et al., 2010), (Langer et al., 2013). However, this data is also related to individuals and privacy issues are an upcoming concern (McDaniel and McLaughlin, 2009), (Simmhan et al., 2011a). Especially the combination of data, e.g., meter values and preferences for DR can exploit serious privacy threats such as the prediction of personal habits. In system engineering, privacy is a cross-cutting concern that has to be taken into account throughout the entire development life-cycle, which is also referred to as *privacy by design* (Cavoukian et al., 2010).

Model-driven privacy assessment is especially useful when applied in software engineering. In (Boehm, 2006), the author thoroughly investigates the phases in software engineering and the expected costs for error correction and change requests. Costs double with every phase and once an application or a service is delivered, the additional adding of crosscutting concerns such as privacy is tied to enormous costs. As a result, design time privacy assessment is preferred in early phases of the software engineering process. Therefore, a framework is needed to (i) model the system, including high-level use cases and concrete components and communication flows; and (ii) to assess the system's privacy impact using expert knowledge from the domain. Related work in the domain of automated assessments in the smart grid mainly focuses on security aspects and is not primarily concerned with privacy and the modeling in adherence to reference architectures.

In this paper we address these issues and present an approach for the model-driven assessment of privacy for smart grid applications. The framework proposed in this paper is designed to assist system engineers to evaluate use cases in the smart grid in an early design phase. For evaluation only meta-information is used and no concrete data is needed. We use Data Flow Graphs (DFG) to formally define use cases according to a standardized smart grid reference architecture. The assessment is based on an ontology driven approach taking into account expert knowledge from various domains, including customer views on privacy as well as system engineering concerns. The output is a set of threats and a quantitative analysis of risks, i.e., a number indicating the strength of that threat. To evaluate the system we draw on insights from the University of Southern California microgrid. The primary contributions of this paper are (i) the use of DFGs to model use cases in the smart grid; (ii) the usage of DFGs for a quantitative privacy assessment; and (iii) the use of an ontology driven approach to capture domain knowledge.

The remainder of this paper is structured as follows: In Section 2 related work in the area of smart grid reference architectures, privacy evaluation and automated assessment tools is presented. In Section 3 the architecture of the proposed framework and its components are described. This includes the concept of DFGs for modeling use cases in the smart grid, the principal design of the ontology and the mapping of data flow graphs to the ontology, the methodology for defining threat patterns and finally, how these patterns are matched to use cases. The framework is evaluated with a set of representative use cases in Section 4. Section 5 summarizes this paper and gives an outlook to further work in this area.

2 RELATED WORK

In this section related work in the field of smart grid reference architectures, privacy evaluation and assessment as well as automated assessment tools are presented. Often, privacy and security are used interchangeably. For the purpose of this paper we refer to privacy as legally accessing data but not using it for the intended purpose. Security, by contrast, would involve the illegal acquisition of data. In both cases, the well established and widely understood terminology from security assessment is used, i.e., *threat*, *attacker*, *vulnerability* and *countermeasure*.

2.1 Reference Models

Stakeholders in the smart grid come from historically different areas, including electrical engineering, computer science and economics. To ensure interoperability and to foster a common understanding, standardization organizations are rolling out reference models and road maps. In the US the NIST Framework and Roadmap for Smart Grid Interoperability Standards (National Institute of Standards and Technology, 2012) and in the EU the Smart Grid Reference Architecture (CEN, Cenelec and ETSI, 2012b) were published. The European Smart Grid Architecture Model (SGAM) is based on the NIST Framework, but extends the model to better meet European requirements, such as distributed energy resources. In this paper we investigate use cases from the US. In particular we are focusing on use cases from the University of Southern California microgrid and we thoroughly discuss a typical DR use case. Investigations have, however, shown that for the purpose of this project all use cases from the US can be directly mapped to the European SGAM without the loss of information. Therefore we propose the utilization of the SGAM for two reasons: (i) the SGAM builds on the NIST model and allows to capture both, use cases from the US and the EU; and (ii) with the SGAM Toolbox (Dänekas et al., 2014) present a framework for modeling use cases based on the SGAM; in that way formally modeled use cases are the input for the evaluation.

2.2 Privacy

Privacy (and security) issues in the smart grid are addressed by standards in the US (National Institute of Standards and Technology, 2010) and the EU (CEN, Cenelec and ETSI, 2012a). Privacy, in specific, has no clear definition. According to a thorough analysis in (Wicker and Schrader, 2011), privacy can be defined as the right of an individual's control over personal information. More formally this is defined by (Barker et al., 2009) in a four dimensional privacy taxonomy. The dimensions are *purpose*, *visibility*, *granularity* and *retention*. The *purpose* dimension refers to the intended use of data, i.e., what personal information is released for. The purpose ranges from single, a specific use only, to any. *Visibility* refers to who has permitted access. The range is from owner to all/world. *Granularity* describes to what extent information is detailed. The *retention* dimension finally is the period for storage of data. In any case, privacy is assured if all these dimensions are communicated clearly and fully disclosed to data owners and the compliance to the principles is governed. Hence, data is collected and processed for the intended purpose only, and the degree of visibility, granularity and retention is at the necessary minimum.

2.3 Assessment Tools

To measure the degree to which systems adhere to privacy requirements, approaches for automated qualitative assessments (resulting in statements of possible privacy impacts due to privacy critical actions or relationships) and quantitative assessments (resulting in a numeric value that determines the risk of privacy impacts) exist.

In (Ahmed et al., 2007), the authors present an approach towards ontology based risk assessment. The authors propose three ontologies, the *user environment ontology* capturing where users are working, i.e., software and hardware, the *project ontology* capturing concepts of project management, i.e., work packages and tasks and the *attack ontology* capturing possible attacks, e.g., non-authorized data access, virus distribution or spam emails. For a risk assessment, attacks (defined in the attack ontology) are matched with information available from the other ontologies. For a quantitative assessment, the annual loss expectancy is calculated by combining a set of harmful outcomes and the expected impact of such an outcome with the frequency of that outcome. The approach presented by Ahmed et al. is designed for security issues and does not explicitly cover privacy assessments.

In (Kost et al., 2011) and (Kost and Freytag, 2012) an ontology driven approach for privacy evaluation is presented. The aim of these papers is to integrate privacy in the design process. High-level privacy statements are matched to system specifications and implementation details. The proposed *privacy by design* process includes the following phases: identification of high-level privacy requirements, translation of abstract privacy requirements to formal privacy descriptions, realization of the requirements and modeling of the system and analyzing the system by matching formal privacy requirements to the formal system model. Contrary to our work this approach is not focused on use cases in the smart grid and therefore does not model systems based on a standardized reference architecture.

A workflow oriented security assessment is presented in (Chen et al., 2013). This approach is not based on ontologies but on argument graphs. The presented framework uses *security goal*, *workflow and system description*, *attacker model* and *evidence* as an input. This information is aggregated in a discriminative set of argument graphs, each taking into account additional input. Nodes in the graph are aggregated using boolean expressions and the output is a quantitative assessment of the system. Instead of focusing on workflow analysis using graphs, we model systems as a whole in adherence to the standardized reference architecture using an ontology driven approach to integrate expert knowledge.

A considerably broader approach for an assessment tool that incorporates both, the balancing of privacy requirements and operational capabilities is presented in (Knirsch et al., 2015). This work presents a graph based approach that allows the modeling of systems with respect to the operational requirements of certain nodes (e.g. metering at a certain frequency) and the impact of privacy restrictions on subsequent nodes. The authors further present an optimum balancing algorithm, i.e. to what extent restrictions gained from privacy enhancing technologies and the necessary operational requirements can be combined. However, this needs sufficient information on how privacy is impacted by certain use cases which is provided by this work.

3 ARCHITECTURE

This section is dedicated to an architectural overview as well as a detailed discussion of the components. Figure 1 shows the principal components of the proposed architecture, including input and output. For a privacy assessment, the framework accepts two inputs, a use case UC modeled as a DFG in adherence to the SGAM and a set of threat patterns T. In order to qualitatively analyze this input the use case is mapped to individuals – i.e., instances of classes – of an ontology (sometimes referred to as the *assertion box, ABox* (Shearer et al., 2008)). The corresponding class model (sometimes referred to as the *terminological box, TBox* (Shearer et al., 2008)) is based on the SGAM. This qualitative analysis provides explicit and implicit information about the elements from the DFG: actors, components, information objects and their interrelation. The results of the qualitative assessment are the input for the subsequent quantitative analysis. The output of that analysis is finally a class c from a set of classes C that the use case is assigned to. A threat pattern t is used to describe potential threats, where $t \in T$ and a class c represents a subset of threats T^*. A class c describes how threat patterns and the qualitative results are combined, which is presented as a threat matrix as an output. Note that the terminology *threat matrix* is borrowed from security analysis and that the output is not a matrix in the mathematical sense. A threat matrix compares a set of threats and the risk for these threats. Formally, the classifier is defined as Assign UC to c_i if $t \in T_i^*, \forall t \in T, 1 \leq i \leq \{C\}$. A threat exploits a set of vulnerabilities and is mitigated by a set of countermeasures. Each threat pattern can be evaluated for itself or multiple patterns are combined to classes of threats. A vulnerability is any kind of privacy impact for any kind of stakeholder or actor. Threats are evaluated using the attack vector model which is adapted from security analysis and defined in detail later in this paper. In general, an attack is feasible, if given (i) an attacker; (ii) a privacy asset; and (iii) the resources to perform the

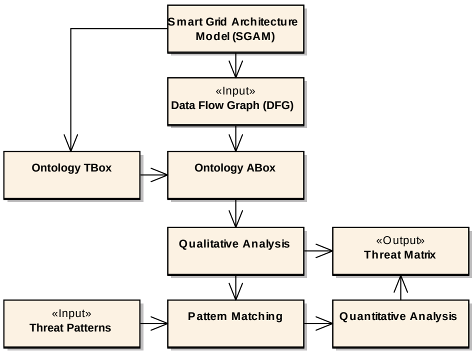


Figure 1: Architecture overview showing input, output, components and principal information flows of the framework.

attack. Hence, a receiver or collector of privacy critical data items is potentially able to access these assets and to use them in a way not corresponding to the original purpose. This is formally represented as $\langle \text{data access}, \text{privacy asset}, \text{attack resources} \rangle$.

3.1 Data Flow Graphs

In order to qualitatively and quantitatively assess the privacy impact of a use case a formalization is crucial. In this section we introduce the concept of Data Flow Graphs (DFG) for the smart grid based on a model-driven design approach originally presented in (Dänekas et al., 2014) and (Neureiter et al., 2013). DFGs formally capture all aspects of use cases in the smart grid in adherence to the SGAM. They contain high-level business cases as well as detailed views of a system's characteristics such as encryption and protocols. DFGs are a powerful tool as they allow both, easy modeling and full adherence to the reference architecture. Furthermore, in the graph relationships between actors, as well as the transported information objects (IO) are modeled. Nodes in a graph represent business actors, system actors or components and edges represent data flows annotated with IOs. In accordance to the standard (CEN, Cenelec and ETSI, 2012b), DFGs consist of the following five layers:

1. Business Layer. In a DFG this layer is a high level description of the business case. Business actors, their common business goal and their business requirements are modeled.
2. Function Layer. The function layer details the business case by mapping business actors to system actors and by dividing the high level business goals in use cases and steps.
3. Information Layer. This layer describes information flows in detail. System actors communicate to each other through IOs. IOs are characterized by describing information attributes on a meta-level. An IO is one of the key data used for classification and is discussed in greater detail below.
4. Communication Layer. The communication layer is a more detailed view on communication taking into account network and protocol specifications.
5. Component Layer. In a DFG this layer contains concrete components. Therefore system actors are mapped to components and devices.

Each layer is a directed graph. Both, nodes and edges can have attributes. The semantics, however, are varying. For instance, where attributed edges in the business layer describe a business case, in the information layer concrete meta-data of communication flows are captured. Even though implicitly covered in the model presented above, for automated evaluation we introduce two additional layers: Between business and function layer we include the *Business Actor to System Actor Mapping* and between communication and component layer the *System Actor to Component Mapping*. This allows to capture the complexity of use cases on different levels while still maintaining the cross-layer relationship between high-level business actors and their representation as components. These layers are directed graphs as well, with edges indicating the mapping. The mapping defines a one to many relationship from business actors to system actors and from system actors to components. In the European Smart Grid Reference Architecture with the SGAM Methodology an approach for mapping use cases to the reference model is suggested. DFGs build on this methodology focusing on actors and their interrelation. An implementation for modeling DFGs in UML is available as the *SGAM Toolbox*[1]. Data Flow Graphs contain explicit information (what is modeled) and implicit information (what can be concluded). Conclusions are drawn using ontology reasoning.

3.2 Ontology Design

The ontology driven approach for classification has been chosen for two main reasons: (i) ontologies are powerful for capturing domain knowledge explicitly; and (ii) through logic reasoning (Shearer et al., 2008) ontologies are a source for implicit knowledge. The power of ontologies to formally capture knowledge and how to draw conclusions is discussed in (Guarino et al., 2009). The power of reasoning for gaining

[1] http://www.en-trust.at/downloads/sgam-toolbox/

additional, implicit knowledge can easily be outlined with two examples: In a DFG, information objects may be sent from an actor *A* to an actor *B* and from there to another actor *C*. This is explicitly modeled in the DFG. A reasoner in an appropriate ontology, however, may conclude directly the transitivity, hence that actor *A* in fact sends information to actor *C*. Another example is concerned with compositions of data. An information object I_1 may contain sensitive data and it may be used by an actor *D* to compose another information object I_2 that is sent to a collecting actor *E*. It is not explicitly modeled in the DFG, but it can be concluded by the reasoner, that *E* receives an information object which is of type sensitive data since I_2 is a composition of I_1. The ontology we propose here is designed to capture all aspects of a DFG. The ontology is modeled in OWL[2] and class expressions are stated in Manchester Syntax[3]. Therefore, all components available for modeling DFGs are represented either directly or as an abstraction in the ontology (referred to as the *TBox*). The DFG is represented in the ontology as a set of individuals (referred to as the *ABox*). Figure 2 depicts the principal classes and relationships of the ontology and therefore the most relevant concepts for mapping a DFG to the ontology. This view shows the main classes and relationships for illustration purposes only; our current ontology comprises more than 60 classes, data properties and object properties. Crucial concepts represented immediately, include which actor sends or receives which data and IO and how these IOs are composed. Furthermore, a set of pre-classifiers is defined to determine implicit knowledge.

These classifiers are OWL classes using an equivalent class expression in Manchester Syntax. For instance, to determine if some aggregation consists of direct personal data, the following expression is used: `Data and isAggregationOf some DirectPersonalData`. To determine the multiplicity of the sending actor and if the data is a composition sent by many of such actors, more elaborate expressions can be phrased: `Data and isSentBy some Actor and Multiplicity value "n" and isCompositionOfMany some Data`.

3.3 Threat Patterns

In this paper we evaluate the privacy impact on customers, thus we identified the following list of typical high-level threats based on literature reviews (Cavoukian et al., 2010), (Langer et al., 2013), (Simmhan et al., 2011a). These threats have been modified in order to be more representative for the use cases from the University of Southern California microgrid that are investigated in this paper. Subsequently, IOs that may cause these threats are determined.

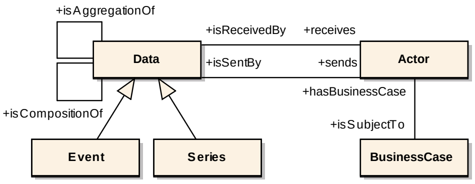


Figure 2: Principal components of the ontology, showing a subset of the relationships between actor and data.

Customer Presence at Home. This privacy concern is discussed in (Cavoukian et al., 2010). To potentially determine a person's presence at home, some device in the customer premises is needed. This device collects data at a certain frequency, high enough to have a resolution that allows to draw conclusions on the energy usage of specific devices. Furthermore, data collected from that device needs to be sent to another actor (i.e., a utility). At the utility an individual or a system needs to have access to the data in an appropriate resolution. Since we always assume that data is accessed legally, we do not focus on unallowed data access. Additionally, the total delay of the data transmission is of relevance. If data is collected and transmitted in almost real time the presence at home can be determined immediately. If data is available with a delay only, the analysis of past events and predictions might be possible. If this information is published, an attacker might exploit this vulnerability in order to break in the house.

Tracking Customer Position. This threat is especially interesting for electric vehicle charging. Assuming the customer has some identification towards the charging station, at least the location, a timestamp and the amount of energy consumed will be recorded for billing. Depending on the design of the infrastructure only little information will be sent to the operator or a very detailed profile of the customer is maintained. Here, the multiplicity of the actors is crucial and the fact that different actors have access to the same data. Attacks for this threat are described in (Langer et al., 2013), e.g., using information for targeted ads, for tracking movements to certain places or to infer the income based on recharges.

3.4 Pattern Matching

Actual classification is done in the pattern matching process. For each actor in the DFG and the ontology,

[2] http://www.w3.org/TR/owl-features/

[3] http://www.w3.org/TR/owl2-manchester-syntax/

respectively, the attack vector is determined, i.e., to which resources does an actor have access and what is the effort. If that shows feasible matching this is seen as a threat. It can be retrieved immediately from the ontology if an actor has access to a certain IO. This is done by evaluating actor and data object properties and by incorporating information from the pre-classifiers. Furthermore, relationships on the business layer and data properties such as encryption are taken into account. The following, discriminative set of classifiers is used to determine potential threats: first, for each information object the data provider and the data collector are determined (according to the terminology defined in (Barker et al., 2009)) and it is assessed who has access to the data. This yields a list of three-tuples in the form $\langle$information object (IO), data provider (DP), data collector (DC)$\rangle$. Then it is determined if an information object either contains sensitive or direct personal data (according to the terminology defined in (The European Parliament and the Council, 1995)). This yields another three-tuple in the form $\langle$information object (IO), sensitive (S), direct personal (DP)$\rangle$. Finally it is determined if the attacker has actual data access, yielding one more three-tuples in the form $\langle$information object (IO), data collector (DC), access (A)$\rangle$. Data access depends on the relationship of actors, on data resolution, retention and encryption. Matching these tuples to each other results in the components of the attack vector, recalling $\langle$data access, privacy asset, attack resources$\rangle$ yields $\langle\langle \mathrm{IO}, \mathrm{DP}, \mathrm{DC}\rangle, \langle \mathrm{IO}, \mathrm{S}, \mathrm{DP}\rangle, \langle \mathrm{IO}, \mathrm{DC}, \mathrm{A}\rangle\rangle$. An exemplary attack vector for a DR use case where DR preferences are sent to the utility is $\langle\langle \text{DR preferences}, \text{customer}, \text{utility}\rangle, \langle \text{DR preferences}, \text{false}, \text{false}\rangle, \langle \text{DR preferences}, \text{utility}, \text{true}\rangle\rangle$. This already provides thorough qualitative analysis. It is possible to determine which actor can potentially threaten the privacy of another actor. It is even possible to conclude how and where this might happen. However, for a quantitative assessment the risk for a particular threat is calculated. While a qualitative assessment is useful in supporting detailed system design decisions and evaluation, for a very first outline of the overall system characteristics, a quantitative value is much more expressive. Further, providing a numeric value for the system's privacy impact helps to easily compare and contrast proposed designs.

Risk is calculated as the product of the *probability of occurrence* (PO) and the *expected loss* (EL). For the set T^* a number of patterns $t_{v,1} \dots t_{v,N}$ and $t_{c,1} \dots t_{c,M}$, respectively is defined. A pattern therefore contains a set of conditions for vulnerabilities $t_{v,i}$ and countermeasures $t_{c,i}$. Conditions are SPARQL ASK queries[4] that return either *true* or *false* if the pattern applies or not. For brevity, t'_v denotes the number of vulnerabilities that apply, t'_c the number of countermeasures that apply and t_v and t_c denote the total number of vulnerabilities and countermeasures, respectively. In this paper we propose the following approach for determining values for the probability of occurrence $PO(t'_v, t'_c)$ and the expected loss $EL(t'_v, t'_c)$: $PO(t'_v, t'_c)$ is determined by defining a plane that satisfies the following conditions: $PO(t'_v = t_v, t'_c = 0) = 1$, $PO(t'_v = 0, t'_c = t_c) = 0$ and $PO(t'_v = 0, t'_c = 0) = \frac{1}{2}$. This yields $PO(t'_v, t'_c) = \frac{1}{2}(\frac{t'_v}{t_v} - \frac{t'_c}{t_c} + 1)$. A linear model is chosen due to its simplicity and might be extended by more complex approaches in future. A condition that is of type *vulnerability* increases $EL(t'_v, t'_c)$, a condition of type *countermeasure* decreases $EL(t'_v, t'_c)$. The value of $EL(t'_v, t'_c)$ is defined in the pattern. Risk R is finally defined by $R = PO(t'_v, t'_c)EL(t'_v, t'_c)$.

To feed in the results gained from the qualitative analysis, certain variables in the query can be bound to instances. For example, given the following fraction of a query (where `usc` denotes the namespace prefix for actors and IOs in the University of Southern California microgrid) `$io usc:isSentBy ?systemactor . ?systemactor usc:isRealizationOf ?businessactor . ?businessactor a usc:BusinessActor` to determine if *some* information object is sent by *some* business actor. It is now possible to bind the variable `$io` to a concrete value as determined in the qualitative assessment, e.g., `$io` ← `InformationObject.CustomerName`. This allows to assess a particular impact on a particular information object or component/actor based on the previously calculated attack vectors.

We developed generic patterns for *typical* threats, i.e., such as the ones mentioned above. The framework is, however, not limited to this set of patterns and allows the definition of an arbitrary number of additional patterns to meet the individual needs of the application scenario. The output of the framework is a threat matrix contrasting the results from the qualitative analysis and from the quantitative risk assessment. For a UC, a threat matrix contains the attack vector and the assigned risk for the determined class c.

For illustrative purposes, the following listing shows an example pattern for *customer presence at home*. This includes the vulnerability *device in customer premises* (exemplary assigned an EL of 4) and

[4] http://www.w3.org/TR/sparql11-query/

the countermeasure *aggregation of data from multiple customers* (exemplary assigned an EL of -6).

```
<Pattern name="customer presence at home">
  <Vulnerability
    name="device in customer premises">
    <EL>4</EL>
    <Condition>
      ?device x:isRealizationOf $ba .
      $ba a x:BusinessActor .
      ?device x:Zone
      "Customer Premises"^^xsd:string
    </Condition>
  </Vulnerability>
  <Countermeasure
    name="aggregation of data from multiple
      customers">
    <EL>-6</EL>
    <Condition>
      $io x:manyAreAggregatedBy ?io2 .
      ?io2 x:isReceivedBy ?ba1 .
      $io x:isRecevivedBy ?ba2
      FILTER (?ba1 != ?ba2)
    </Condition>
  </Countermeasure>
</Pattern>
```

4 EVALUATION

For evaluating the framework new, previously unused use cases are applied. The set of threat patterns and their impact on privacy is based on the aforementioned literature reviews. We are therefore using a representative set of use cases describing typical applications in the smart grid. This includes, but is not limited to, smart metering, electric vehicle charging and DR. In this section a real-life use case from the University of Southern California microgrid is evaluated as an example. This use case has been chosen as it is (i) simple enough to verify results based on literature reviews; and (ii) complex enough to have an interesting combination of actors and information flows. We are focusing on a DR scenario similar to the one described in (Simmhan et al., 2011b). This scenario is outlined in Figure 3. A customer interested in DR creates an online profile stating on which DR actions the customer is interested to participate (e.g., turning down air condition). When the utilities want to curtail load with DR, a customer whose profile fits the current requirements is sent a text message to, e.g., turn down the air condition. This message is acknowledged by the customer and the utility further reads the meter values to track actual power reduction. Besides the data flows mentioned, this further involves the storing of the profile and the past behavior of the customer for a more accurate prediction. For modeling this use case as a DFG, the following actors and IOs are identified. Evaluation is performed with a prototypical implementation that uses DFGs and threat patterns as an input and produces a threat matrix as an output.

4.1 Data Flow Graph

Actors. Business actors are the *user* and the *utility*. The user is mapped to the system actors *smart meter*, *device* and *portal*. DR requests are sent to the user device (e.g., a cell phone) and the user's DR preferences are set in the portal (e.g., a web service). The smart meter is used to measure actual curtailment. The utility is mapped to a *DR repository*, containing preferences for each user and past behavior, to a *prediction unit* predicting DR requests based on the preferences and a *control unit* to meter user feedback and actual curtailment.

Information Objects. Cross-domain/zone information flows include user preferences sent to the utilities, DR requests sent to the user from the utility and both, the user acknowledge/decline and the meter values sent back to the utility. Information flows within the utilities' premises are from the DR repository to the prediction unit and from the control unit to the DR repository. Given the threat patterns introduced in Section 3, we use our framework to determine the privacy impact of this use case which provides the following results.

Customer Presence at Home. The qualitative analysis shows that in the DR repository of the utility information about both, past customer behavior and customer data is brought together, i.e., direct personal data is composed with a detailed history of a person's actions. Furthermore, the customer's acknowledge/decline and the measured curtailment reveal if a customer (i) responded to the DR request; and (ii) actually participated in DR; both is a indication for the presence at home. For this threat we identified four vulnerabilities (device in customer premises, collecting data at a certain frequency, receiver has access to data, data retention is unlimited) and one countermeasure (aggregation of data from multiple customers), resulting in a *PO* of 0.9, an *EL* of 11.5 and a risk value of 10.35.

Tracking Customer Position. In our case, this threat might apply in two different scenarios: First, this threat is immediate if the acknowledge/decline response to DR requests contains the customer position (e.g., if sent by a cell phone or other mobile device). This does not only show the customers past and present position, but also if the customer is able to remotely control devices in his premises. Second, when the customer is represented by an additional

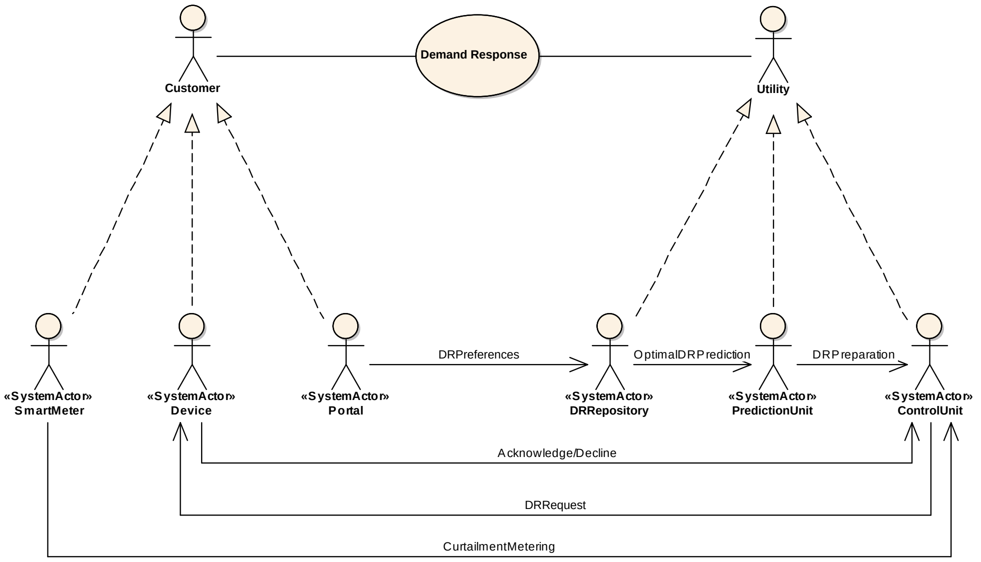


Figure 3: Outline of the DR use case that is discussed for evaluation.

component *electric vehicle charging station*. Assuming that DR requests are also sent with respect to the charging behavior. Based on the amount of energy the customer is willing to DR it might be possible to estimate the consumption of the electric vehicle and subsequently the traveled distance. For this threat we identified two vulnerabilities (composition of location and timestamp, different actors have access to the same data) and one countermeasure (aggregation of data from multiple customers), resulting in a *PO* of 0.66, an *EL* of 5 and a risk value of 3.33.

The mode-driven assessment of the DR use case has shown that the risk of tracking customer position is low compared to the risk of determining customer presence at home. This result stems from the fact that there apply a number of vulnerabilities with high expected loss value, hence a device in the customer premises, data collected at a certain frequency, receiver has access to data and unlimited data retention.

5 CONCLUSION AND FUTURE WORK

In this paper we introduced a framework for the model-driven privacy assessment in the smart grid. The framework builds on an ontology driven approach matching threat patterns to use cases that are modeled in adherence to standardized reference architectures. The approach presented here builds on meta-information and high-level data flows. It has been shown how to utilize this framework to successfully assess the privacy impact on use cases in early design time. Exemplary threats and exemplary use cases draw on insights from the University of Southern California microgrid. Future work will include an evaluation of the systems ability to generalize to arbitrary kinds of threats in the smart grid. Furthermore the system will be extended to serve as a policy decision point for system developers and customers in a smart grid IT infrastructure.

ACKNOWLEDGEMENT

The financial support of the Josef Ressel Center by the Austrian Federal Ministry of Economy, Family and Youth and the Austrian National Foundation for Research, Technology and Development is gratefully acknowledged. Funding by the Austrian Marshall Plan Foundation is gratefully acknowledged. The authors would like to thank Norbert Egger for his contribution to the prototypical implementation.

This material is based upon work supported by the United States Department of Energy under Award Number number DE-OE0000192, and the Los Angeles Department of Water and Power (LA DWP). The views and opinions of authors expressed herein do not

necessarily state or reflect those of the United States Government or any agency thereof, the LA DWP, nor any of their employees.

REFERENCES

Ahmed, M., Anjomshoaa, A., Nguyen, T., and Tjoa, A. (2007). Towards an ontology-based risk assessment in collaborative environment using the semanticlife. In *Proceedings of the The Second International Conference on Availability, Reliability and Security*, ARES 07, pages 400–407, Washington, DC, USA. IEEE Computer Society.

Barker, K., Askari, M., Banerjee, M., Ghazinour, K., Mackas, B., Majedi, M., Pun, S., and Williams, A. (2009). A data privacy taxonomy. In *Proceedings of the 26th British National Conference on Databases: Dataspace: The Final Frontier*, BNCOD 26, pages 42–54, Berlin, Heidelberg. Springer.

Boehm, B. (2006). A view of 20th and 21st century software engineering. In *Proceedings of the 28th International Conference on Software Engineering*, ICSE 2006, pages 12–29, New York, NY, USA. ACM.

Cavoukian, A., Polonetsky, J., and Wolf, C. (2010). Smartprivacy for the smart grid: embedding privacy into the design of electricity conservation. *Identity in the Information Society*, 3(2):275–294.

CEN, Cenelec and ETSI (2012a). Smart Grid Information Security. Technical report, CEN/Cenelec/ETSI Smart Grid Coordination Group Std.

CEN, Cenelec and ETSI (2012b). Smart Grid Reference Architecture. Technical report, CEN/Cenelec/ETSI Smart Grid Coordination Group Std.

Chen, B., Kalbarczyk, Z., Nicol, D., Sanders, W., Tan, R., Temple, W., Tippenhauer, N., Vu, A., and Yau, D. (2013). Go with the flow: Toward workflow-oriented security assessment. In *Proceedings of New Security Paradigm Workshop (NSPW)*, Banff, Canada.

Dänekas, C., Neureiter, C., Rohjans, S., Uslar, M., and Engel, D. (2014). Towards a model-driven-architecture process for smart grid projects. In Benghozi, P.-J., Krob, D., Lonjon, A., and Panetto, H., editors, *Digital Enterprise Design & Management*, volume 261 of *Advances in Intelligent Systems and Computing*, pages 47–58. Springer International Publishing.

Guarino, N., Oberle, D., and Staab, S. (2009). *What Is an Ontology?* Handbook on Ontologies – International Handbooks on Information Systems. Springer, 2nd edition.

Knirsch, F., Engel, D., Frincu, M., and Prasanna, V. (2015). Model-based assessment for balancing privacy requirements and operational capabilities in the smart grid. In *Proceedings of the 6th Conference on Innovative Smart Grid Technologies (ISGT2015)*. to appear.

Kost, M. and Freytag, J.-C. (2012). Privacy analysis using ontologies. In *CODASPY '12 Proceedings of the second ACM conference on Data and Application Security and Privacy*, pages 205–2016, San Antonio, Texas, USA. ACM.

Kost, M., Freytag, J.-C., Kargl, F., and Kung, A. (2011). Privacy verification using ontologies. In *Proceedings of the 2011 Sixth International Conference on Availability, Reliability and Security*, ARES '11, pages 627–632, Washington, DC, USA. IEEE Computer Society.

Langer, L., Skopik, F., Kienesberger, G., and Li, Q. (2013). Privacy issues of smart e-mobility. In *Industrial Electronics Society, IECON 2013 - 39th Annual Conference of the IEEE*, pages 6682–6687.

McDaniel, P. and McLaughlin, S. (2009). Security and privacy challenges in the smart grid. *Security Privacy, IEEE*, 7(3):75–77.

National Institute of Standards and Technology (2010). Guidelines for smart grid cyber security: Vol. 2, privacy and the smart grid. Technical report, The Smart Grid Interoperability Panel – Cyber Security Working Group.

National Institute of Standards and Technology (2012). NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 2.0. Technical Report NIST Special Publication 1108R2, National Institute of Standards and Technology.

Neureiter, C., Eibl, G., Veichtlbauer, A., and Engel, D. (2013). Towards a framework for engineering smart-grid-speficic privacy requirements. In *Proc. IEEE IECON 2013, Special Session on Energy Informatics*, Vienna, Austria. IEEE.

Shearer, R., Motik, B., and Horrocks, I. (2008). Hermit: A highly-efficient owl reasoner. In Dolbear, C., Ruttenberg, A., and Sattler, U., editors, *OWLED*, volume 432 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Simmhan, Y., Kumbhare, A., Cao, B., and Prasanna, V. (2011a). An analysis of security and privacy issues in smart grid software architectures on clouds. In *IEEE International Conference on Cloud Computing (CLOUD), 2011*, pages 582–589. IEEE.

Simmhan, Y., Zhou, Q., and Prasanna, V. (2011b). Semantic information integration for smart grid applications. In Kim, J. H. and Lee, M. J., editors, *Green IT: Technologies and Applications*, pages 361–380. Springer, Berlin Heidelberg, Germany.

The European Parliament and the Council (1995). Official Journal L 281, 23/11/1995 P. 0031 - 0050 – Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995. Online.

Wicker, S. and Schrader, D. (2011). Privacy-aware design principles for information networks. *Proceedings of the IEEE*, 99(2):330–350.

3.8 ENGEL11A

- D. Engel. Conditional access smart meter privacy based on multi-resolution wavelet analysis. In *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies*, pages 45:1–45:5, New York, NY, USA, 2011. ACM.

Conditional Access Smart Meter Privacy Based on Multi-Resolution Wavelet Analysis

Dominik Engel
Salzburg University of Applied Sciences
Urstein Sued 1
Salzburg, Austria
dominik.engel@fh-salzburg.ac.at

ABSTRACT

Smart Metering is an important component of Smart Grids. Detailed load profiles are available through smart metering at a high resolution. Load profiles allow inferring detailed information on the end user by non-intrusive load monitoring. Therefore, these load profiles need to be regarded as sensitive data, and treated with security and privacy in mind. We propose a method that allows conditional access to different resolution levels of the load data, allowing access on a "need-to-know" basis only. For this purpose, a multiresolution representation of the load data is created using the simple Haar wavelet transform. Securing the portions of the wavelet representation pertaining to each resolution with a unique key allows to implement conditional access for smart meter data.

Categories and Subject Descriptors

E.3 [**Data**]: Data Encryption

1. INTRODUCTION

Smart Grids have recently come to the focus of attention of a large number of research programmes. The powerful combination of communication technology with electrical grids leads the energy infrastructure into a new paradigm.

However, this transition is not without challenges. There are security concerns: An overview of current research in smart grid security can be found in [3], a comprehensive proposal for securing smart grid infrastructure is given by [16]. Privacy is another critical area, related to security. Smart meters form a core component of smart grids. Each of these devices contains a processor, as well as storage and communication facilities and is capable of transmitting detailed usage statistics to the energy provider. While the exact granularity of the transmitted data is not finally specified, and may differ by country, it seems likely that the interval between single measurements will lie between 1 and 30 minutes. The availability of per-customer load profiles on such a fine granularity raises privacy concerns [17, 15, 8, 1]. A comprehensive discussion of such privacy concerns can be found in [14, 13].

There are a number of approaches for matching appliance signatures to load profiles to determine which appliances were used at what time and for how long, e.g. [6, 9, 12]. This type of method is termed "non-intrusive load monitoring" (NILM) or "non-intrusive appliance load monitoring" (NALM). Detection based on NILM is remarkably accurate: [14] reports over 90% accuracy in detecting presence and sleep cycle intervals. The results reported in [13] show that "even with relatively unsophisticated hardware and data-extraction algorithms, some information about occupant behavior can be estimated with a high degree of accuracy". [10] uses genetic algorithms for identification and report flawless identification for up to 10 types of appliances.

In [14] results of a collaboration between researchers from law and engineering are reported. The authors argue that there "exist strong motivations for entities involved in law enforcement, advertising, and criminal enterprises to collect and repurpose power consumption data." For example, burglars could use the data to determine occupancy patterns of houses to time break-ins, or NILM may be used to identify specific brands of appliances, which can then be used for targeted advertising. In summary, while there are many useful applications of smart meter data, such as energy saving and tailor-made energy rates, the privacy of this kind of data needs to be secured.

In this paper, we discuss the utility of wavelet multiresolution analysis (MRA) to afford privacy to smart meter load profiles. We evaluate MRA-based privacy based on multi-layer conditional access. Access to resolutions can be defined individually, ranging from a low frequency dataset over multiple refinement datasets to the highest resolution representation. Conditional access to the different resolutions has the advantage that a information can be made accessible on a "need-to-know" principle.

Thereby it is possible to use data at lower resolutions, i.e. a (daily) average, for accounting purposes with the energy provider, while the high resolution data remain secured from access. Access to these higher resolutions can selectively be granted to third parties, e.g. to serve as input to energy-saving tools, which match load signatures to determine appliances with high power consumption. The contribution of the proposed approach is to provide both security and privacy to the level specified as needed. For evaluation we use data from a test project conducted by Salzburg AG, an Austrian energy provider.

ISABEL '11, October 26-29, Barcelona, Spain
Copyright 2011 ACM ISBN 978-1-4503-0913-4/11/10 ...$10.00.

The rest of this paper is organized as follows: Section 2 gives an overview of related approaches for smart meter privacy and security. Section 3 introduces the application of multi-resolution wavelet analysis for the representation of smart meter data. Section 4 details the proposed conditional access encryption scheme. Section 5 evaluates security and privacy provided by the proposed scheme, in comparison to other schemes. In Section 6 the complexity and computational demands of the proposed are discussed. Section 7 concludes and gives an outlook on future research.

2. RELATED WORK

There are a number of proposals for secure transmission of smart meter data, e.g. [19] proposes a secure multi-cast protocol that automatically derives group memberships and verifies configuration performance; [2] proposes a security protocol for smart meter aggregation that provides hop-by-hop security, while still providing end-to-end security. In principle, the approach proposed here is compatible with many of the basic methods used in secure transmission, such as aggregation along a spanning tree. The advantage of the approach proposed here, is the possibility to determine the available granularity of the data along the spanning tree.

There are suggestions for security methods that also preserve privacy. [11] proposes to employ homomorphic encryption for smart meter data. Specifically a Paillier cryptosystem is used, which supports the additive homomorphic property, to enable aggregation of smart meter data in the encrypted domain. The approach is evaluated in an honest-but-curios adversary model.

There are some approaches that propose to install rechargeable batteries at the end user home to mask the load profile [7, 18]. While in theory this is an effective approach, the practical applicability remains questionable due to the high costs of installing batteries and the energy loss introduced by using a battery buffer.

The authors of [14] use their NILM algorithm on power consumption data from a two-week experiment to infer individual information and usage patterns. This endeavor is highly successful on high resolution data (15 second intervals). The authors then investigate the performance of their algorithm in the face of downsampled data, i.e. decreased resolution. They report that the algorithm performance “degrades quite gracefully”. “Meaningful estimates” are possible even for 20 minute intervals. The observed graceful degradation for load signature detection supports the representation of the smart meter data in different resolutions, as each resolution will exhibit different detection properties.

[5] proposes an anonymization scheme that is based on two different resolutions. This scheme employs a trusted third party escrow service. Two smart meter data sets are generated: One of low frequency that can be used for billing purposes, and one of high frequency that allows further investigation. Only for the low frequency data set a mapping to the corresponding user is provided, foremost to allow the energy provider to invoice the user for the consumed energy. The high frequency data set is not attributable to a user, at least not by the energy provider.

In this scenario, two IDs have to be used by the smart meter hardware: HFID, or High-Frequency ID, which remains anonymous, and LFID, or Low-Frequency ID, which is attributable. Each message that is communicated from the smart meter, needs to include one of these IDs. In order to keep the mapping HFID and smart meter hidden from the energy provider, the authors propose an agnostic data aggregator (operated by the third party escrow service) that collects high frequency profiles from a number of smart meters. The validity of the used IDs is verified by the third party escrow service. Low-frequency data is forwarded including the LFID. HFID data is aggregated across multiple smart meters and forwarded without the corresponding HFIDs.

As [3] points out, there are a number of security issues with the approach proposed by [5]. As with all systems that rely on a trusted third party, a compromise of this party is devastating. Furthermore, an attacker could try to match high-frequency data to low-frequency data (and thereby to a unique user) by collecting high-frequency information over time, summing up this information and matching it to observed low-frequency data. A third issue is the fact that both LFID and HFID need to be stored in hardware which is in the sphere of control of the users. This may lead to the possibility to tamper the hardware and manipulate the IDs.

3. WAVELET MULTI-RESOLUTION-ANALYSIS OF SMART METER DATA

In [5] representations of the smart meter data in two resolutions are created, a low frequency and a high frequency resolution. In principle, this is a valid approach in terms of granting each party access only to the information that is needed for the processing demands of this party.

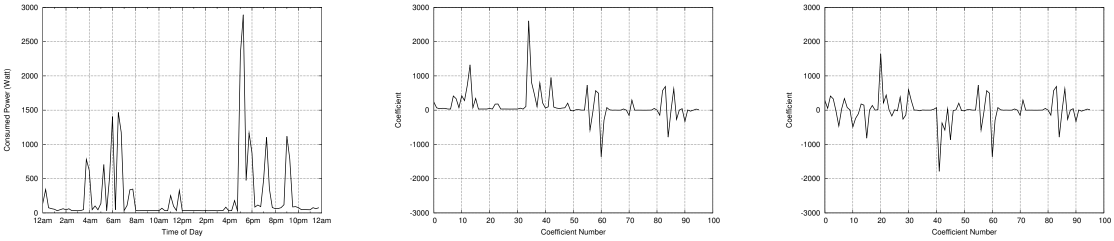
As a first stage of our proposed approach, we generalize this idea. Instead of creating only two variants that are separate, we generate a hierarchy of resolutions in an integrated representation. The best suited tool for this endeavor is the discrete wavelet transform (DWT).

A suitable wavelet transform is applied to the original smart meter data. This leads to a low frequency and a high frequency band. As an example, we use the Haar wavelet transform, which in principle only consists of calculating averages and differences.

To obtain a multi-resolution representation of the original signal, the wavelet analysis step is recursively applied to the low pass subband, up to a maximum level m. The low-frequency portion at each step presents the data at a resolution with half the number of samples of the next higher resolution. The resolution level corresponding to the highest decomposition depth m is referred to as R_0, and has a size of 2^{-m} samples.

The synthesis step of the inverse wavelet transform starts with R_0. Each next higher resolution can be obtained by applying the inverse wavelet transform to the low-pass subband (i.e., the lower resolution) and the corresponding high-pass subband. In this way, m further resolutions can be obtained. Typically, the resulting subbands are represented in a single bitstream.

Figure 1 shows an example of the wavelet decomposition of a smart meter signal. Figure 1(a) shows the original signal. Figure 1(b) shows the first level of decomposition into a low-pass and a high-pass subband. Figure 1(c) shows a wavelet decomposition with a maximum decomposition depth $m = 5$. In this example, the interval between smart metering values is 15 minutes. Therefore, 96 values are produced within 24 hours. The lowest resolution of a level 5 wavelet decomposition in this case contains 3 values, which roughly correspond to the average load during



(a) Original Data (b) Level 1 Decomposition (c) Level 5 Decomposition

Figure 1: Example for Wavelet Decomposition of Smart Meter Data

the first, second and third eight-hour interval, respectively, of the 24-hour interval. By combining subbands L_0 and H_1 the resolution R_1 can be obtained, which contains double the number of samples as R_0. Depending on the needed granularity, higher resolutions can be reconstructed by applying the inverse wavelet transform, up to the highest resolution R_5, which contains the original 96 values.

To implement multi-resolution analysis in a manner that is suitable for smart metering devices, wavelet lifting [4] is the best approach. This view on the wavelet transform factors wavelet filters into lifting steps, which for many filters encompass only basic operations.

For the Haar wavelet, the lifting steps can easily be implemented by hardware with lowest computing power, such as the smart meter environment. Furthermore the transformation is lossless, and the aggregation is equivalent to subsampling.

4. CONDITIONAL ACCESS FOR SMART METER DATA

The idea of conditional access stems from the context of multimedia entertainment data. Entertainment content usually exists in various resolutions (e.g. mobile content, standard definition, high definition), which may be priced differently. A multi-resolution representation of the multimedia data allows the efficient representation of the resolutions in a single bitstream. This is an advantage as only one version of the bitstream needs to be handled and transmitted. Conditional access allows end-users to pay only for the resolutions they are interesting in. For example, the owner of a standard definition television has no need to pay for the high-definition version of the content. Through conditional access only the bitstream portion relevant for the desired resolution is decrypted, the rest of the bitstream is ignored.

We propose to use the conditional access paradigm for smart-metering data in multi-resolution representation. Each high-pass subband is encrypted with a different key. The lowest resolution is left unencrypted. The whole datastream can be transmitted over the Smart Grid communications network. Access to the different resolutions is thereby only granted to parties that hold the needed keys, as illustrated by Figure 2. The lowest resolution remains accessible to the network provider at all times to enable invoicing. In the example above this is done by leaving R_0 unencrypted (which in principle mirrors the situation in current energy networks). Alternatively, R_0 could also be encrypted with an appropriate key that allows access to the energy provider.

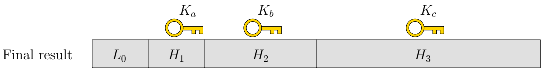


Figure 2: Final Bitstream Produced by Smart Meter

This scheme allows flexible control by the end-user how access is granted to smart meter data. For example, the energy network operator may be granted access to the lowest resolution for billing purposes, but the end-user may not be willing to provide any detailed usage statistics to the energy network operator. Through the conditional access scheme, end-users can also decide to use the services of third parties, by encrypting the relevant portions of their metering data so that the third party offering a service can access the needed data. Such services could include load signature matching to identify appliances with high energy consumption. Aggregation services through a trusted third party are another possible type of service.

The proposed scheme also enables the relaying of data. For example, the data needed by a third party analysis lab can be forwarded in encrypted form by an aggregator, or even the energy network operator. While the parties forwarding the data on route to the destination can access the low-frequency data, there is no access to the higher resolutions.

Of course, the proposed scheme requires the smart meter hardware to provide functionality for wavelet lifting and encryption, and to support the manual or automatic setting of encryption keys for the higher resolutions.

5. SECURITY

Following [11], we assume an “honest-but-curious” adversary model, i.e. all parties are assumed to follow the protocol (“honest”), but within this limitation try to infer as much information about the other participating parties as possible (“curious”). In this setting, the proposed scheme is very successful, as detailed data of the high resolutions remains protected from unauthorized access, including eavesdroppers. There are no methods to infer the higher resolutions by using the information from lower resolutions. Therefore, the higher resolutions are secure, provided that state-of-the art cryptographic ciphers are used.

If the adversary model is changed to malicious, an adversary will also tamper the data to change aggregation or billing data. Under these settings, the homomorphic encryption model proposed by [11] suffers from the fact that all homomorphic encryption schemes are malleable and allows in-transit tampering. The scheme proposed here is widely agnostic to the used encryption scheme, and supports traditional, non-homomorphic encryption, which do not suffer from this malleability. Integrity checks can be added to prevent in-transit tampering. However, as in the scheme proposed by [11], the proposed scheme does not prevent tampering of smart meter data at the point of origin, i.e. if a tampered smart meter produces fake data, this is not recognized. To prevent this kind of tampering, the proposed scheme needs to be combined with trusted computing (e.g. [16]).

Other than the scheme proposed by [11], the method proposed here does not allow a non-trusted aggregator. While the proposed method allows to control the level of detail an aggregator is allowed to process, there is no restriction on readability on this level of data for the aggregator. The recipients of the data, i.e. the key holders for a certain resolution, need to be trusted with the data at the given resolution.

Regarding successful privacy protection of the higher resolutions, the proposed scheme has an advantage over the scheme proposed by [5]. As stated by [3], the privacy afforded by the scheme proposed by [5] may be compromised by data aggregation through an eavesdropper. The eavesdropper could link LFID and HFID by summing up high frequency data that he observes. Such an attack is not possible in the scheme proposed here, as all high frequency data is transmitted in encrypted form.

6. COMPUTATIONAL COMPLEXITY

As discussed above, implementing the wavelet transform as lifting steps is computationally inexpensive. Generally, the discrete wavelet transform has a complexity of $\mathcal{O}(n)$. Due to the simple operations used in the lifting implementation, the transformation part can be realized by inexpensive smart meter hardware.

The computational demands for encryption depends on the used encryption scheme. For standard encryption schemes, efficient hardware implementations exist that can be integrated into the smart meter hardware. Depending on the desired scenario, symmetric encryption alone can be used, or in combination with asymmetric encryption. The latter case is computationally more demanding but benefits from the support for public key infrastructures, such as proposed by [16].

Some overhead is introduced for key management, and potentially for the creation of session keys. Both are computationally inexpensive, and implementations should be easily transferable to smart meter hardware, given that the other tasks above should be implementable on the underlying smart meter hardware.

To explore the complexity of the proposed approach empirically, we use smart meter data from an Austrian energy provider, which was generated by real households over a period of 18 months. In our tests, we use 400 load profiles. The load profiles originate from Siemens smart meter hardware, model TD3510 (3 phase, 100 Amp.). The sampling interval is 15 minutes, i.e. 96 readings a day. Three scenarios were implemented: (i) Symmetric encryption: AES with 128 bit keys, (ii) Asymmetric encryption only: RSA with 2048 bit keys, (iii) Hybrid encryption: 128 bit AES session keys encrypted with 2048 bit RSA keys. In each scenario the following steps are executed: (i) Level 5 wavelet transform of the load profile, (ii)) Generation of different keys to encrypt resolutions R_1 through R_5 (R_0 is left unencrypted), (iii) Encryption of R_1 through R_5, each with a different key.

	WAV	AES	RSA	HYB
Avg. Exec. Time (ms)	35.9	246.4	1366.3	1425.5
Std. Deviation	3.4	17.7	18.9	76.8

Table 1: Execution Times for Test Implementation: Total for 400 load profiles, averaged over 1000 executions

The implementation was done in Java (OpenJDK 64-Bit Server VM). The Java standard implementation of the cryptographic routines were used. The Haar wavelet transform was implemented as lifting steps. No special optimization was performed. The tests were run on a Sun Fire V20z with two AMD Opteron Processors 244 and 8GB of RAM, running 64-bit Ubuntu Linux 10.04.1 with kernel version 2.6.32. We note that the execution environment for smart meter hardware will be extremely restricted and generally not be comparable to this test setup. However, the test setup is suited for obtaining empirical data on the comparative performance of the possible scenarios, and to estimate the computational complexity of the wavelet transform in relation to encryption.

Table 6 shows timing results in milliseconds comparing wavelet transform only (WAV) with AES (128 bit), pure RSA (2048 bit) and hybrid encryption (HYB) using an AES 128-bit session key encrypted with RSA (2048 bit). Due to the small amount of time needed for the transformation and encryption of a single load profile, in order to get valid results, timing was done for a batch of 400 load profiles. Each load profile was subjected to 1000 encryptions in each of the aforementioned categories. In Table 6, each entry corresponds to the total transformation/encryption time of 400 different load profiles, averaged over 1000 encryptions.

It can be seen that indeed the computational demands for the wavelet transform are negligible. On average, the transformation of 400 load profile takes 36 ms. Regarding demand for encryption, as expected, symmetric encryption with AES is the fastest method by far. In the used test setup, this approach outperforms the other two encryption approaches by a factor of 5. In many application scenarios, where the superior key management of public key cryptography is not needed, this advantage will make symmetric encryption a prime candidate. Due to the limited size of the subbands, public key cryptography can be used directly on the load data. For our test setup, all subbands can be encrypted using 2048 bit RSA keys. We compare this approach to a hybrid approach, in which a symmetric session key is encrypted with a public key. For larger plaintext data sets the hybrid approach allows to combine the speed of symmetric encryption with asymmetric key management. It can be observed that the hybrid approach in our scenario is slower than the pure asymmetric approach. This is due to the fact that the load profile subbands are limited in size. Of course, for larger data sets using pure asymmetric encryption is not feasible and the hybrid approach would have to be used. However, it can be rated an advantage that the multi-resolution representation of the load profiles allows the direct application of public key cryptography.

7. CONCLUSION AND OUTLOOK

We have shown that a multi-resolution representation of smart-meter data is a way to balance the need for privacy with the additional functionality introduced by the smart meter load profiles. By using multiple keys to encrypt each resolution separately, the scheme affords end-user control of access to different granularities of the data. Apart from privacy, due to encrypting the higher resolutions, the proposed scheme also implements secure transmission of the load profiles and prevents unauthorized access by eavesdroppers.

The scheme proposed here fits neatly into the larger frameworks proposed to date, such as [16], as it is compatible with other approaches for securing smart grid communication, including authentication, integrity checking, and the integration into smart grid public key infrastructure.

Regarding computational complexity, some overhead is introduced. However, we have shown that the simple Haar wavelet transform, implemented as lifting steps, has very low demands. The computational demands for encryption of the higher resolution subbands are higher, especially if an asymmetric or hybrid approach is chosen. However, communication in the smart grid will have to be secured by cryptographic means, and smart meters are no exception. Therefore, it is likely that smart metering hardware will be required to support encryption. The additional computational overhead for multi-resolution representation and multiple key-handling is acceptable, especially when seen with the background of this requirement for secure communication and authentication.

Future work will focus on the details of integrating the proposed approach into larger smart grid communication frameworks. To support privacy preserving data aggregation by non-trusted data aggregators, we will investigate if, based on the simple operations used in Haar wavelet lifting, the proposed scheme is compatible with the homomorphic encryption approach proposed by [11]. Finally we will explore possible advantages of using more sophisticated wavelets for representing the load profiles.

8. ACKNOWLEDGMENTS

The author would like to thank Salzburg AG for providing test data.

9. REFERENCES

[1] R. Anderson and S. Fuloria. On the security economics of electricity metering. In *Proceddings of the Ninth Workshop on the Economics of Information Security (WEIS 2010)*, June 2010.

[2] A. Bartoli, J. Hernández-Serrano, M. Dohler, A. Kountouris, and D. Barthel. Secure lossless aggregation for smart grid m2m networks. In *Proceedings of First IEEE International Conference on Smart Grid Communications*, pages 333–338, Gaithersburg, Maryland, USA, Oct. 2010.

[3] T. Baumeister. Literature review on smart grid cyber security. Technical report, University of Hawaii at Manoa, Dec. 2010.

[4] I. Daubechies and W. Sweldens. Factoring wavelet transforms into lifting steps. *J. Fourier Anal. Appl.*, 4(3):247–269, 1998.

[5] C. Efthymiou and G. Kalogridis. Smart grid privacy via anonymization of smart metering data. In *Proceedings of First IEEE International Conference on Smart Grid Communications*, pages 238–243, Gaithersburg, Maryland, USA, Oct. 2010.

[6] G. Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12), 1992.

[7] G. Kalogridis, C. Efthymiou, S. Z. Denic, T. A. Lewis, and C. R. Privacy for smart meters: Towards undetectable applicance load signatures. In *Proceedings of First IEEE International Conference on Smart Grid Communications*, pages 232–237, Gaithersburg, Maryland, USA, Oct. 2010.

[8] H. Khurana, M. Hadley, N. Lu, and D. Frincke. Smart-grid security issues. *IEEE Security & Privacy*, 8(1):81–85, 2010.

[9] H. Y. Lam, G. S. K. Fung, and W. K. Lee. A novel method to construct taxonomy of appliances based on load signatures. *IEEE Transactions on Consumer Electronics*, 53(2):653–660, 2007.

[10] S. K. J. Leung, S. H. K. Ng, and W. M. J. Cheng. Identifying appliances using load signatures and genetic algorithms. In *Proceedings International Conference on Electrical Engineering (ICEE*, Hong Kong, July 2007.

[11] F. Li, B. Luo, and P. Liu. Secure information aggregation for smart grids using homomorphic encryption. In *Proceedings of First IEEE International Conference on Smart Grid Communications*, pages 327–332, Gaithersburg, Maryland, USA, Oct. 2010.

[12] J. Liang, S. Ng, G. Kendall, and J. Cheng. Load Signature Study Part I: Basic concept, structure, and methodology. *IEEE Transactions on Power Delivery*, 25(2):551–560, 2010.

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

[14] M. A. Lisovich and S. B. Wicker. Privacy concerns in upcoming residential and commercial demand-response systems. *IEEE Proceedings on Power Systems*, 1(1), 2008.

[15] P. McDaniel and S. McLaughlin. Security and privacy challenges in the smart grid. *IEEE Security Privacy Magazine*, 7(3):75–77, 2009.

[16] A. R. Metke and R. L. Ekl. Security technology for smart grid networks. *IEEE Transactions on Smart Grid*, 1(1):99–107, June 2010.

[17] E. L. Quinn. Privacy and the new energy infrastructure. *Social Science Research Network (SSRN)*, Feb. 2009.

[18] D. Varodayan and A. Khisti. Smart meter privacy using a rechargeable battery: minimizing the rate of information leakage. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2011)*, Prague, Czech Republic, May 2011. to appear.

[19] J. Zhang and C. A. Gunter. Application-aware secure multicast for power-grid communications. In *Proceedings of First IEEE International Conference on Smart Grid Communications*, pages 339–344, Gaithersburg, Maryland, USA, Oct. 2010.

3.9 ENGEL13A

- D. Engel. Wavelet-based load profile representation for smart meter privacy. In *Proceedings IEEE PES Innovative Smart Grid Technologies (ISGT'13)*, pages 1–6, Washington, D.C., USA, Feb. 2013. IEEE.

Wavelet-based Load Profile Representation for Smart Meter Privacy

Dominik Engel
Josef Ressel Center for
User-Centric Smart Grid Privacy, Security and Control
Salzburg University of Applied Sciences
Urstein Sued 1, A–5412 Urstein/Salzburg, Austria
Email: dominik.engel@en-trust.at

***Abstract*—A significant portion of (potential) end-users at this point in time are wary about possible disadvantages of smart grid technologies. A critical issue raised by end-users in various studies is the lack of trust in the level of privacy. Smart metering is the component in the end-user domain around which the most intense debate on privacy revolves, because load profiles are made available at high resolutions. Non-intrusive load monitoring (NILM) techniques allow the analysis of these load profiles to infer user behaviour, such as sleep-wake cycles. We investigate and compare the utility of different variants of the wavelet transform for creating a multi-resolution representation of load profiles. In combination with selective encryption, this multi-resolution representation allows end-users to grant or deny access to different resolutions on a "need-to-know" basis. Access to the different resolutions is thereby only granted to parties holding the needed keys. The whole datastream can be transmitted over the smart grid communications network. The lifting implementation of the wavelet transform has computationally low demands and can be run in embedded environments, e.g. on ARM-based architectures, in acceptable time. The proposed approach is evaluated based on the provided level of security, computational demands and feasibility in an economic sense.**

I. Introduction

The move towards smart grids has spawned a large number of industry initiatives, research programmes and standardization efforts, see e.g. [1] for a current overview. Many of the earlier contributions focused on the smart grid ecosystem at a larger scale, without exploring in detail the ramifications of the move towards smart grid technology for the end-user. More recent programmes increasingly accommodate the user perspective, cf. [1]. Addressing the topic of user acceptance is pointed out as a key issue by almost all authors.

Spreading Smart Grid technologies will be inherently difficult without addressing user concerns and actively managing user acceptance by providing secure methods and demonstrating safety of user data and privacy. Methods for privacy and security will be a critical in establishing end user *trust* and thereby enabling end user participation.

Smart meters form a core component of smart grids. Each of these devices contains a processor, as well as storage and communication facilities and is capable of transmitting detailed load profiles on a daily basis or even in real-time. The exact granularity of the transmitted load profiles is not finally specified, and may differ by country. The intervals between single measurements will lie between a few seconds and 30 minutes.

The availability of data at such fine granularities has raised privacy concerns: Apart from the data needed for regular operation, a number of other information items can be extracted from this data, some of them related to sensitive and personal information on the end user. Especially in the area of individual high resolution load profiles made available by smart meters, severe privacy concerns have been expressed in numerous contributions, e.g. [2]–[6]. In [5] results of a collaboration between researchers from law and engineering are reported. The authors argue that there "exist strong motivations for entities involved in law enforcement, advertising, and criminal enterprises to collect and repurpose power consumption data" [5, p. 1]. For example, burglars could use the data to determine occupancy patterns of houses to time break-ins. Marketing agencies could identify specific brands of appliances used, which could then be used for targeted advertising.

There are a number of approaches for matching appliance signatures to load profiles to determine which appliances were used at what time and for how long, e.g. [7]–[9]. This type of method is termed "non-intrusive load monitoring" (NILM) or "non-intrusive appliance load monitoring" (NALM). Detection based on NILM is remarkably accurate: In [5] over 90% accuracy are reported in detecting presence and sleep cycle intervals. The results show that "personal information can be estimated with a high degree of accuracy, even with relatively unsophisticated hardware and algorithms" [5, p. 2]. The authors of [10] use genetic algorithms for identification and report flawless identification for up to 10 types of appliances. In [11] successful identification of appliances in low resolution load profiles is reported, e.g. 30 minute intervals, with the use of data-mining techniques.

In summary, while there are many useful applications of smart meter data, such as energy saving and tailor-made energy rates, the privacy of this kind of data needs to be secured, even within communication environments secured against unauthorized external access. The authors of [1] make the case for a system in which insiders will access "data in an authorized manner and will only use this data in an *acceptable* manner" [1, p. 8].

In this paper we evaluate wavelet-based multi-resolution representations to secure load profiles and to provide a user-centric privacy approach. In previous work [12], the Haar wavelet was used in a preliminary proof of concept. In this paper, we detail the approach, show that aggregates are preserved and apply the approach in an embedded environment. Furthermore, we investigate the utility of integer-based wavelet filters and compare the filter variants. We provide a detailed evaluation regarding computational demands, and investigate the preservation of aggregates in real-world conditions, as well as the level of security provided and feasibility from an economic perspective.

The rest of this paper is organized as follows: Section II discusses the state of the art as well as prior and related work. The proposed scheme is detailed in Sections III and IV and evaluated in Section V. Section VI summarizes the most important findings and conludes.

II. Related Work

There are two kinds of privacy approaches: regulatory-based and technology-based [1]. An important source for regulatory scenarios and recommendations are the reports of the European Commission Smart Grid Expert Group Two for regulatory recommendations for data safety, data handling and data protection, e.g. [13]. Other sources include Common Criteria for Information Technology Security Evaluation (ISO/EIC 15408) and country-specific recommendations, such as the Federal Office for Information Security (BSI) in Germany, e.g. [14].

In the context of smart grid privacy, there is a number of contributions that deal with technological approaches to end-user privacy in general, for an overview see [15]. In [16] an anonymization scheme that is based on two different resolutions is proposed. This scheme employs a trusted third party escrow service. Two smart meter data sets are generated: One of low frequency that can be used for billing purposes, and one of high frequency that allows further investigation. The authors of [17] propose the anonymization of smart metering readings through the use of aggregation, i.e. high resolution smart meter readings are aggregated at the NAN level and only the aggregate is sent to the utility provider. They introduce two solutions both with and without involvement of trusted third parties. In [18] a scheme that allows to obfuscate smart meter data is proposed that does not affect the performance of overall state estimation. The authors of [19] propose a scheme that trades off interests of utility and users based on lossy source coding. In [20] the use of random sequences in compressed sensing of load profiles to provide privacy and integrity is proposed. The authors of [9] propose a zero-knowledge protocol for privacy enhanced-smart metering. The authors of [21] propose a privacy-preserving protocol for general calculations on meter readings on high resolutions. They use simple cryptography on the meters to certify readings and propose to off-load high-integrity calculations to other user devices. The authors show correctness through cryptographic verification.

Secure transmission of smart meter data is a key topic addressed by many contributions. A security protocol for smart meter aggregation that provides hop-by-hop security, while still providing end-to-end security, is proposed by [22]. In [23], a comprehensive proposal for securing smart grid infrastructure is given, including a proposal for a key infrastructure. The authors of [24] propose a scheme for authentication in the smart grid that is privacy aware. In [25] a secure transport protocol for smart grid data collection in general is presented. The authors of [26] propose a model-based access control system. In [27] a zero-configuration identity-based signcryption scheme for the smart grid is proposed.

Privacy-enabling encryption for smart meter data by the use of homomorphic encryption is suggested by both [28] and [29]. Specifically, a Paillier [30] cryptosystem is used in both contributions, which supports the additive homomorphic property, to enable aggregation of smart meter data in the encrypted domain. The approach suggested by [29] is evaluated in an honest-but-curious adversary model. The system proposed by [31] uses multi-party computation in combination with homomorphic encryption.

The need to deal with multiple resolutions of the available data has been widely acknowledged, e.g. [1], [16]. Furthermore multi-resolution representation can serve to protect privacy while at the same time preserving essential statistics of the underlying data [32]. We have previously proposed the use of the Haar wavelet to create a multi-resolution representation and to use selective encryption to grant conditional access to the individual resolutions on a "need-to-know" principle [12].

III. Multi-resolution Representation of Load Profiles

The basis for both, regulatory-based and technology-based approaches to preserve privacy is detailed knowledge of what information can be extracted with which tools from the available user data. To date, there is little systematic research on this subject in the context of smart grids. In [33] an information theoretic approach to abstract privacy and utility requirements is used. The authors aim at providing a measure for the amount of information leaked, and also for the utility that is retained in the data at different levels of abstraction.

In [9] the information revealed from load profiles at different granularities is investigated. The authors show that with off-the shelf statistical methods detailed information on the behavior of users can be inferred from load profiles without prior knowledge or precomputed appliance signatures. They argue that "the information leaks directly correlate with the time granularity that a meter measures power consumption" [9, p.61] and list a number of privacy-relevant questions that can be answered using load profiles at granularities ranging from hours to seconds.

While a detailed empirical investigation of the exact amount of information that can be extracted from load profiles at each granularity is missing, current results, such as reported by [9], indicate that it is safe to assume an increase in the order of magnitude in detection accuracy each time the number of available samples for a specific time are doubled. In other words, based on existing investigations it seems that classes of detection accuracy can be based on a resolution increases of powers of two.

A representation of load profiles in different resolutions corresponds to these classes of detecting accuracy. The classical wavelet transformation in the lifting implementation is the ideal tool to create integrated, dyadic multi-resolution representations of load profiles. Each resolution contained in the multi-resolution load profile can be tailored to correspond to a class of detection accuracy. Granting access to third party based on this multi-resolution representation allows informed, privacy-aware data exchange to the user.

A. *Wavelet-based Representation*

A suitable wavelet transform is applied to the original load profile. This leads to a low frequency and a high frequency band. To obtain a multi-resolution representation of the original signal, the wavelet analysis step is recursively applied to the low pass subband, up to a maximum level m. The low-frequency portion in each step presents the data at a resolution with half the number of samples of the next higher resolution. The resolution level corresponding to the highest decomposition depth m is referred to as R_0, and has a size of 2^{-m} samples.

The synthesis step of the inverse wavelet transform starts with R_0. Each next higher resolution can be obtained by applying the inverse wavelet transform to the low-pass subband (i.e., the lower resolution) and the corresponding high-pass subband. In this way, m further resolutions can be obtained. The resulting subbands are represented in a single bitstream.

To implement multi-resolution analysis in a manner that is suitable for smart metering devices, wavelet lifting [34] is the best approach. This view on the wavelet transform factors wavelet filters into lifting steps, which for many filters rely on simple operations only.

B. *Haar Wavelet Filter*

The Haar wavelet filter realizes low-pass filtering as averaging of the sample values. The high-pass step is realized by the corresponding differences to allow for lossless reconstruction. Let x_l be the input signal, and s_l and d_l be the low-pass and high-pass output signals, respectively. The lifting steps for the forward transform with the Haar wavelet can be written as follows [34]:

$$\begin{aligned} s_l^{(0)} &= x_{2l} && (1) \\ d_l^{(0)} &= x_{2l+1} && (2) \\ d_l &= d_l^{(0)} - s_l^{(0)} && (3) \\ s_l &= s_l^{(0)} + \frac{1}{2} d_l && (4) \end{aligned}$$

with the inverse transform written as:

$$\begin{aligned} s_l^{(0)} &= s_l - \frac{1}{2} d_l && (5) \\ d_l^{(0)} &= d_l + s_l^{(0)} && (6) \\ x_{2l+1} &= d_l^{(0)} && (7) \\ x_{2l} &= s_l^{(0)}. && (8) \end{aligned}$$

The Haar wavelet filter perfectly preserves[1]. the first moment, i.e. the average of the whole sequence is preserved in the lowpass signal with each transformation step:

$$\sum_l x_l = \frac{1}{2} \sum_k s_k. \qquad (9)$$

This is an important property as it allows the use of lower resolutions for functions like accurate billing, as the sum of the original sequence can be derived from any of the lower resolutions.

Furthermore, the transformation is lossless, and the aggregation is equivalent to subsampling. In effect, the Haar wavelet in the proposed approach is equivalent to reducing the sampling rate.

C. *LeGall* $5/3$ *Wavelet Filter*

The LeGall 5/3 wavelet filter [35] is a biorthogonal wavelet filter, frequently used in image coding. An interesting property of this filter is that its lifting implementation can be realized using integer operations only. With the background of an advanced metering infrastructure with limited computational capacity this can be seen as an advantage.

On the other hand, the LeGall 5/3 filter uses more samples for prediction in the lifting implementation than the Haar wavelet. This may result in longer processing times. Furthermore, and also due to this fact, the LeGall 5/3 filter always requires zero-padding.

The LeGall 5/3 also preserves the first moment. However, due to the necessary border handling, the sum is not perfectly preserved. It depends on the intended application if the loss in accuracy is acceptable. Empirical results on this issue are discussed in Section V.

IV. Conditional Access of Multi-resolution Load Profiles

The idea of conditional access stems from the context of multimedia entertainment data. Entertainment content usually exists in various resolutions (e.g. mobile content, standard definition, high definition), which may be priced differently. A multi-resolution representation of the multimedia data allows the efficient representation of the resolutions in a single bitstream. This is an advantage as only one version of the bitstream needs to be handled and transmitted. Conditional access allows users to pay only for the resolutions they are interesting in. For example, the owner of a standard definition

[1] There may be small discrepancies due to border handling, depending on the length of the input signal. However, this can be resolved by using zero-padding.

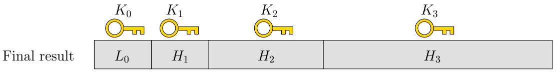


Fig. 1. Final Bitstream Produced by Smart Meter

	$l = 1$	$l = 2$	$l = 3$	$l = 4$	$l = 5$
Haar	0%	0%	0%	0%	0%
LeGall 5/3	0.44%	1.16%	2.24%	6.26%	11.6%

TABLE I
RELATIVE DIFFERENCE IN AGGREGATION OVER 400 LOAD PROFILES

television has no need to pay for the high-definition version of the content. Through conditional access only the bitstream portion relevant for the desired resolution is decrypted, the rest of the bitstream is ignored.

We propose to use the conditional access paradigm for smart-metering data in multi-resolution representation. Each high-pass subband is encrypted with a different key. If desired, the lowest resolution can remain unencrypted to be accessible for each party, e.g. for billing purposes. The whole datastream can be transmitted over Smart Grid communications infrastructure. Access to the different resolutions is thereby only granted to parties that hold the needed keys, as illustrated by Figure 1.

The lowest resolution remains accessible to the energy provider at all times to enable billing. In the example above this is done by leaving R_0 unencrypted (which in principle mirrors the situation in current energy networks). Alternatively, R_0 could also be encrypted with an appropriate key that allows access to the energy provider.

This scheme allows flexible control by the end-user how access is granted to smart meter data. For example, a particular energy provider may be granted access to the lowest resolution for billing purposes, but the end-user may not be willing to provide any detailed usage statistics. A third-party service providing energy saving advice by employing NILM methods may be granted access to the highest resolution by the user. Thereby, a hierarchical keying scheme (e.g., MIKEy – Multimedia Internet KEYing [36]) needs to be employed, allowing parties who hold K_n to access data encrypted with K_i for $i \leq n$.

The proposed scheme also enables relaying of data. For example, the data needed by a third party analysis lab can be forwarded in encrypted form by an aggregator, or even the utility provider. The parties forwarding the data on route to the destination can only access data in the resolution for which they have been cleared by the owner of the data. This may also mean no access at all, i.e. only forwarding is permitted.

Of course, the proposed scheme requires the smart meter hardware to provide functionality for wavelet lifting and encryption, and to support the manual or automatic setting of encryption keys for the higher resolutions. Furthermore solutions for key management, revocation and a key infrastructure need to be provided.

V. EVALUATION

In the following we evaluate the discussed wavelet filters for use in the proposed approach. We use smart meter data from an Austrian energy provider, which was generated by real households over a period of 18 months. In our tests, we use 400 load profiles. The load profiles originate from Siemens smart meter hardware, model TD3510 (3 phase, 100 Amp.). The sampling interval is 15 minutes, i.e. 96 readings a day. This allows a maximum wavelet decomposition depth of 5. With maximum decomposition, the lowest resolution consists of 3 values per day.

A. Aggregation Preservation

For real-world applicability, the lower resolutions need to be created in a way that preserves the original sum. Both investigated wavelet filters preserve the first moment and the original sum can be derived from wavelet decompositions of arbitrary depth. However, due to the necessary border handling, for the LeGall Filter a loss in accuracy is incurred. Table V-A shows the average relative difference for the sum of the lowpass subband compared to the sum of the original data for the 400 load profiles in the test set for different decomposition levels l.

B. Security

There are no methods to infer the higher resolutions by using the information from lower resolutions. Therefore, the higher resolutions are secure, provided that state-of-the art cryptographic ciphers are used.

The proposed scheme does not prevent tampering of smart meter data at the point of origin, i.e. if a tampered smart meter produces fake data, this is not recognized. To prevent this kind of tampering, the proposed scheme needs to be combined with trusted computing (e.g. [23]).

Regarding successful privacy protection of the higher resolutions, the proposed scheme has an advantage over the scheme proposed by [16]. As stated by [37], the privacy afforded by the scheme proposed by [16] may be compromised by data aggregation through an eavesdropper: The eavesdropper could link the tow IDs for low and high frequency data (LFID and HFID, respectively) by summing up high frequency data that he observes. Such an attack is not possible in the scheme proposed here, as all high frequency data is transmitted in encrypted form.

C. Complexity

As discussed above, implementing the wavelet transform as lifting steps is computationally inexpensive. Generally, the discrete wavelet transform has a complexity of $\mathcal{O}(n)$. Due to the simple operations used in the lifting implementation, the transformation part can be realized by inexpensive smart meter hardware.

The computational demands for encryption depends on the used encryption scheme. For standard encryption schemes,

	WAV	AES	RSA	HYB
Average Execution Time (ms)	0.3092	2.36	89.27	92.12
Standard Deviation	0.0356	0.42	4.92	6.59

TABLE II
EXECUTION TIMES FOR HAAR WAVELET ON A BEAGLEBOARD: AVERAGE FOR 400 LOAD PROFILES WITH 1000 EXECUTIONS EACH

	WAV	AES	RSA	HYB
Average Execution Time (ms)	0.2684	2.34	89.15	91.69
Standard Deviation	0.0495	0.42	3.26	1.53

TABLE III
EXECUTION TIMES FOR LEGALL 5/3 WAVELET ON A BEAGLEBOARD: AVERAGE FOR 400 LOAD PROFILES WITH 1000 EXECUTIONS EACH

efficient implementations exist that can even be integrated into smart meter hardware. Depending on the desired scenario, symmetric encryption alone can be used, or in combination with asymmetric encryption. The latter case is computationally more demanding but benefits from the support for public key infrastructures, such as proposed by [23]. Some overhead is introduced for key management, and potentially for the creation of session keys.

Three scenarios are investigated for each wavelet filter: (i) Symmetric encryption: AES with 128 bit keys, (ii) Asymmetric encryption only: RSA with 2048 bit keys, (iii) Hybrid encryption: 128 bit AES session keys encrypted with 2048 bit RSA keys. In each scenario the following steps are executed: (i) Level 5 wavelet transform of the load profile, (ii) Generation of different keys to encrypt resolutions R_1 through R_5 (R_0 is left unencrypted), (iii) Encryption of R_1 through R_5, each with a different key.

The implementation was done in Java (OpenJDK 1.6). The Java standard implementation of the cryptographic routines were used. Lifting implementations were used for both, the Haar wavelet and the LeGall 5/3 wavelet transforms. No special optimization was performed. The tests were run on an low cost embedded environment (Beagleboard BB-XM-00 with a TI DM3730 ARM processor and 512MB of RAM) running Ubuntu Linux 12.04. An ARM-based environment can be envisioned to be used as the central unit for processing and communication in a AMI Home Area Network or even as part of the smart meter.

Tables II and III show the results for the Haar Wavelet and the LeGall Wavelet, respectively. The timing results are given in milliseconds comparing wavelet transform only (WAV) with AES, pure RSA and hybrid encryption (HYB) using an AES session key encrypted with RSA. In each category, 400 load profiles were investigated, each of which was transformed and encrypted 1000 times. The results present the average time needed for processing one load profile.

It can be seen that compared to the computational demands of the encryption stage, the computational demands for the wavelet transform are almost negligible. On average, the transformation of a load profile takes 0.31 ms for the Haar wavelet and 0.27 ms for the LeGall wavelet. The fact that the LeGall uses integer lifting operations accounts for the slightly faster performance.

As expected, symmetric encryption outperforms asymmetric and hybrid encryption by a factor of nearly 40. In application scenarios, that do not require public key management, this advantage will make symmetric encryption a prime candidate.

Due to the limited size of the subbands, public key cryptography can be used directly on the load data. For our test setup, all subbands can be encrypted using 2048 bit RSA keys. It can be observed that the hybrid approach in our scenario is slower than the pure asymmetric approach. This is due to the fact that the load profile subbands are limited in size. Of course, for larger data sets using pure asymmetric encryption is not feasible and the hybrid approach would have to be used. However, it can be rated an advantage that the multi-resolution representation of the load profiles allows the direct application of public key cryptography.

VI. CONCLUSION

Multi-resolution wavelet representation of smart-meter data is a way to balance the need for privacy with the additional functionality introduced by the smart meter load profiles. By using multiple keys to encrypt each resolution separately, the proposed scheme provides end-user control of access to different granularities of the data. Apart from providing user-centric privacy, due to encrypting the higher resolutions the proposed scheme also implements secure transmission of the load profiles and prevents unauthorized access by eavesdroppers.

In terms of choice of wavelet filter, the Haar filter offers the advantage of preserving the aggregate exactly over the different resolutions, which makes functions like billing possible. The fact that the LeGall 5/3 wavelet offers slightly faster computation cannot counterbalance this advantage.

The scheme fits neatly into the larger frameworks proposed to date, such as [23], as it is compatible with other approaches for securing smart grid communication, including authentication, integrity checking, and the integration into smart grid public key infrastructure.

Regarding computational complexity, some overhead is introduced. However, both employed wavelet transforms have very low demands, when implemented as lifting steps. The computational demands for encryption of the higher resolution subbands are higher, especially if an asymmetric or hybrid approach is chosen.

In terms of economic feasibility, it has been shown that the proposed privacy-aware encryption scheme can be employed on inexpensive ARM-based hardware, even running a non-optimized Java implementation on Linux. In dedicated chipsets that offer hardware acceleration for the cryptographic routines the scheme can easily be integrated into smart meters or the corresponding communication gateways.

Acknowledgement

The financial support by the Austrian Federal Ministry of Economy, Family and Youth and the Austrian National Foundation for Research, Technology and Development is gratefully acknowledged.

The author would like to thank Salzburg AG for providing extensive test data.

References

[1] Z. Fan, P. Kulkarni, S. Gormus, C. Efthymiou, G. Kalogridis, M. Sooriyabandara, Z. Zhu, S. Lambotharan, and W. Chin, "Smart grid communications: Overview of research challenges, solutions, and standardization activities," *IEEE Communications Surveys & Tutorials*, vol. PP Issue:99, no. 99, pp. 1–18, 2012.

[2] E. L. Quinn, "Privacy and the new energy infrastructure," *Social Science Research Network (SSRN)*, Feb. 2009.

[3] P. McDaniel and S. McLaughlin, "Security and privacy challenges in the smart grid," *IEEE Security Privacy Magazine*, vol. 7, no. 3, pp. 75–77, 2009.

[4] H. Khurana, M. Hadley, N. Lu, and D. Frincke, "Smart-grid security issues," *IEEE Security & Privacy*, vol. 8, no. 1, pp. 81–85, 2010.

[5] M. A. Lisovich and S. B. Wicker, "Privacy concerns in upcoming residential and commercial demand-response systems," *IEEE Proceedings on Power Systems*, vol. 1, no. 1, 2008.

[6] M. Lisovich, D. Mulligan, and S. Wicker, "Inferring personal information from demand-response systems," *IEEE Security & Privacy*, vol. 8, no. 1, pp. 11–20, 2010.

[7] G. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992.

[8] H. Y. Lam, G. S. K. Fung, and W. K. Lee, "A novel method to construct taxonomy of appliances based on load signatures," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 2, pp. 653–660, 2007.

[9] 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*, ser. BuildSys '10. New York, NY, USA: ACM, 2010, pp. 61–66. [Online]. Available: http://doi.acm.org/10.1145/1878431.1878446

[10] S. K. J. Leung, S. H. K. Ng, and W. M. J. Cheng, "Identifying appliances using load signatures and genetic algorithms," in *Proceedings International Conference on Electrical Engineering (ICEE)*, Hong Kong, Jul. 2007.

[11] G. Kalogridis and S. Z. Denic, "Data mining and privacy of personal behaviour types in smart grid," in *Proc. IEEE 11th Int Data Mining Workshops (ICDMW) Conf*, 2011, pp. 636–642.

[12] D. Engel, "Conditional access smart meter privacy based on multiresolution wavelet analysis," in *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies*. New York, NY, USA: ACM, 2011, pp. 45:1–45:5.

[13] European Commission Task Force Smart Grids, Expert Group 2: Regulatory Recommendations for Data Safety, Sata Handling and Data Protection, "Report," http://ec.europa.eu/energy/gas_electricity/smartgrids/doc/expert_group2.pdf, Feb. 2011, online.

[14] Bundesamt für Sicherheit in der Informationstechnik (German Federal Office for Information Security), "Protection profile for the gateway of a smart metering system," https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/SmartMeter/PP-SmartMeter.pdf, Aug. 2011, final Draft Version 01.01.01. [Online]. Available: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/SmartMeter/PP-SmartMeter.pdf

[15] G. Iachello and J. Hong, "End-user privacy in human-computer interaction," *Found. Trends Hum.-Comput. Interact.*, vol. 1, pp. 1–137, January 2007. [Online]. Available: http://dl.acm.org/citation.cfm?id=1324103.1324104

[16] C. Efthymiou and G. Kalogridis, "Smart grid privacy via anonymization of smart metering data," in *Proceedings of First IEEE International Conference on Smart Grid Communications*, Gaithersburg, Maryland, USA, Oct. 2010, pp. 238–243.

[17] J.-M. Bohli, C. Sorge, and O. Ugus, "A privacy model for smart metering," in *Proc. IEEE Int Communications Workshops (ICC) Conf*, 2010, pp. 1–5.

[18] Y. Kim, E. C.-H. Ngai, and M. B. Srivastava, "Cooperative state estimation for preserving privacy of user behaviors in smart grid," in *Proc. IEEE Int Smart Grid Communications (SmartGridComm) Conf*, 2011, pp. 178–183.

[19] L. Sankar, S. Kar, R. Tandon, and H. V. Poor, "Competitive privacy in the smart grid: An information-theoretic approach," in *Proc. IEEE Int Smart Grid Communications (SmartGridComm) Conf*, 2011, pp. 220–225.

[20] H. Li, R. Mao, L. Lai, and R. C. Qiu, "Compressed meter reading for delay-sensitive and secure load report in smart grid," in *Proc. First IEEE Int Smart Grid Communications (SmartGridComm) Conf*, 2010, pp. 114–119.

[21] A. Rial and G. Danezis, "Privacy-preserving smart metering," in *Proceedings of the 10th annual ACM workshop on privacy in the electronic society*, ser. WPES '11. New York, NY, USA: ACM, 2011, pp. 49–60. [Online]. Available: http://doi.acm.org/10.1145/2046556.2046564

[22] A. Bartoli, J. Hernández-Serrano, M. Dohler, A. Kountouris, and D. Barthel, "Secure lossless aggregation for smart grid M2M networks," in *Proceedings of First IEEE International Conference on Smart Grid Communications*, Gaithersburg, Maryland, USA, Oct. 2010, pp. 333–338.

[23] A. R. Metke and R. L. Ekl, "Security technology for smart grid networks," *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 99–107, Jun. 2010.

[24] T. W. Chim, S. M. Yiu, L. C. K. Hui, and V. O. K. Li, "Pass: Privacy-preserving authentication scheme for smart grid network," in *Proc. IEEE Int Smart Grid Communications (SmartGridComm) Conf*, 2011, pp. 196–201.

[25] Y.-J. Kim, V. Kolesnikov, H. Kim, and M. Thottan, "SSTP: A scalable and secure transport protocol for smart grid data collection," in *Proc. IEEE Int Smart Grid Communications (SmartGridComm) Conf*, 2011, pp. 161–166.

[26] H. Cheung, A. Hamlyn, T. Mander, C. Yang, and R. Cheung, "Role-based model security access control for smart power-grids computer networks," in *Proc. IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, 2008, pp. 1–7.

[27] H. K.-H. So, S. H. Kwok, E. Y. Lam, and K.-S. Lui, "Zero-configuration identity-based signcryption scheme for smart grid," in *Proceedings of First IEEE International Conference on Smart Grid Communications*, Gaithersburg, Maryland, USA, Oct. 2010, pp. 321–326.

[28] F. Garcia and B. Jacobs, "Privacy-friendly energy-metering via homomorphic encryption," in *Security and Trust Management*, ser. Lecture Notes in Computer Science, J. Cuellar, J. Lopez, G. Barthe, and A. Pretschner, Eds. Springer Berlin / Heidelberg, 2011, vol. 6710, pp. 226–238. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-22444-7_15

[29] F. Li, B. Luo, and P. Liu, "Secure information aggregation for smart grids using homomorphic encryption," in *Proceedings of First IEEE International Conference on Smart Grid Communications*, Gaithersburg, Maryland, USA, Oct. 2010, pp. 327–332.

[30] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Proceedings of Eurocrypt '99, Advances in Cryptology*, ser. Lecture Notes in Computer Science, J. Stern, Ed., vol. 1592. Prague, Czech Republic: Springer, May 1999, pp. 223–238.

[31] C. Thoma, T. Cui, and F. Franchetti, "Secure multiparty computation based privacy preserving smart metering system," in *44th North American Power Symposium (NAPS)*, 2012. to appear.

[32] L. Liu, J. Wang, and J. Zhang, "Wavelet-based data perturbation for simultaneous privacy-preserving and statistics-preserving," in *Proc. IEEE Int. Conf. Data Mining Workshops ICDMW '08*, 2008, pp. 27–35.

[33] S. R. Rajagopalan, L. Sankar, S. Mohajer, and H. V. Poor, "Smart meter privacy: A utility-privacy framework," in *Proc. IEEE Int Smart Grid Communications (SmartGridComm) Conf*, 2011, pp. 190–195.

[34] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," *J. Fourier Anal. Appl.*, vol. 4, no. 3, pp. 247–269, 1998.

[35] D. Le Gall and A. Tabatabai, "Sub-band coding of digital images using symmetric short kernel filters and arithmetic coding techniques," in *Proc. Int Acoustics, Speech, and Signal Processing ICASSP-88. Conf*, 1988, pp. 761–764.

[36] J. Arkko, E. Carrara, F. Lindholm, M. Naslund, and K. Norrman, "MIKEY: Multimedia Internet KEYing," RFC 3830 (Proposed Standard), Internet Engineering Task Force, Aug. 2004. [Online]. Available: http://www.ietf.org/rfc/rfc3830.txt

[37] T. Baumeister, "Literature review on smart grid cyber security," University of Hawaii at Manoa, Tech. Rep., Dec. 2010.

3.10 ENGEL13E

- D. Engel and G. Eibl. Multi-resolution load curve representation with privacy-preserving aggregation. In *Proceedings of IEEE Innovative Smart Grid Technologies (ISGT) 2013*, pages 1–5, Copenhagen, Denmark, Oct. 2013. IEEE.

Multi-Resolution Load Curve Representation with Privacy-preserving Aggregation

Dominik Engel and Günther Eibl
Josef Ressel Center for User-Centric Smart Grid Privacy, Security and Control
Salzburg University of Applied Sciences
Urstein Sued 1, A–5412 Urstein/Salzburg, Austria
Email: {dominik.engel, guenther.eibl}@en-trust.at

***Abstract*—The availability of individual load curves per household in the smart grid end-user domain combined with non-intrusive load monitoring to infer personal data from these load curves has led to privacy concerns. Two types of approaches show high potential to resolve this issue: (i) secure aggregation and (ii) multi-resolution representation with conditional access. In this paper a combination of these two principle approaches is proposed. It is shown formally that secure aggregation and wavelet-based multi-resolution representation are compatible. Furthermore, it is shown that that the wavelet transformation is compatible with existing privacy-preserving protocols and can be used to extend them with additional degrees of freedom. An implementation of the proposed approach is used for evaluation of feasibility in a low-cost embedded environment.**

I. Introduction

Intelligent energy systems, so-called smart grids, revolutionize existing energy grids by combining them with information and communication technology. Smart grids demand accurate and fine-grained data on network status. The widespread roll-out of smart meters is one of the consequences. Smart meters record energy consumption in a specified granularity (usually the time between readings is between 1 second and 15 minutes) and have the ability to transmit these load curves in a specified interval (e.g., once a day).

It has been shown that personal information on the end-user can be inferred from fine-grained load curves [1], [2], and this has led to privacy concerns [3], [4]. The accuracy of the inferred information is directly connected to the available resolution of the load data. A number of methods have been proposed to balance the need for privacy with the information needed for correct operation of smart grids. Two types of approaches show high potential to resolve this issue: (i) secure aggregation of encrypted load curves, and (ii) representation of load curves in multiple resolutions, each associated with different access levels.

Approaches of the first type can again be divided into two categories: protocols using masking [5], [6] and protocols using homomorphic encryption. In this paper the focus is put on the second kind of protocols. Privacy-enabling encryption for smart meter data by the use of homomorphic encryption is suggested by, e.g., [7]–[10], allowing the aggregation of encrypted signals, also termed "secure signal processing". A recent overview of secure signal processing, covering four proposals for privacy-preserving smart metering aggregation is given by [11].

Approaches of the second type suggest to represent load curve data in multiple resolutions, where each resolution can be used for a different purpose, e.g., low resolution for billing, and is therefore disclosed to selected parties only, e.g., [12]. Using the wavelet transform to produce an integrated bitstream supporting multiple resolutions has been proposed by [13]. Combined with conditional access, i.e., different encryption keys for each resolution, this wavelet-based representation allows user-centric privacy management: access can be granted or revoked for each resolution. Access to high resolutions, which are privacy-sensitive, may be reserved to a small number of trusted entities only, whereas resolutions of medium granularity may be provided more freely, e.g., to contribute to network stability (in exchange for lower energy prices or other incentives).

In this paper, a privacy-preserving smart metering method that combines the two types of approaches, namely homomorphic encryption and multi-resolution representation, is proposed. This enhances the possibilities for managing privacy requirements, as the combination of both methods significantly increases the degrees of freedom. Access control does not relate to the aggregated signal as a whole anymore, but access can be granted on the aggregate on each resolution *individually*. This is an important feature, as it allows to grant access to participants in the smart grid system, based on their roles and the functions they have to fulfill. Each role can be assigned access to the aggregate on the minimum resolution necessary to fulfill the functions associated with this role.

The rest of this paper is structured as follows. Section II summarizes the principle of wavelet-based load curve representation and homomorphic encryption. In Section III the combination of the two approaches is introduced and their compatibility is proven mathematically. Results are discussed in Section IV-A. The usability of the wavelet transformation with existing protocols is discussed in IV-B. Section V concludes the paper.

II. Background

A. Wavelet-based Representation

A wavelet transform starts with the original load curve and is recursively performed in S steps. In each step s half of the

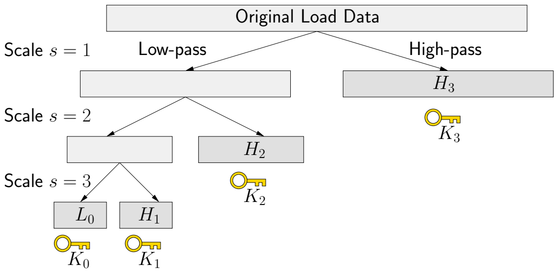


Fig. 1. Wavelet transformation and encryption of load curve

data (the highpass data) $\tilde{H}_s$ are remembered as the wavelet coefficients (subband) of scale s and the next step is performed for the lowpass data. At the end of the transformation the final subband $\tilde{H}_S$ consists of 2^{-S} samples. The higher the scale s, the lower the time resolution $r := S - s$. Reindexing $H_r = \tilde{H}_{S-s}$, at the end of the transformation one obtains a sequence $L_0,\ H_1, \ldots,\ H_S$, see Fig. 1.

The synthesis step of the inverse wavelet transform W^{-1} starts with the lowest resolution $r = 0$. To get the next higher resolution of the signal the next higher resolution subband is needed, so that in a series of S steps one finally obtains the original load curve (since we only consider lossless transformations). In order to provide a signal m_R with maximum resolution R only R synthesis steps must be performed and only the subbands with resolution $r \leq R$, i.e., $L_0,\ H_1, \ldots,\ H_R$ are needed. Denoting the selection of the R highest resolutions as T_R this can be written as

$$m_R = W^{-1}\left[T_R(W[m])\right] \qquad (1)$$

Making the signal available at the needed resolution instead of the full resolution increases privacy because less (personal) information can be deduced. For allowing differentiated access control, each subband (i.e., each resolution) of the resulting wavelet decomposition is encrypted with a different key, as illustrated by Fig. 1. A combination with public key infrastructures to allow fine-grained access control is possible. A hierarchical key creation scheme can be used to minimize overhead for key exchange. The result of this process is an encrypted bitstream that forms an integrated representation of all resolutions. Note that no data expansion occurs, i.e., the size of the final bitstream equals the size of the original data.

The operator T_R can be generalized to be any transformation T of the wavelet coefficients to be used for example for denoising. In the simplest case, using a global threshold η it could be defined as $T(W[m]) = W[m]\delta(m-\eta)$ with δ denoting Dirac's delta function, for more sophisticated denoising methods, see [14]. Using denoising transformations could turn out to be valuable for transmission of signal aggregations.

Wavelets are also used for the generation of features in load forecasting [15], [16]. The representation suggested in this paper may be of advantage for load forecasting methods based on wavelets. However, the main focus of this paper is load aggregation with access control to different resolutions.

Key generation: generate public keys g and n and private key λ
- Generate private key prime numbers p and q randomly
- Set private $\lambda = \mathrm{lcm}(p-1, q-1)$
- Set public key $n := pq$
- Select public key g with the property
  $\gcd(L(g^{\lambda} \mod n^2),\ n) = 1$ for $L(u) := \frac{u-1}{n}$

Encryption: given message $m \in \mathbb{Z}_n$
- Generate random number $r \in \mathbb{Z}_n^*$
- Ciphertext $c = E(m; g, n) = g^m \cdot r^n \mod n^2 \in \mathbb{Z}^*_{n^2}$

Decryption: given ciphertext $c \in \mathbb{Z}^*_{n^2}$
- $m = D(c; g, n, \lambda) = \frac{L(c^{\lambda} \mod n^2)}{L(g^{\lambda} \mod n^2)} \mod n \in \mathbb{Z}_n$

TABLE I
HOMOMORPHIC ENCRYPTION

B. Homomorphic Encryption

Following previous proposals [7]–[9], a Paillier cryptosystem [17] is employed. The whole encryption and decryption process can be split into three parts: key-generation, encryption and decryption. It is described in Table I. Note that the numbers g, n and λ are kept fixed and are omitted for simplicity.

Homomorphic encryption has the following important property, which is called the *additive property*:

$$D\left(E(m_1)E(m_2) \mod n^2\right) = m_1 + m_2 \mod n. \qquad (2)$$

This property means that the decryption of the product of the *ciphertexts* is the sum of the original plaintext messages.

C. Privacy Preserving Protocols

In [7]–[10], protocols using homomorphic encryption are proposed as tools for privacy conserving aggregation of load curves. As it is done there, the smart grid network considered consists of N households each having one smart meter installed and an aggregator (Fig. 2).

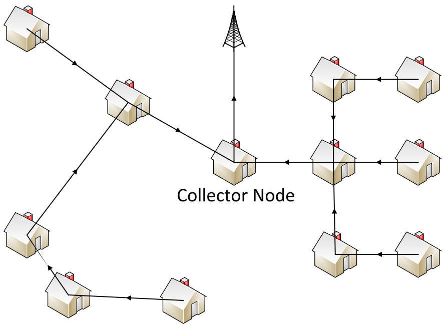


Fig. 2. Aggregation of encrypted signals

The network is assumed to have tree-like connections. Each smart meter i sends its measured load m_i in encrypted form to its parent smart meter. The parent smart meter multiplies the obtained encrypted signals with its own encrypted signal and in turn sends this product to its parent node. Finally, the aggregator multiplies the obtained signals and decrypts the product. Due to the homomorphic property, the result is the sum of the measured loads. With E and D denoting Pailler encryption and decryption this can be stated as

$$D\left(\prod_i E(m_i) \bmod n^2\right) = \sum_i m_i \mod n. \tag{3}$$

Privacy is preserved because of the distributed way of processing. Smart meters only have the plaintext information of their own messages, because they cannot decrypt the messages they get. The aggregator can decrypt messages, but, as it receives the product of the individual ciphertexts, can only decrypt the sum of the load curves.

III. Aggregation of encrypted wavelet-transformed signals

The goal of this paper is an extension of the distributed homomorphic encryption process in a way that is compatible with the wavelet transformation. In particular it is shown that when homomorphic encryption is applied to a signal represented in the wavelet domain, homomorphic additivity is not only preserved, but can be separately exploited for each resolution.

In [13], a variety of wavelet filters regarding their utility for the multi-resolution representation of load curves was evaluated. Only lossless transformations are useful in the context of smart metering. The Haar wavelet filter preserves the average over all resolutions, which is an important property for many use cases. Using the lifting implementation of the Haar wavelet, the transformation can be realized efficiently. The lifting steps for the forward transform with the Haar wavelet have been formulated by [18]. As the original Haar wavelet uses real coefficients, it is ill-suited for use with homomorphic encryption. Therefore, for the combination with homomorphic encryption a modified version of the Haar wavelet is used that only produces integer values for the transformed load curve (where $\tilde{L}_0 = X[i]$ is the input signal, $\tilde{H}_s[i]$ and $\tilde{L}_s[i]$ are the resulting high-pass and low-pass subband at scale s, respectively, with i denoting the position within the signal):

$$\begin{aligned}
\tilde{L}^{(0)}_{s+1}[i] &= \tilde{L}_s[2i] && (4)\\
\tilde{H}^{(0)}_{s+1}[i] &= \tilde{L}_s[2i+1] && (5)\\
\tilde{H}_{s+1}[i] &= \tilde{H}^{(0)}_{s+1}[i] - \tilde{L}^{(0)}_{s+1}[i] && (6)\\
\tilde{L}_{s+1}[i] &= 2\tilde{L}^{(0)}_{s+1}[i] + \tilde{H}_{s+1}[i]). && (7)
\end{aligned}$$

The inverse transform can be written as:

$$\begin{aligned}
\tilde{L}^{(0)}_{s}[i] &= \frac{1}{2}\tilde{L}_{s+1}[i] - \frac{1}{2}\tilde{H}_{s+1}[i] && (8)\\
\tilde{H}^{(0)}_{s}[i] &= \tilde{H}_{s+1}[i] + \tilde{L}^{(0)}_{s}[i] && (9)\\
\tilde{L}_s[2i+1] &= \tilde{H}^{(0)}_{s}[i] && (10)\\
\tilde{L}_s[2i] &= \tilde{L}^{(0)}_{s}[i]. && (11)
\end{aligned}$$

Note that the average of the original series is still preserved over all resolutions for the modified Haar filter:

$$\forall s: \ \sum_i X[i] = 2^{-s} \sum_k \tilde{L}_s[k].$$

Fig. 3 shows the aggregation of a number of multi-resolution load curves at a collector node. Homomorphic encryption is applied to each resolution r separately with a different key $K_r = (g_r,\ n_r)$. The resulting signal m is the sum of all signals m_i (each of which has a maximum resolution of R) at resolution $r \leq R$, whereby W denotes a wavelet transformation. The collector node can perform aggregation (i.e., multiply) in the encrypted domain, i.e., it does not have any keys. This ensures that the aggregator cannot get information about the loads of its children, e.g., by divisions.

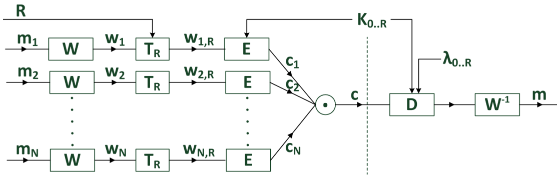


Fig. 3. Aggregation of encrypted multi-resolution load curves

Writing the procedure mathematically yields the following calculation of the ciphertext c

$$c = \prod_i E(T_R(W[m_i])) \bmod n^2.$$

The ciphertext c is decrypted by the aggregator in the following way

$$m = W^{-1}[D(c) \bmod n]$$

Using this procedure the wavelet transformation is compatible with homomorphic encryption, i.e., the homomorphic property that the message m equals the sum of the messages is preserved (choosing $R = S$). Even more, choosing $R < S$, the decrypted message m equals the sum of the messages of resolution R:

$$m = W^{-1}[\prod_i E(T_R(W[m_i])) \bmod n] = \sum_i m_{R,i} \bmod n. \tag{12}$$

The aggregator gets the product of the encrypted messages and can therefore not extract any information about the individual messages. However, it can calculate the sum of the messages which is the information needed, e.g., for load forecasting. Note again that the product of the ciphertexts is calculated in a distributed way by the smart meters and not by the aggregator. The number n must be chosen big enough so that $\prod_i E(T_R(W[m_i])) < n^2$ and $\prod_i E(T_R(W[m_i])) < n$ hold. For sake of readability the modulo parts of the calculations are therefore omitted in the proof.

Proof: Without loss of generality two messages are considered. To simplify the analysis the notation $y_i := T_R(W[m_i])$ is used, so $E(T_R(W[m_i])) = E(y_i)$. The aggregator calculates the signal $W^{-1}[D(c)]$. Using the fact that the ciphertext c is the product of the individual ciphertexts and the homomorphic encryption property leads to

$$\begin{aligned} W^{-1}[D(c)] &= W^{-1}[D(c_1c_2)] \\ &= W^{-1}[D(E(y_1)E(y_1))] \\ &= W^{-1}[y_1 + y_2] \end{aligned}$$

Substituting for the y_i, using the linearity of the wavelet transform and the definition of m_R yields

$$\begin{aligned} W^{-1}[D(c)] &= W^{-1}[T_R(W[m_1]) + T_R(W[m_2])] \\ &= W^{-1}[T_R(W[m_1])] + W^{-1}[T_R(W[m_2])] \\ &= m_{R,1} + m_{R,2} \end{aligned}$$

So in general for I different messages and ciphertext $c = \prod_i c_i$ the desired property (12)

$$W^{-1}[D(c)] = \sum_{i=1}^{I} m_{R,i} \tag{13}$$

is obtained. ■

An example use-case scenario is the use of aggregated load information for energy monitoring by the network operator, as, e.g., suggested by [11]. The approach proposed here adds an additional layer of flexibility by making the aggregates available at different resolutions with access being granted to parties on the resolutions with the necessary granularity to fulfill a specific task. In combination with suitable key management, this approach implements the "need-to-know" principle of access for aggregated signals.

IV. Results

A. Cost and complexity

The proposed method has been implemented as a proof of concept in Java (Oracle Java v8 preview with ARM-extensions) and evaluated in a low-cost ARM-based environment (Beagleboard BB-XM-00, Rev C, with a TI DM3730 1Ghz ARM processor and 512MB of RAM) running Ubuntu Linux 12.04.

Results are shown in Table II: Each value represents the execution time for a single load curve consisting of 96 values for the wavelet transform combined with different encryption settings, averaged over 400 load curves with 100 encryptions each (acquisition of the load curve and key generated are not considered in the timing results). WAV denotes the wavelet transform only, without any encryption applied. AES denotes the wavelet transform followed by encryption with the symmetric AES cipher with a 128 bit key for each subband. HYB denotes hybrid encryption, which adds RSA 2048 bit public key encryption of the AES keys with a different public key for each subband. Finally, PAI-n denotes Pailler encryption with a module of n bits and a different key for each subband.

It can be seen that by using a lifting implementation the transformation is very fast and the computational overhead is negligible compared to the encryption step. Homomorphic encryption comes at the cost of a significant increase in computational overhead compared to conventional encryption.

	WAV	AES	HYB	PAI-256	PAI-512	PAI-1024
Exec. time	**0.15**	1.91	72.4	1,649	11,452	85,355
Std. dev.	**0.01**	0.03	0.1	16	22	133

TABLE II
Execution time in milliseconds for transforming/encrypting a single load curve (average over 400 load curves with 100 encryptions each)

The results show that the computational demands grow exponentially with the module size. Considering that 256 and 512 bit modules will in most use-cases not be sufficient in terms of security, the increased execution time for module sizes that are more secure provides a challenge. While AES encryption only takes 1.9 ms, for the used (non-optimized) implementation, Paillier encryption of a load curve with 96 values takes nearly 90 seconds for a module of 1024 bit. It needs to be pointed out that this drawback also affects all previously proposed methods for homomorphic load curve encryption that rely on a Paillier cryptosystems. Optimization of the implementation is one option to be considered. Another option is to investigate the utility of alternative homomorphic encryption schemes.

The approach proposed here adds the possibilities offered by wavelets to distributed homomorphic encryption and decryption schemes. It is therefore compatible with any homomorphic encryption scheme. The wavelet transformation can be seen as an add-on which is compatible with homomorphic encryption. Since the computational cost of the wavelet transformation is small, the computational cost of the main privacy preserving protocol dominates the overall cost. Thus, the complexity evaluation given in [11] can be used as a complexity assessment for different kinds of privacy preserving protocols.

B. Usability with existing protocols

The extension of the privacy preserving protocol was designed for the protocol used in [7]. Thus it can readily be used within privacy preserving protocols, which directly rely on the homomorphic encryption property such as [7], [9]. Here we study, if wavelets can also be used together with other protocols found in the literature.

The method in [8] combines Paillier's homomorphic encryption with additive secret sharing. Generally, additive masking terms need no adjustment since they cancel out in the decryption step before the inverse transformation takes place. Thus, the method is compatible with the wavelet transformation.

The method in [10] extends [7] by preserving data integrity. The wavelet transformation is compatible with this method since it is mostly based on the ciphertext. There, it is irrelevant if the encrypted message is in its original or in a transformed form. Decryption is only done in the incremental verification process where the compatibility can be verified for each individual step.

Other existing protocols need a homomorphic property but do not use Paillier's homomorphic encryption [6], or they use other principles as for example masking [6], [5]. Next, it will be checked, if the wavelet transformation is also compatible with these methods.

In [6], the modulo operation is used for homomorphic encryption instead of Paillier's homomorphic encryption scheme. Privacy is achieved by masking. The second main feature is the addition of Laplacian noise for differential privacy. This encryption scheme can be made compatible with the wavelet transformation by the following modifications: the multiplication in the aggregation step must be substituted by an addition. The signal $T_R(W(m_i))$ corresponds to the signal X_i in [6]. As already stated above, the additive masking terms need no adjustment. The same argument holds for the keys added for ensuring confidentiality with the aggregator. However, the terms for differential privacy need to be modified. The added noise must be adapted in two ways due to the inverse transformation W^{-1} arising in the decryption step: first, the parameter λ must be chosen suitable for the signals $W^{-1}(T_R(W(m_i))) = m_{R,i}$. Second, the noise added to each signal $T_R(W(m_i))$ which consists of the subtraction of two gamma distributions (with the adapted parameter λ) must be transformed by W which later cancels the W^{-1} in the decryption step. With these changes wavelets are compatible with the method of [6].

In [5], four different protocols which rely on masking are described. These protocols can be categorized into so-called aggregation and comparison protocols. The aggregation protocols are compatible with wavelets. However, in the comparison protocols, the transformed sum of the values is in the exponent of the generating element of the Diffie-Hellman group. As the reverse transformation cannot be calculated, wavelets are not compatible with these comparison protocols.

Summarizing, the wavelet method is compatible with existing privacy preserving protocols except comparison protocols. Adaptations are needed for differential privacy.

V. Conclusion and Outlook

The proposed approach enables access models on a "need-to-know" basis for secure signal processing. This adds flexibility to existing approaches and enhances privacy. Access control to encrypted aggregates is not binary for the whole signal anymore, but instead can be granted to parties for individual resolutions, based on their roles and the associated specific needs in terms of data resolution. A proof has been given that shows that the wavelet transform is compatible with any homomorphic encryption method. Furthermore, the proposed approach can be included in most existing privacy-preserving protocols to enhance the degrees of freedom. Computational demands of homomorphic encryption schemes in general remain a challenge. The overhead for multi-resolution processing is negligible compared to the complexity of encryption.

Like most papers this paper focuses on methodological aspects. In the future we will extend existing work [19] and investigate how these methods can be applied to the relevant use cases like energy feedback, billing or grid stability including practical aspects such as robustness against losing the connection to individual smart meters.

Acknowledgements

The financial support by the Austrian Federal Ministry of Economy, Family and Youth and the Austrian National Foundation for Research, Technology and Development is gratefully acknowledged.

The authors thank company partner Salzburg AG for providing anonymized real-world load curves for testing.

References

[1] G. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992.

[2] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin, "Private memoirs of a smart meter," in *Proc. 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, ser. BuildSys '10. New York, NY, USA: ACM, 2010, pp. 61–66.

[3] P. McDaniel and S. McLaughlin, "Security and privacy challenges in the smart grid," *IEEE Security Privacy Magazine*, vol. 7, no. 3, pp. 75–77, 2009.

[4] M. Lisovich, D. Mulligan, and S. Wicker, "Inferring personal information from demand-response systems," *IEEE Security & Privacy*, vol. 8, no. 1, pp. 11–20, 2010.

[5] K. Kursawe, G. Danezis, and M. Kohlweiss, "Privacy-friendly aggregation for the smart grid," in *Privacy Enhanced Technology Symposium*, 2011, pp. 175–191.

[6] G. Acs and C. Castelluccia, "I have a dream! (differentially private smart metering)," in *Proc. Information Hiding Conference*, 2011, pp. 118–132.

[7] F. Li, B. Luo, and P. Liu, "Secure information aggregation for smart grids using homomorphic encryption," in *Proc. of First IEEE International Conference on Smart Grid Communications*, Gaithersburg, Maryland, USA, Oct. 2010, pp. 327–332.

[8] F. Garcia and B. Jacobs, "Privacy-friendly energy-metering via homomorphic encryption," in *Security and Trust Management*, ser. Lecture Notes in Computer Science, J. Cuellar, J. Lopez, G. Barthe, and A. Pretschner, Eds. Springer Berlin / Heidelberg, 2011, vol. 6710, pp. 226–238.

[9] Z. Erkin and G. Tsudik, "Private computation of spatial and temporal power consumption with smart meters," in *Proceedings of the 10th international conference on Applied Cryptography and Network Security*, ser. ACNS'12. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 561–577.

[10] F. Li and B. Luo, "Preserving data integrity for smart grid data aggregation," in *Smart Grid Communications (SmartGridComm), 2012 IEEE Third International Conference on*, 2012, pp. 366–371.

[11] Z. Erkin, J. Troncoso-Pastoriza, R. Lagendijk, and F. Perez-Gonzalez, "Privacy-preserving data aggregation in smart metering systems: An overview," *Signal Processing Magazine, IEEE*, vol. 30, no. 2, pp. 75–86, March.

[12] C. Efthymiou and G. Kalogridis, "Smart grid privacy via anonymization of smart metering data," in *Proceedings of First IEEE International Conference on Smart Grid Communications*, Gaithersburg, Maryland, USA, Oct. 2010, pp. 238–243.

[13] D. Engel, "Wavelet-based load profile representation for smart meter privacy," in *Proc. IEEE PES Innovative Smart Grid Technologies (ISGT'13)*, Washington, D.C., USA, Feb. 2013, pp. 1–6.

[14] A. Anestis, J. Bigot, and T. Sapatinas, "Wavelet estimators in nonparametric regression: a comparative simulation study," *Journal of Statistical Software*, vol. 6, pp. 1–83, 2001.

[15] C. Chen, B. Das, and D. J. Cook, "Energy prediction based on resident's activity," in *Proceedings of the 4th International Workshop on Knowledge Discovery from Sensor Data*, Jul. 2010.

[16] C. Guan, P. Luh, L. Michel, Y. Wang, and P. Friedland, "Very short-term load forecasting: Wavelet neural networks with data pre-filtering," *IEEE Transactions on Power Systems*, vol. 28, pp. 30–41, 2013.

[17] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Proceedings of Eurocrypt '99, Advances in Cryptology*, ser. Lecture Notes in Computer Science, J. Stern, Ed., vol. 1592. Prague, Czech Republic: Springer, May 1999, pp. 223–238.

[18] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," *J. Fourier Anal. Appl.*, vol. 4, no. 3, pp. 247–269, 1998.

[19] M. Jawurek, F. Kerschbaum, and G. Danezis, "Privacy technologies for smart grids - a survey of options," Microsoft Research, Tech. Rep., 2012.

3.11 PEER14A

- C. Peer, D. Engel, and S. Wicker. Hierarchical key management for multi-resolution load data representation. In *Proceedings of 5th IEEE International Conference on Smart Grid Communications (SmartGridComm 2014)*, pages 926–932, Venice, Italy, Nov. 2014. IEEE.

2014 IEEE International Conference on Smart Grid Communications

Hierarchical Key Management for Multi-resolution Load Data Representation

Christian D. Peer
Josef Ressel Center for
User-Centric Smart Grid Privacy,
Security and Control
Salzburg University of
Applied Sciences, Austria
Email: christian.peer@en-trust.at

Dominik Engel
Josef Ressel Center for
User-Centric Smart Grid Privacy,
Security and Control
Salzburg University of
Applied Sciences, Austria
Email: dominik.engel@en-trust.at

Stephen B. Wicker
School of Electrical and
Computer Engineering
Cornell University
Ithaca, New York
Email: wicker@ece.cornell.edu

***Abstract*—It has been shown that information about a consumer's actions, beliefs and preferences can be extracted from high resolution load data. This information can be used in ways that violate consumer privacy. In order to increase consumer control over this information, it has been suggested that load data be represented in multiple resolutions, with each resolution secured with a different key. To make this approach work in the real-world, a suitable key management needs to be employed. In this paper, we consider a combination of multi-resolution load data representation with hierarchical key management. Emphasis is placed on a privacy-aware design that gives the end-user the freedom to decide which entity is allowed to access user related data and at what granularity.**

I. Introduction

Increasing energy needs accompanied by an emphasis on alternative energy production creates a need for efficient power grid management and regulated power consumption. This so-called Smart Grid enables load balancing and forecasting within the power grid. In addition it is able to influence the consumer's energy consumption by offering real-time pricing information. Based on this information, consumers can decide when to use devices so as to manage energy costs. Studies show that Smart Grid Infrastructure can reduce peak load during summertime by as much as 20% [1]. To fulfill this task, the Smart Grid relies on Advanced Metering Infrastructure (AMI), a sensor network collecting fine-grained power consumption data. Smart Meters form the core component of an AMI. These devices collect fine-grained consumption data, so-called load data, from a single household. While this data plays an essential part in load balancing and real-time pricing, its collection also creates serious privacy concerns.

It has been shown that apart from information needed for grid operation, other pieces of information can be obtained from fine-grained load data that are sensitive and private to the end user [2]–[4]. Occupancy or sleeping patterns can be determined and certain appliances within the household can be identified and a usage pattern can be drawn. This information can be valuable for targeted marketing as well as criminal purposes. With regard to the former, techniques for matching appliance signatures to load data are called non-intrusive load monitoring (NILM) or non-intrusive appliance load monitoring (NALM) [3].

Acting on privacy concerns, customers and governments are rejecting the deployment of Smart Meters and therefore blocking the deployment of the Smart Grid [5]. To address this issue, privacy preserving methods have to be implemented. Two types of approaches show great potential for ensuring privacy within the smart grid: (i) Secure aggregation of encrypted load data and (ii) consumer control over load data in multiple resolutions, each resolution associated with different access levels. In terms of secure aggregation, Erkin et al. give an overview of the recent development in [5].

This paper will focus on the representation and securement of load data in multiple resolutions. NILM/NALM techniques need high resolution load data to gain accurate results. By lowering the resolution of the load data, NILM/NALM techniques can only achieve limited results. While a low resolution on a daily average is sufficient for accounting purposes, applications like load forecasting or energy saving tools require high resolution load data to achieve useful results. This is where multi-resolution load data representation is needed. Each resolution is encrypted with a different key. Trusted services or third parties are only granted access to the resolution level necessary to fulfill their role. Access can be controlled by a trusted authority, or better, by the user. This adds a new degree of freedom, as the user can decide which party gains access to which data.

An approach on how to represent load data in multiple resolutions can be found in [6]. While this approach describes how to split load data in multiple resolutions, it leaves the question about suitable key generation and management unanswered. In this paper, a key management system suitable for accessing multi-resolution load data within the Smart Grid Infrastructure will be introduced. Furthermore this paper will suggest the use of hierarchical keys to keep key management efforts as low as possible.

This paper also proposes a general communication infrastructure fulfilling the requirements within the Smart Grid Infrastructure. It allows secure communication between entities and third party entities without exposing the Smart Meters

978-1-4799-4934-2/14/$31.00 ©2014 IEEE

to a public network. Hence, it minimizes the risk of possible attacks on the Smart Grid Infrastructure.

The rest of the paper is organized as follows: Related Work and the state of the art are discussed in Section II. The proposed key management system is discussed in detail in Section III. Section IV introduces the idea of hierarchical key management and generation. Finally Section V summarizes the most important findings.

II. Related Work

The following section gives a short overview of the technologies on which this paper is based.

A. Privacy Preserving Architecture

There are different possibilities to enforce privacy protection. One is by regulation and law. While this basic idea is essential for a modern society, it still offers the potential to violate privacy using legal or illegal means. As long as system design and architecture offer the possibility to collect personally identifying information, there is a possibility to violate privacy protection. Therefore a better approach is to ensure privacy protection by design. In [7], Wicker et al. propose a framework for privacy aware design tailored to the development of demand response architectures. They suggest five major elements:

1) *Provide Full Disclosure of Data Collection:* Information on which data is collected, collection purpose and duration of storage has to be provided to the consumer
2) *Require Consent to Data Collection:* User must agree to data collection
3) *Minimize Collection of Personal Data:* Only collect data necessary for functionality of technology, use data as close as possible to the point of collection
4) *Minimize Identification of Data with Individuals:* Anonymize data wherever possible, separate functional records and personally identifying records.
5) *Minimize and Secure Data Retention:* Store data only if necessary and in a way that is not useful in any other context. Notify user if data is lost or stolen.

The system proposed in this paper will take these design principals into account.

B. Multi-resolution load data representation

To preserve users' privacy, the resolution of load data generated by a Smart Meter can be reduced. As different use cases within the Smart Grid require different resolutions, it is difficult to determine a resolution suitable for all use cases. In addition, according to the framework for privacy aware design proposed by Wicker et al. in [7], there is no need for entities to get access to load data in a higher resolution than actually needed. To solve this problem, Engel [6] proposes to provide a Smart Meter's load data in multiple resolutions. Access to a certain resolution is only granted according to an entity's need. Furthermore, the user can decide, if access to a certain resolution is granted or revoked. Engel [6] suggests to use the wavelet transform based on the Haar wavelet and lifting scheme. The Haar wavelet calculates averages and deltas recursively, therefore adding only low computational costs. In addition, transformation is lossless and preserves the aggregate, meaning the whole consumption can be calculated using any resolution.

C. Key Management

To ensure message integrity and prevent eavesdropping, a secure way for communication between the single nodes is required within a Smart Grid. A system guaranteeing both, integrity and confidentiality for the communication channel and authentication and authorization for accessing provided services has to be implemented. A key management system can be seen as the base of such a system.

In the literature, there are different approaches on how to design a key management scheme for a secure communication within a Smart Grid.

Long et al. [8] propose an encryption scheme based on shared secrets. They divide the Smart Grid control architecture into two levels, each with its own key management system, tailored to the computational resources of the devices. While, at a first glance, shared keys seem to be an easy solution, within a growing infrastrucutre, the number of keys is growing rapidly. Every entity has to maintain one key per secure connection to another entity. Hence, causing high efforts for key management, renewal and distribution.

To solve this key management issue and to keep the number of secret keys to a minimum, the use of public keys is recommended. As Smith points out in [9], due to the use of digital signatures enabled by public key cryptography, the secret known by each device cuts down to exactly one, its own private key. Public key cryptography needs a Public Key Infrastructure (PKI) used for establishing, maintaining and distributing the public/private key pair and its assignment to a certain identity. According to Smith, a PKI doesn't have good scalability properties. Therefore, deploying a PKI within a Smart Grid Infrastructure can raise serious issues on how to manage a vast amount of Certification Authorities (CA), maintain the trust path and on how to revoke already issued certificates.

To address these scalability issues, in [10], Baumeister proposes a PKI using multiple CAs, including a CA acting as a bridge between different PKIs. Baumeister also pointed out that several PKIs have been standardized and widely accepted for many years, hence guaranteeing reliability, stability and security.

The same CA topology is also recommended by the United States National Institute of Standards and Technology in [11]. It suggests that every Grid Operator maintains its own PKI based on a hierarchical CA topology. Compatibility, communication and policy enforcement between different PKIs are ensured using bridges.

Through compromising the private key or changing certificate information, a certificate can become invalid before its lifetime is over, in which case it must be revoked. A PKI can publish revoked certificates in a Certificate Revocation

List (CRL). During the verification of a certificate, each entity has to download the CRL to check if the certificate is listed and therefore being revoked. CRLs tend to be large files generating high overhead and hence are hard to process for low resource entities. [12]

A better solution is the implementation of the Online Certificate Status Protocol (OCSP). During certificate validation, the entity sends a query about the revocation status of the certificate to a OCSP server. The provided information is up to date and communication overhead is reduced. The accessibility of the OCSP server can result in a high availability issue. OCSP stapling[1] can be used to solve this problem. An entity obtains a OCSP response for its own certificate and provides the cached response to any entity requesting the certificate. [11], [12]

D. Hierarchical Key Generation

Already in 1981, Lamport [14] suggested to use a hash chain generating a series of One Time Passwords (OTP) to address the problem of identification by sending a secret password over an insecure communication channel. To construct a hash chain of length N, a one-way hash function F is applied to an initial seed value s N-times.

$$F^2(s) = F(F(s)) \tag{1}$$

$$F^N(s) = F(F^{N-1}(s)) \tag{2}$$

F^N is used as the initial value and therefore sent to the server in a secure way. The remaining OTPs $F^1...F^{N-1}$ are stored in a secure manner on the client. The client can use F^{N-1} as the next OTP. Knowing F^N, the server can verify the OTP by calculating $F^N = F(F^{N-1})$, but neither the server nor any eavesdropper can determine the next valid OTP as F is a one-way hash function. After a successful authentication, the server stores F^{N-1} as the next value to compare with and F^{N-2} is used for the next authentication attempt. The S/KEY One-Time Password System is one example on how to use OTP for authentication [15].

The idea of hash chains can be found in many security systems [16]. Hash chains or hash trees are also used for access control to JPEG2000 coded images or H.264/scalable coded video (H.264/SVC) [17]–[19].

Imaizumi et al. propose a scheme for hierarchical access control to JPEG2000 coded images in [17]. Image properties are encrypted with different keys. According to the keys gained, a certain resolution or property can be decrypted. To minimize the number of managed keys, a hierarchical key management is introduced. All keys used are derived from one managed master key using hash chains and cyclic shifts. For decryption, the key for the highest resolution, is used. As the used hash function is no secret, the keys needed to decrypt the requested resolution can be derived from the one key provided. It is impossible to decrypt the image in a higher resolution, as the needed keys can't be derived from the one provided. In [18], Wu et al. propose a similar system for access control to JPEG2000 coded images.

In [19], Asghar et al. suggest to use key derivation for encrypting multi-layered coded video (H.264/SCV). The aim is the same as with Imaizumi et al. [17]. A user should be able to watch his/her subscribed layer data with just holding one key. For key generation and distribution, Asghar et al. use the Multimedia Internet Keying Protocol (MIKEY) [20]. Key derivation is done within the MIKEY key generation process. After key generation and distribution, an Advanced Encryption Standard - Counter Mode (AES-CM) Cipher algorithm is used for encryption.

Access control to a multi-resolution representation of load data has similar requirements as for JPEG2000 coded images or H.264/SCV encoded videos. Techniques used for these use cases can be adopted to the Smart Grid. As many successful security systems build on hash chains and one-way hash functions, they can be seen as well-established and secure.

III. Smart Grid Communication Infrastructure

To preserve privacy and to ensure secure communication, a system guaranteeing integrity, confidentiality, and authentication is needed within the smart grid. Encrypted communication between two entities must be confidential, therefore no other entity should be capable of decrypting this communication channel. In addition, third party entities should also be able to use provided services if access is granted to them. It is essential that the system is designed following the framework for privacy aware design proposed in [7]. Each entity should only have access to services and resources on a need to know basis. Information is only stored as long as needed and the user has to be informed how his/her data is being used. Access should be granted on an opt-in basis as opposed to the more prevalent (and less privacy-enabling) opt-out basis.

Possible attacks on the Smart Grid Communication Infrastructure can come from many different sides, namely the user or neighbor, the Grid Operator, Utility or any third party with or without intended access to the Grid. Independent of their origin, attacks can be classified into following groups: altering/forging messages, eavesdropping, data misusage, altering firmware or theft of private keys and denial of service. The approach proposed in this paper addresses these attacks by relying on well-established techniques for content and communication encryption. Hence these techniques can be assumed to be safe.

In Section II, different approaches on designing a suitable key management system for the Smart Grid have been discussed. A PKI is the only suitable key management system with the capability to manage a big infrastructure with a vast amount of issued certificates. The approach proposed in this paper relies on a certificate based Public Key Infrastructure (PKI). Several PKIs are standardized and well-established, therefore guaranteeing reliability and security. This approach also allows third parties to access the Smart Grid Infrastructure in a secure manner.

[1] see RFC 4366 [13]

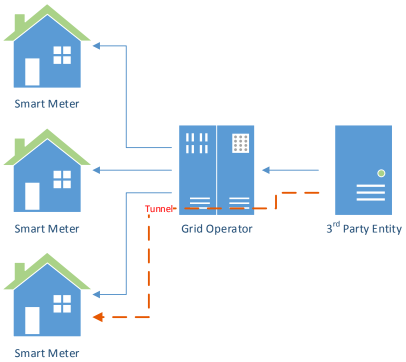


Fig. 1. Smart Meters are connected directly to the Smart Grid Operator. Third Party Entities can access a Smart Meter only via the Smart Grid Operator

The proposed PKI is managed by the Grid Operator and uses bidges to enable communication with other PKIs, therefore simplifying the certificate management as well as the trust path. Each entity acting within the Smart Grid needs to have a valid certificate proving its identity.

The Smart Meter plays a main role in the proposed system and is therefore a trusted device. A Smart Meter must be capable to generate strong keys and store these keys in a manner, that they can't be read or altered from outside. In addition, a Smart Meter must be able to compute cryptographic functions. Like suggested by the United States National Institute of Standards and Technology (NIST) [11] and Wicker et al. [7], a Hardware Security Module (HSM) or a Trusted Platform Module (TPM) can be used to fulfill these requirements. Another requirement is tamper resistance. It must be guaranteed that nobody can intrude or tamper with the Smart Meter without authorization. This embraces changes in hardware as well as in software/firmware. For identification and content encryption, each Smart Meter holds a valid certificate including a private/public key pair.

The assumed Smart Grid Infrastructure is shown in Figure 1. Smart Meters are connected directly to the Grid Operator. Third parties can access the Smart Meter via an Application Programming Interface (API) provided by the Grid Operator. This approach has two benefits: On the one hand, the Grid Operator can choose the technology on how to communicate with the Smart Meters. On the other hand, not exposing Smart Meters directly to a public network improves security as the Grid Operator can act as firewall only allowing authorized entities to communicate with the Smart Meters. Smart Meters are devices with low computational power, vulnerable to Denial of Service Attacks (DoS Attacks). An attack can result in serious issues on grid balancing and pricing. Monitoring and blocking unauthorized traffic by the Grid Operator is an essential part of increasing reliability and availability within the Smart Grid Infrastructure.

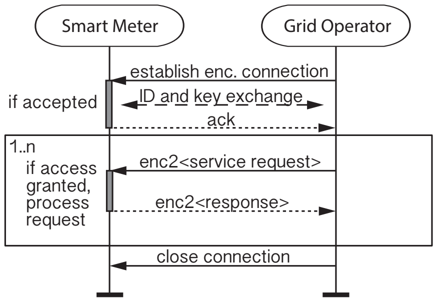


Fig. 2. The Grid Operator can communicate with the Smart Meter using an encrypted connection.

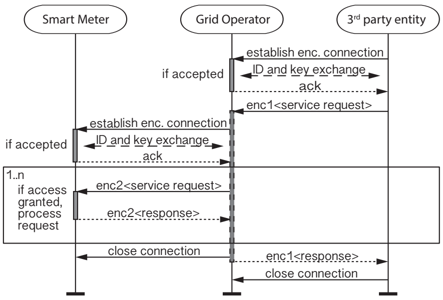


Fig. 3. To establish a connection with a Smart Meter, a third party has to send the request to the Grid Operator. If access is granted by the Grid Operator, it requests the resource from the Smart Meter. If the request is accepted by the Smart Meter, too, the Smart Meter processes the request and sends the response back to the Grid Operator. The Grid Operator forwards the response to the third party entity. Note: encrypted communication is established between third party entity and Grid Operator (enc1) and Grid Operator and Smart Meter (enc2). Hence, the Grid Operator can read the response. Content encryption has do be applied in addition, if necessary.

Figure 2 shows the communication sequence for establishing a connection between the Smart Grid Operator and a Smart Meter. First, the Grid Operator establishes an encrypted connection to the Smart Meter using Transport Layer Security (TLS)[2]. The Smart Meter accepts the connection if the Grid Operator provides a valid certificate. As soon as the encrypted connection is established successfully, the Grid Operator can use the Smart Meter's API to place a service request. If the Grid Operator has permission to access the service, the Smart Meter processes the request and sends the result back to the Grid Operator. The Grid Operator can place multiple service requests. The Grid Operator closes the connection as soon as the connection is not needed any more.

Whereas the Grid Operator can connect directly to a Smart Meter, third party entities must connect via the Grid Operator's API with the Grid Operator acting as a proxy. As shown in Figure 3, first the third party entity establishes an

[2] see IETF RFC 5246 [21]

encrypted connection to the Grid Operator and identifies itself. If the third party entity hast permission to access the Smart Grid Infrastructure, the Grid Operator accepts the connection. Now, using the encrypted channel, the third party sends a service request including the target Smart Meter ID to the Grid Operator. After verifying the service request, the Grid Operator establishes an encrypted connection to the Smart Meter and forwards the service request. Based on the third party entity's certificate, the Smart Meter grants or denies access to the requested service. If access is granted, the Smart Meter processes the request and sends the response back to the Grid Operator. The Grid Operator then forwards the response to the third party entity. The third party can place multiple service requests. As soon as the connection is not needed any more, the Grid Operator closes the encrypted connection to the Smart Meter and the third party entity closes the encrypted connection to the Grid Operator. Note that an encrypted communication is established between the third party entity and the Grid Operator as well as between the Grid Operator and the Smart Meter. Since these two connections are independent, the Grid Operator can read the whole communication between third party entity and Smart Meter. The proposed sequence only guarantees communication encryption preventing eavesdropping. For content encryption and hence privacy protection, the Smart Meter can encrypt the response using the third party entity's public key. An example for content encryption is given in Section IV. It is necessary for grid stability and reliability to differ between communication and content encryption. Within the Smart Grid, there are multiple data flows used for load balancing and controlling/managing the grid. Intruding and altering these data flows can cause severe damage to the grid. Hence, it is necessary that the Grid Operator can monitor and control the data flows within the Smart Grid Infrastructure, requiring the Grid Operator to read the sent messages. For data flows containing private information, content encryption has to be applied, preventing the Grid Operator from reading these data flows. However, it must be ensured, that these data flows can not harm the grid.

The Smart Grid is aiming to alter consumption behavior by providing fine-grained pricing information to the consumer encouraging the consumer to use energy when it is cheapest. Therefore, Wicker et al. [7] propose to broadcast real-time pricing information to the Smart Meters. Each Smart Meter is therefore accumulating price-weighted consumption data. The Electricity Provider can than access the aggregate on a daily, weekly or monthly basis. This proposal ensures protection of consumers' privacy and also fits perfectly in the scheme, proposed in this paper.

Access to and encryption of load data is discussed in Section IV.

IV. Load data encryption and distribution

As discussed in Section II, Engel et al. propose a multi-resolution representation of load data to increase privacy [6], [22]. Access to a certain resolution is based on the conditional access paradigm. A given entity is granted access to a resolution necessary to fulfill its role. As a NILM or NALM algorithm needs high resolution data to achieve accurate results, reducing the resolution of the provided load data reduces the potential for abuse. In addition, the consumer can decide, which entity is granted access to a certain resolution. This adds another degree of freedom as entities have to explain their data usage to gain users' trust.

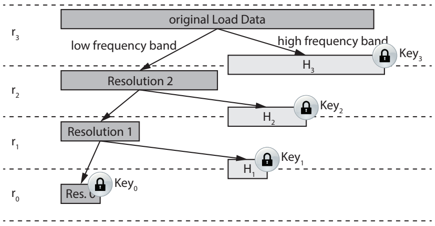


Fig. 4. The Wavelet transform splits load data into high and low frequency bands. The low frequency band equals load data with reduced resolution.

Load data can be represented in multiple resolutions using a suitable wavelet transform, as suggested by Engel et al. in [6]. The Haar wavelet transform suits the requirements best. It consists of calculating averages and deltas, therefore needing few computational resources. The Haar wavelet is a lossless transform; under each resolution, the total consumption over the whole timespan can be derived.

The wavelet transform splits load data in a high and low frequency band recursively up to a certain level. Where the low frequency band is used for the next recursive operation, the high frequency band is preserved. The low frequency band represents the data at a certain resolution with half the number of samples of the next higher resolution. The high frequency band represents the delta of a sample to the according sample of the low frequency band. The values from the high frequency band and the remaining value from the low frequency band are called wavelet coefficients. The wavelet coefficients are needed to do the inverse wavelet transform and restore the load data to a certain resolution. The described steps can be seen in Figure 4.

To restore a certain resolution, the inverse wavelet transform is performed on the low frequency band and its according high frequency band. The inverse starts with the coefficients of the lowest resolution and works its way up to the desired resolution.

To fulfill the conditional access paradigm introduced prior in this section, wavelet coefficients have to be encrypted with a different key for each resolution (from now on resolution key). Granting access to a certain resolution means to distribute the resolution keys for the certain resolution and for all lower resolutions to the requesting entity. A high number of resolution keys has to be managed and distributed, therefore introducing significant overhead for key management and storage.

To address the problem of high key management costs, Hierarchical Keys are introduced. Hierarchical Keys allow the decryption of multiple ciphertexts with a single key although the messages were encrypted with different keys. For example encrypting three messages m_1, m_2, m_3 each with a different hierarchical key k_1, k_2, k_3. In terms of decryption, using k_1 just decrypts m_1, but k_2 or k_3 can be used to decrypt m_1, m_2 or m_1, m_2, m_3, respectively. Hierarchical Keys therefore simplify key management, as less keys have to be known to decrypt multiple messages. Key generation and sample use cases have been already discussed in Section II.

As the use case of multi-resolution representation of load data is quite similar to H.264/SVC and JPEG2000 encryption, techniques proposed in [17]–[19] can be adopted. A hierarchical resolution key is generated for each level of resolution. Resolution keys are derived from a master key using hash chains. Any appropriate one-way hash function can be used. Resolution key renewal can be done within a certain time period, e.g., daily. Wavelet coefficients are encrypted using the appropriate resolution key. The wavelet transform itself is performed on a cyclic basis, e.g., hourly, covering a fixed time span, e.g., the last 24 hours. The wavelet coefficients are packed into a single stream (see Figure 5) and transfered to any entity requesting it. According to the entity's resolution key, the entity is only capable to decrypt the wavelet coefficients of the resolution, to which access was granted to. As the one-way hash function is no secret, the entity can derive the resolution keys to encrypt the wavelet coefficients of a lower encryption but it can't encrypt any wavelet coefficients of a higher resolution.

Figure 6 shows the service requests needed for obtaining load data. This sequence is based on the communication sequence shown in Figure 3. Before sending a service request to the Smart Meter, the entity has to establish a connection via the Grid Operator, as described in section III. To obtain load data, the entity has to go through two steps, (i) obtaining a suitable resolution key and (ii) retrieving the load data. To obtain the resolution key, the entity has to request access for a certain resolution. Therefore, it sends a service request including the certificate and the requested resolution to the Smart Meter. The Smart Meter has to decide, if access is granted. If this is the entity's first access request, the Smart Meter forwards the request to the consumer as he/she can decide, if access for a certain resolution is granted to a certain entity. If the entity is known by the Smart Meter, access can be granted/denied based on previous consumer decision. In case access is granted, the Smart Meter encrypts the resolution key using the entities public key and sends it to the entity. In a second step, the entity sends a load data request to the Smart Meter. The Smart Meter returns a stream containing the encrypted wavelet coefficients, as shown in Figure 5. There is no additional authentication process needed, as the stream is worthless without the resolution key obtained in step one. By decrypting the wavelet coefficients and performing an inverse wavelet transform, the entity can now restore load data up to the resolution, to which access was granted. Load data can be obtained as long as the issued resolution key is valid. To ensure content security, the resolution key is encrypted using the requesting entity's public key. As only the entity knows it's private key, the resolution key cannot be decrypted by the Grid Operator working as a proxy.

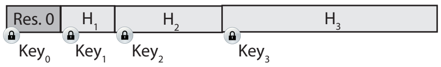


Fig. 5. All wavelet coefficients needed for the inverse transformation are encrypted with different keys and transmitted as a single stream.

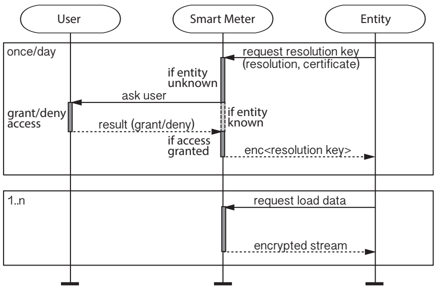


Fig. 6. To access load data, the entity has to request the resolution key for the desired resolution. The user has to decide, if access is granted or denied. After receiving a valid resolution key, the entity can request load data as long as the resolution key is valid. To guarantee content security, the resolution key is encrypted using the entity's public key.

V. Conclusion

Secure communication plays an important role within the Smart Grid. It is essential to ensure authentication, authorization and integrity to prevent unauthorized parties from eavesdropping or altering communication. As consumer related data is collected and transferred, privacy protection is another important issue to address.

In this paper, a secure way of communication, suitable to be used within a Smart Grid Infrastructure, has been introduced. The approach uses a PKI to ensure a secure communication between Smart Meters, the Grid Operator and third party entities. For communication between third party entities and Smart Meters, the Grid Operator acts as a proxy. Hence, the Grid Operator protects the Smart Grid Infrastructure from possible attacks.

To preserve privacy, load data is represented in multiple resolutions. The consumer can decide which entity can access data and at which specific resolution. For multi-resolution representation, the wavelet transform is used, as it adds just a small computational overhead and the transformation process is lossless. Each resolution is encrypted using a different resolution key. Key management efforts are reduced by introducing a hierarchical key management using one-way hash functions for key derivation.

The proposed scheme offers a secure way of communication within the Smart Grid. Methods are used to preserve consumer's privacy. A new degree of consumer freedom is added, as the consumer can decide to whom and at what level his or her personal data can be provided.

Acknowledgment

Financial support by the Austrian Marshall Plan Foundation as well as by the Austrian Federal Ministry of Economy, Family and Youth and the Austrian National Foundation for Research, Technology and Development is gratefully acknowledged. Also, funding by the Austrian Federal Ministry for Transport, Innovation and Technology and the Austrian Research Promotion Agency (FFG) under Project 838793, "INTEGRA", is gratefully acknowledged. Thanks are due to Thomas Stütz for his clarifying comments. Thanks are also due to the School of Electrical and Computer Engineering at Cornell University for hosting and supporting the first author.

References

[1] "A national assessment of demand response potential," in *Staff Report, Federal Energy Regulatory Commission*, June 2009. [Online]. Available: http://www.ferc.gov/legal/staff-reports/06-09-demand-response.pdf

[2] M. Lisovich and S. Wicker, "Privacy concerns in upcoming residential and commercial demand-response systems," in *Clemson University Power Systems Conference 2008*. Clemson University, March 2008. [Online]. Available: http://www.truststc.org/pubs/332.html

[3] G. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec 1992.

[4] G. Eibl and D. Engel, "Influence of data granularity on nonintrusive appliance load monitoring," in *Proceedings of the Second ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec '14)*. Salzburg, Austria: ACM, 2014, pp. 147–151. [Online]. Available: http://doi.acm.org/10.1145/2600918.2600920

[5] Z. Erkin, J. Troncoso-Pastoriza, R. Lagendijk, and F. Perez-Gonzalez, "Privacy-preserving data aggregation in smart metering systems: an overview," *Signal Processing Magazine, IEEE*, vol. 30, no. 2, pp. 75–86, 2013.

[6] D. Engel, "Wavelet-based load profile representation for smart meter privacy," in *Proceedings IEEE PES Innovative Smart Grid Technologies (ISGT'13)*, Washington, D.C., USA, Feb. 2013, pp. 1–6.

[7] S. Wicker and R. Thomas, "A privacy-aware architecture for demand response systems," in *Proceedings 44th Hawaii International Conference on System Sciences (HICSS'11)*, Jan 2011, pp. 1–9.

[8] X. Long, D. Tipper, and Y. Qian, "An advanced key management scheme for secure smart grid communications," in *Proceedings IEEE International Conference on Smart Grid Communications (SmartGridComm'13)*, Oct 2013, pp. 504–509.

[9] S. Smith, "Cryptographic scalability challenges in the smart grid (extended abstract)," in *Proceedings IEEE PES Innovative Smart Grid Technologies (ISGT'12)*, Jan 2012, pp. 1–3.

[10] T. Baumeister, "Adapting PKI for the smart grid," in *Proceedings IEEE International Conference on Smart Grid Communications (SmartGridComm'11)*, Oct 2011, pp. 249–254.

[11] NIST, "Smart grid cybersecurity strategy, architecture, and high-level requirements," in *NISTIR 7628 Guidelines for Smart Grid Cybersecurity*. National Institute of Standards and Technology, U.S. Department of Commerce, 2013, vol. 1.

[12] J. Buchmann, E. Karatsiolis, and A. Wiesmaier, *Introduction to Public Key Infrastructures*. Springer Berlin Heidelberg, 2013. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-40657-7_5

[13] S. Blake-Wilson, M. Nystrom, D. Hopwood, J. Mikkelsen, and T. Wright, "Transport Layer Security (TLS) Extensions," in *IETF Request for Comments*, no. 4366, April 2006. [Online]. Available: http://www.ietf.org/rfc/rfc4366.txt

[14] L. Lamport, "Password authentication with insecure communication," *Commun. ACM*, vol. 24, no. 11, pp. 770–772, Nov. 1981. [Online]. Available: http://doi.acm.org/10.1145/358790.358797

[15] N. Haller, "The S/KEY One-Time Password System," in *IETF Request for Comments*, no. 1760, 1995. [Online]. Available: http://www.ietf.org/rfc/rfc1760.txt

[16] V. Goyal, "How to re-initialize a hash chain," Cryptology ePrint Archive, Report 2004/097, 2004.

[17] S. Imaizumi, M. Fujiyoshi, H. Kiya, N. Aoki, and H. Kobayashi, "A key derivation scheme for hierarchical access control to JPEG2000 coded images," in *Advances in Image and Video Technology*, ser. Lecture Notes in Computer Science, Y.-S. Ho, Ed. Springer Berlin Heidelberg, 2012, vol. 7088, pp. 180–191. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-25346-1_17

[18] Y. Wu, D. Ma, and R.-H. Deng, "Progressive protection of JPEG2000 codestreams," in *International Conference on Image Processing (ICIP '04)*, vol. 5, Oct 2004, pp. 3447–3450 Vol. 5.

[19] M. Asghar and M. Ghanbari, "Cryptographic keys management for H.264/scalable coded video security," in *8th International ISC Conference on Information Security and Cryptology (ISCISC'11)*, Sept 2011, pp. 83–86.

[20] J. Arkko, E. Carrara, F. Lindholm, M. Naslund, and K. Norrman, "MIKEY: Multimedia Internet KEYing," in *IETF Request for Comments*, no. 3830, 2004. [Online]. Available: http://tools.ietf.org/html/rfc3830

[21] T. Dierks and E. Rescorla, "The Transport Layer Security (TLS) Protocol Version 1.2," in *IETF Request for Comments*, August 2008.

[22] D. Engel and G. Eibl, "Multi-resolution load curve representation with privacy-preserving aggregation," in *Proceedings of IEEE Innovative Smart Grid Technologies (ISGT) 2013*. Copenhagen, Denmark: IEEE, Oct. 2013, pp. 1–5.

3.12 LAUSENHAMMER15A

- W. Lausenhammer, D. Engel, and R. Green. A game theoretic software framework for optimizing demand response. In *Proceedings of the 6th Conference on Innovative Smart Grid Technologies (ISGT)*, pages 1–5, Feb 2015.

A Game Theoretic Software Framework for Optimizing Demand Response

Wolfgang Lausenhammer and Dominik Engel
Josef Ressel Center for User-Centric Smart Grid Privacy, Security and Control
Salzburg University of Applied Sciences
Salzburg, Austria
{wolfgang.lausenhammer,dominik.engel}@en-trust.at

Robert Green
Dept. of Computer Science
Bowling Green State University
Bowling Green, OH 43403
greenr@bgsu.edu

***Abstract*—Demand response (DR) is a crucial and necessary aspect of the smart grid, particularly when considering the optimization of both, power consumption and generation. While many benefits of DR are currently under study, an issue of particular concern is optimizing end-users' power consumption profiles at various levels. This study proposes a fundamental, game theoretic software framework for DR simulation that is capable of investigating the effect of optimizing multiple electric appliances by utilizing game theoretic algorithms. Initial results show that by shifting the switch-on time of three household appliances provides a savings of up to 6%.**

I. Introduction

With the increasing pervasiveness of renewable energy new challenges have arisen: Energy is no longer exclusively produced in large power plants, but also in the homes of ordinary people. Eventually, this development will lead to a paradigm shift, away from the hierarchical top-down oriented system to a decentralized structure with volatile renewable energy sources, such as wind turbines, photovoltaic cells and plug in electric vehicles (PEV) [1,2].

By coordinating household appliances and PEVs, off-peak usage could result in cheaper electricity prices. With respect to coordination, demand response (DR) management could pose an ideal solution to this problem [3-5].

Within the vast amount of different approaches to simulate and model DR, game theory proves to be a capable method of modeling and describing complex interactions between different rational players. The goal of a game theoretic approach in DR management is to develop a model and proof that if every agent tries to maximize its own profit, an equilibrium point is found. By acting selfishly, players reach a global optimum [6]. Publications in this area range from load shifting approaches [7,8] to using storage devices such as PEVs in micro-grid storage games [9] to games that focus on utility companies [10,11].

This study proposes a new software framework – *Okeanos*[1] – that enables the simulation and study of these issues through the provision of an extensible, open source simulation platform that can both, model different types of loads and be configured to work with different game theoretic DR management approaches. By providing a very lightweight interface for users to plug in their own control algorithms, the framework also allows for new strategies to be tested.

The remainder of the paper is structured as follows: An overview of related work in game theory in DR management, and software frameworks for DR management is presented in Section II; This is followed by introducing the novel DR simulation platform, *Okeanos*, and highlighting its key concepts in Section III; Initial results of using *Okeanos* are described in Section IV; and, finally, Section V concludes this work.

II. Literature Review

This section provides an overview of the state-of-the-art in game theory in DR management and software frameworks for DR management.

A. Game Theory in Demand Response Management

Game theory, in its essence, aims to help understand situations in which several decision-makers interact. Being a mathematical framework and analytical tool, game theory helps study the relationships and actions among rational players [6,12].

Saad and co-authors evaluate the available approaches for applying game theory to timely open and relevant smart grid related problems in [6]. They focus on three emerging areas, particularly: micro-grid systems, demand-side management, and communications.

One way to reduce the *peak-to-average-ratio* (PAR) of an energy system is to change the schedule of shiftable household appliances. Traditionally, multi-objective functions [13] and non-linear models [14,15] are used to determine a (near) optimal, (near) real-time schedule.

In contrast to that, Mohsenian-Rad and co-authors utilize game theory and propose an energy consumption game to optimize energy costs in [7]. Their aim is to change the daily schedule of shiftable household appliances. Although the schedule could be calculated centrally, calculation is done in a decentralized way. This is the preferable way, as it requires significantly less communication effort and does not provide a single point of failure [7].

[1] The project is released as open source and can be accessed at https://github.com/wolfgang-lausenhammer/Okeanos

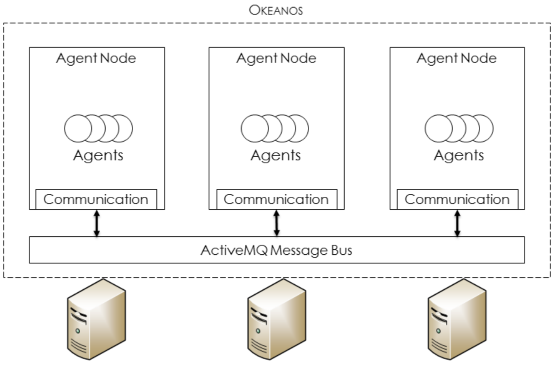


Fig. 1. Physical mapping of agents in Okeanos.

Unlike the aforementioned load shifting approach, the authors of [9] propose a non-cooperative micro-energy-storage game. Here, users decide on a storage profile for their household devices to optimize a utility function reflecting the cost. That is, users decide on a point in time when they want to buy energy, ideally during low-cost periods, and when they want to use this energy to satisfy their demand. Several constraints, such as the maximum capacity, the storage efficiency and running costs are considered [9].

B. Software Frameworks for Demand Response

Recent [16-20], as well as older publications [21-25], propose a multi-agent approach as appropriate to deal with the complex topic of demand-response optimization.

PowerMatcher is a tool used for coordinating a large cluster of distributed energy resource devices within a smart grid in near real-time. Its focus is on end-consumers allowing them to use their appliances to actively participate in the energy market. Thus, by offering flexibility to the grid, customers get the possibility to reduce their energy bills. Appliances, represented by agents, coordinate consumption and production and calculate the market clearing price [16].

Similarly, in DEZENT, agents communicate their demand or supply to a balancing group manager (BGM) at their respective grid hierarchy level. It is the BGMs duty to match similar offers for demand and supply. Unmatched offers are handed over to the next higher level, where this matching starts over again [26].

Okeanos is fundamentally different to these approaches, as it plans consumption and production ahead of time and uses mathematically proved solutions for finding the optimal schedule for household appliances. Indeed, by using game theoretic approaches, it is guaranteed that if every user acts selfishly and optimizes his or her own costs, a global cost-optimum is established.

III. OKEANOS – A MULTI-AGENT GAME THEORETIC DEMAND RESPONSE MANAGEMENT SOFTWARE FRAMEWORK

Okeanos is a novel Java-based multi-agent DR simulation platform with special focus on the compatibility to game theory. That is, not only one particular coordination mechanism as in [16], [17] or [26] is supported, but any mechanism, as long as it complies with the specified interface.

The goal is to allow for a holistic approach to demand response management with a very extensible platform that can host all kind of appliances, as long as there is an appropriate driver available in the system. By defining a clear interface and basing the framework on OSGi, these drivers can be easily developed, deployed and removed from the system.

A. Household appliances as the smallest active unit

Okeanos combines several features of [16], [17] and [26]. That is, *Okeanos* utilizes the multi-agent paradigm to represent household appliances. Thus, every single appliance within a household taking over an active role in DR management is represented by an agent. This implies that every agent can decide on its own and can pursue a target. The target is specified by the currently plugged in game. Likewise, the capabilities are specified by the underlying household device and the corresponding driver, which are both explained later on in this section.

To be able to focus on other aspects of the system, a feature-rich, modularized and easy to use framework is utilized for providing multi-agent features. A comparison between the Java-based multi-agent simulation platforms JADE, *Janus*, *Jason* and JIAC resulted in JIAC as the winner. Criteria included functionality, active development, ease of use and adoption throughout the software developer community.

JIAC’s modern approach to use the *Spring* framework as the basis for the whole system is unique throughout all compared multi-agent frameworks. Additionally, the utilization of this framework assures the system to be future-proof according to the best of the authors' knowledge, thus making it first choice for implementing multi-agent systems in Java.

Okeanos, a JIAC-based multi-agent system, is structured as shown in Fig. 2. That is, the application can consist of several agent nodes, agents and agent beans [27].
Agent nodes are distributed containers providing the necessary infrastructure for agents, such as a communication infrastructure or white and yellow pages services [27].

Several agents, that is, household appliances, can be hosted within one agent node (see Fig. 1) and according to the *service oriented architecture* (SOA) architectural pattern, provide services to other agents. Moreover, as required by [27], every agent comprises several agent beans, which provide the actual functionality like persistent memory and usage of infrastructure services for inter-agent communication. As a consequence, functionality defined by *Okeanos*, such as weather service, pricing service or time service, are implemented as agent beans and OSGi bundles and plugged into the agents as such.

B. Plug in support

To be able to plug in different device drivers or games, flexible and powerful interfaces need to be developed. Furthermore, implementations of these interfaces need to be hooked into the system easily, in order to keep the threshold for developing modules as low as possible. For that reason, it is advantageous to modularize the system as much as possible.

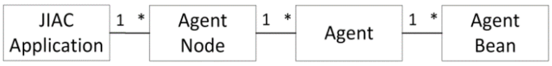


Fig. 2. Structure of a JIAC-based multi-agent system. Adapted from [27].

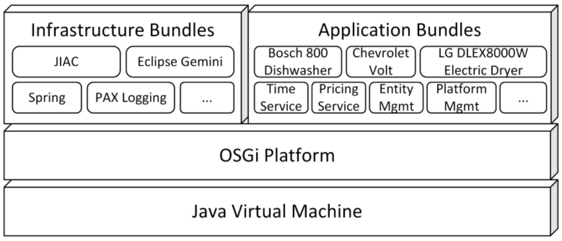


Fig. 3. Okeanos Bundle structure with sample household devices and services

With OSGi designed as a *service oriented architecture*, *Okeanos* features no monolithic core, but is a conglomerate of various bundles (see Fig. 3). According to the OSGi R5 specification [28], it is best practice to keep the interfaces in a separate bundle, to also allow for optional bundles not being present in the OSGi container. Consider, for example, a logging service: The application does not necessarily need an implementation for a correct execution, however, at least the interface needs to be present, otherwise OSGi would not be able to resolve the dependencies of the bundle.

As indicated by Fig. 3, every service in *Okeanos* is represented by its own module and, therefore, separated in its own bundle. Their respective interfaces are all consolidated in interface bundles corresponding to the actual layer the bundle is part of. Likewise, as it is possible to have no implementation present in an OSGi container, it is possible to have multiple implementations present. This is especially true for device drivers, as they all implement the same interface. Therefore, the service user needs to select from the list of available drivers.

The service provider, that is the driver, can specify additional properties, such as year and brand of a household device, as key value pairs to supply the service user with some cues.

C. Game theory in Okeanos

As described in Section II, game theory can be used to understand the result of the dynamism in a game between several interacting players. Every player in such a game is represented by its own agent in *Okeanos*. By that, the prerequisite that players have to act rationally can be assured.

There are a number of published game theoretic approaches to DR management [7-11]. Some take load shifting into consideration, some make use of available storage devices such as PEVs, and others formulate a game with multiple utility companies.

Okeanos is designed to support any game that can be mapped to the specified interface. Therefore, it is crucial to define the interface as general as possible, while at the same time being specific enough that implementations of the interface have a useful basis for doing their optimization and calculation.

Furthermore, it is possible that individual agents use different games. The meaningfulness of such a mixture is, however, questionable, as no guarantee of the existence of a Nash equilibrium can be given under such circumstances.

As a first proof of concept, the game proposed by Mohsenian-Rad and co-authors [7] has been modelled with *Okeanos*. One of the reasons for this is that the authors formulate their algorithm in pseudo code, which allows for accurate adaptation. Moreover, by utilizing load shifting, potentially more devices can be integrated in the first place as it were possible with storage devices.

The decentralized objective function of the game in [7] is given by (1) where $x_{n,a}^h$ represents a one hour energy consumption scheduled for appliance a of user n at hour h. Additionally, the cost functions C_h are increasing and strictly convex. $\mathcal{H}$ is the set of possible hours of the 24h time horizon [7].

$$\begin{array}{ll} \underset{x_{\{n\}}\in\mathcal{X}_n}{minimize} \quad \sum_{h=1}^{H} C_h\left(\sum_{a\in\mathcal{A}_n} x_{n,a}^h + \sum_{m\in\mathcal{N}\{n\}} l_m^h\right) & \\ subject\ to\ \ l_m^h = \sum_{a\in\mathcal{A}_n} x_{m,a}^h \qquad h\in\mathcal{H} & \end{array} \tag{1}$$

The consumption of all other players $l_m^h, m \in \mathcal{N} \setminus \{n\}$ is static, therefore, only the schedule $x_{n,a}^h$ of the local appliances a of player n at each hour h needs to be calculated. That is, an optimal schedule with respect to the consumption patterns of the other players needs to be computed [7].

The algorithm to play this game is given in Algorithm 1. The initial consumption is initialized randomly, because the game guarantees to find the Nash equilibrium regardless of the initial configuration. After that, every appliance finds the best solution to the local optimization problem (1) at random instances, e.g., by using the Interior Point Method (IPM). This randomness is important to allow for another appliance being faster with finding a solution and sending an update of its consumption. If a different solution to the optimal consumption is found, it is broadcast to the other devices. This loop is repeated until no schedules are changed anymore.

```
Randomly initialize $l_n$ and $l_{-n}$
repeat
  at random time instances do
    Solve local problem (1) using IPM.
    if $x_n$ changes compared to current schedule then
      Update $x_n$ according to the new solution.
      Broadcast a control message to announce $l_n$ to the other
          agents across the system.
    end
  end
  if a control message is received then
    Update $l_{-n}$ accordingly.
  end
until no agent announces any new schedule
```

Algorithm 1: Energy consumption game, executed by each user $n \in \mathcal{N}$ [7].

Mohsenian-Rad and co-authors use an hourly consumption schedule for all devices. In contrast, *Okeanos* uses a 15 minutes interval to allow for a more fine-grained control over faster devices.

IV. Initial Results

Okeanos is evaluated to test its applicability to real world problems and use cases. Therefore, this section gives an initial insight into the capabilities of Okeanos. Test results are based on the devices listed in Table 1 in addition to a household load profile. The data for implementing drivers for clothes washer, clothes dryer and a dishwasher is taken from [29]. The household data is based on the H0 load profile provided by the Bundesverband der Energie- und Wasserwirtschaft (Federal Association of the Energy and Water Industry) [30]. The H0 load profile is a standardized profile used to approximate the consumption of customers that cannot be measured otherwise.

The real-time pricing costs are taken from [31]. In order to draw a sound conclusion, the consecutively mentioned experiments were repeated at least 100 times and the reported results are average values. A single household with a 30 kWh load profile is used as a base case.

Starting with multiple devices within one household, the interaction between the devices is tested. The devices search for the point in time which minimizes the electricity costs for that device. The impact of shifting the load profile of a household is depicted in Fig. 4. Devices in the first chart run daily, whereas devices in the second run every third day. Additionally, to make the simulation more realistic and to take the consumption patterns of different households into account, the H0 load profile is shifted 0, ±1h or ±3h.

The major result of this simulation is that the more the regular households differ in their consumption patterns, the more the total load curve evens out. With all households using the standardized H0 load profile, several peaks are present, most notably those at 1 p.m. and 8 p.m. Naturally, considering the price per kWh, it is preferable to, especially at those hours, to reduce the energy consumption.

The only difference between the charts in Fig. 4 is the peak in the morning, when all the load shifting devices are switched on. This difference is due to the fact that the devices run only every third day and, therefore, on average, the consumption at that point should only be one third of that when they are switched on every day.

It can be seen in Table 2 that the effect of varying the load profile of households is negligible. This is valid throughout all compared categories.

Table 1: Overview of drivers used for evaluation. Data from [29] and [30].

Appliance	Model	Rating
Household	Standard load profile	Scaled to 30kWh
Clothes washer	LG WM2016CW	120V, 60Hz, 5A
Clothes dryer	LG DLE2516W	120V, 60Hz, 26A
Dishwasher	Kenmore 665.13242K900	120V, 60Hz, 9.6A

Table 2: Comparison of costs per month per household with load shifting in relation to shifted household load profiles.

	Regular 30kWh household	28kWh household with 2kWh load shifted devices	
		Run daily	Run every third day
0h shifting	$85.80	$82.25 (4.14%)	$80.71 (5,93%)
±1h shifting	$85.72	$82.17 (4,14%)	$80.66 (5,90%)
±3h shifting	$85.10	$81.60 (4,11%)	$80.11 (5,86%)

Actual savings, according to the outcomes (see Table 2), can be noticed between a regular 30kWh household and when load shifting is in place. The average savings is around 4.14% if load shifting is in place.

Naturally, the savings of a household with its devices running only every third day needs to be higher compared to a household where the devices run every day. The savings compared to a regular household with no load shifting are 5.9%.

V. Conclusion

In this paper, we proposed *Okeanos*, a novel multi-agent demand response simulation platform that is capable of evaluating game theoretic approaches. Due to its extensibility, *Okeanos* can support a wide range of different household appliances. Moreover, because the system is based on OSGi, exchanging specific implementations is very easy, as long as it implements the same interfaces.

Initial results show that by optimizing three household appliances of one household, Okeanos can save up to 5.9% of energy costs per month. Future work will focus on studying the impact of more households, as well as integrating plug in electric vehicles in the simulation.

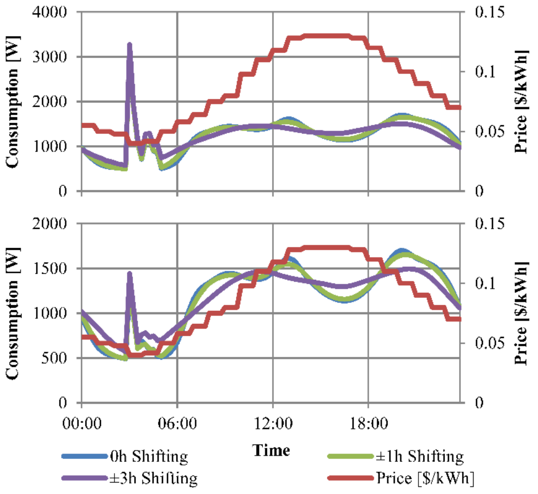


Fig. 4. Optimizing the schedule of one 28kWh/day household. Devices run every day in the first chart, every third day in the second chart.

REFERENCES

[1] P. Palensky and D. Dietrich, “Demand side management: demand response, intelligent energy systems, and smart loads,” *IEEE Trans. Ind. Informat.*, vol. 7, no. 3, pp. 381-388, 2011.

[2] J. Schwarzer, A. Kiefel and D. Engel, “The role of user interaction and acceptance in a cloud-based demand response model,” in *39th Annu. Conf. IEEE Industrial Electronics Society (IECON 2013)*, Vienna, Austria, 2013, pp. 4797–4802.

[3] Z. Fan, P. Kulkarni, S. Gormus, C. Efthymiou, G. a. S. M. a. Z. Z. Kalogridis, S. Lambotharan and W. Chin, “Smart grid communications: Overview of research challenges, solutions, and standardization activities,” *IEEE Commun. Surveys & Tutorials*, vol. 15, no. 1, pp. 21-38, 2012.

[4] A. Patel, J. Aparicio, N. Tas, M. Loiacono and J. Rosca, “Assessing communications technology options for smart grid applications,” in *2011 IEEE Int. Conf. Smart Grid Communications (SmartGridComm 2011)*, Brussels, Belgium, 2011, pp. 126-131.

[5] V. Murthy Balijepalli, V. Pradhan, S. Khaparde and R. Shereef, “Review of demand response under smart grid paradigm,” in *2011 IEEE PES Innovative Smart Grid Technologies-India (ISGT 2011)*, Kollam, Kerala, India, 2011, pp. 236-243.

[6] W. Saad, Z. Han, H. V. Poor and T. Başar, “Game-theoretic methods for the smart grid: an overview of microgrid systems, demand-side management, and smart grid communications,” *IEEE Signal Processing Mag.*, vol. 29, no. 5, pp. 86-105, 2012.

[7] A. Mohsenian-Rad, V. W. Wong, J. Jatskevich, R. Schober and A. Leon-Garcia, “Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid,” *IEEE Trans. Smart Grid*, vol. 1, no. 3, pp. 320-331, 2010.

[8] C. Ibars, M. Navarro and L. Giupponi, “Distributed demand management in smart grid with a congestion game,” in *2010 1st IEEE Int. Conf. Smart Grid Commun. (SmartGridComm 2010)*, Gaithersburg, MD, 2010, pp. 495-500.

[9] P. Vytelingum, T. D. Voice, S. D. Ramchurn, A. Rogers and N. R. Jennings, “Agent-based micro-storage management for the smart grid,” in *Proc. 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, Toronto, Canada, 2010, pp. 39-46.

[10] S. Maharjan, Q. Zhu, Y. Zhang, S. Gjessing and T. Başar, “Dependable demand response management in the smart grid: a Stackelberg game approach,” *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 120-132, 2013.

[11] S. Bu, F. R. Yu and P. X. Liu, “A game-theoretical decision-making scheme for electricity retailers in the smart grid with demand-side management,” in *2011 IEEE Int. Conf. Smart Grid Commun. (SmartGridComm 2011)*, Brussels, Belgium, 2011, pp. 387-391.

[12] C. Rieck, *Spieltheorie; Eine Einführung*, 11th ed (in German). Eschborn, Germany: Christian Rieck Verlag, 2012.

[13] M. A. A. Pedrasa, T. D. Spooner and I. F. MacGill, “Scheduling of demand side resources using binary particle swarm optimization,” *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 1173-1181, 2009.

[14] P. Faria, Z. Vale, J. Soares and J. Ferreira, “Demand response management in power systems using a particle swarm optimization approach,” *IEEE Intell. Syst.*, vol. 28, no. 4, pp. 45-51, 2011.

[15] N. Gudi, L. Wang, V. Devabhaktuni and S. S. S. R. Depuru, “Demand response simulation implementing heuristic optimization for home energy management,” in *North American Power Symposium (NAPS 2010)*, 2010, Arlington, TX, 2010, pp. 1-6.

[16] K. Kok, “The PowerMatcher: Smart Coordination for the Smart Electricity Grid,” Ph.D. dissertation, Free Univ. Amsterdam, Amsterdam, Netherlands, 2013.

[17] O. Struß, “Open-Source-Modellierung und auktionsorientierte Regulierung dezentraler Energienetze,” (in German), diploma thesis, Dept. Computer Science and Mathematics, Univ. Bremen, Bremen, Germany, 2012.

[18] S. Rohjans, S. Lehnhoff, S. Schütte, S. Scherfke and S. Hussain, “mosaic — A modular platform for the evaluation of agent-based Smart Grid control,” in *4th European Innovative Smart Grid Technologies (ISGT 2013)*, Lyngby, Denmark, 2013, pp. 1-5.

[19] S. Lehnhoff, *Dezentrales vernetztes Energiemanagement: ein Ansatz auf Basis eines verteilten adaptiven Realzeit-Multiagentensystems* (In German). Heidelberg, Germany: Springer DE, 2010.

[20] K. Kok, G. Venekamp and P. Macdougall, “Market-based control in decentralized electrical power systems,” in *1st Int. Workshop on Agent Technologies for Energy Systems (ATES 2010)*, Toronto, Canada, 2010, pp. 61-66.

[21] R. Gustavsson, “Agents with power,” *Commun. of the ACM*, vol. 2, no. 3, pp. 41-47, 1999.

[22] P. Carlsson, “Algorithms for electronic power markets,“ Ph.D. dissertation, Uppsala Univ., Uppsala, Sweden, 2004.

[23] F. Ygge and H. Akkermans, “Power load management as a computational market,” in *2nd Int. Conf. Multi-Agent Systems (ICMAS 1996)*, Menlo Park, CA, 1997, pp. 393-400.

[24] K. Kok, Z. Derzsi, M. Hommelberg, C. Warmer, R. Kamphuis and H. Akkermans, “Agent-based electricity balancing with distributed energy resources, a multiperspective case study,” in *Proc. 41st Annu. Hawaii Int. Conf. on System Sciences (HICSS 2008)*, Waikoloa, HI, 2008.

[25] K. Kok, C. Warmer and I. Kamphuis, “PowerMatcher: multiagent control in the electricity infrastructure,” in *Proc. 4th Int. Joint Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2005)*, Utrecht, Netherlands, 2005, pp. 75–82.

[26] S. Lehnhoff, O. Krause, C. Rehtanz and H. Wedde, “Distributed autonomous power management-For a reliable operation under feasibility constraints,” *Automatisierungstechnik*, vol. 59, no. 3, pp. 167-179, 2011.

[27] JIAC Development Team, “JIAC - Java Intelligent Agent Componentware,” DAI-Labor, Tech. Univ. Berlin, Berlin, Germany, Manual. Version 5.1.3, 2012.

[28] *OSGi Core Release 5*, OSGi Alliance, San Ramon, CA, 2012.

[29] Pipattanasomporn, M., Kuzlu, M., Rahman, S., and Teklu, Y., “Load profiles of selected major household appliances and their demand response opportunities", *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 742–750, 2014.

[30] Bundesverband der Energie- und Wasserwirtschaft, “Standardlastprofil H0: Haushalt, Privatverbrauch, gegebenenfalls geringfügig gewerblicher Bedarf”, (in German), 2014.

[31] Office of the Ohio Consumers' Counsel, “Smart grid: dynamic and time-of-use pricing", Office of the Ohio Consumers' Counsel, Columbus, OH, Tech. Rep., 2011.

3.13 EIBL15B

- G. Eibl, D. Engel, and C. Neureiter. Privacy-relevant smart metering use cases. In *Proceedings of IEEE International Conference on Industrial Technology (ICIT) 2015*, pages 1387–1392, Seville, Spain, 2015. IEEE.

Privacy-Relevant Smart Metering Use Cases

Günther Eibl, Dominik Engel and Christian Neureiter
Josef Ressel Center for User-Centric Smart Grid Privacy, Security and Control
Salzburg University of Applied Sciences
Urstein Sued 1, A–5412 Puch/Salzburg, Austria
Email: {guenther.eibl, dominik.engel, christian.neureiter}@en-trust.at

***Abstract*—Privacy in Smart Metering has been discussed extensively, as have privacy enhancing technologies (PETs). However, neither of these items has been put into the perspective of the actual use cases at hand. This perspective is crucial to (i) map the correct PETs to each use case and to (ii) identify gaps, i.e., use cases which are privacy-relevant, but not yet covered by a PET. Beside the construction of such a set of privacy-relevant smart metering use cases, some open research questions have been found. Most importantly, the Smart Metering systems must be described in more detail in order to facilitate a sound development of PETs.**

I. Introduction

Smart Metering, as part of the Smart Grid, is a step towards modernizing our electrical grids. However, the discussion of how to achieve optimal roll-outs of smart meter technology has been accompanied by a – sometimes ferocious – debate on privacy concerns. Numerous contributions have pointed out that the load consumption data produced by a household is privacy-sensitive data, as it allows to deduce behavioral patterns of its inhabitants (e.g., [1], [2], [3]).

Privacy enhancing technologies (PETs) have been proposed to strike a balance between the functional requirements of Smart Metering and the requirement of preserving individual privacy. This paper focuses on methods that are applied near to the customer and that aim at providing the minimum amount of information needed to external parties. An excellent overview of PETs in Smart Metering [4] shows that PETs typically focus on either one of two use cases: billing and aggregation (Table I).

PET	Billing	Aggregation
Anonymization	-	X
Cryptographic Computation	-	X
Perturbation	-	X
Verifiable Computation	X	-
Trusted Computation	X	X

TABLE I
Matching of PETs and use cases [4]

The goal of the billing use case is a rather infrequent calculation of the bill for a single customer using the consumption values and the tariff as input. Verifiable computing (VC) methods are cryptographic methods that enable the computation of a function by another, untrusted party. The output consists of the result and a zero knowledge proof (ZKP) that the calculation has been done properly. In the Smart Metering case the bill could be computed at the customer's site as a function of the consumption of the customer and the tariff provided by the energy provider [5]. Since the consumption does not leave the customer's site, privacy is preserved. On the other hand, the energy provider can be sure that the bill has been calculated correctly.

The aggregation use case aims at a frequent calculation of consumption values that are averaged over either space, typically a neighborhood, or time. Average consumption is believed to be sufficient for use cases concerning the network operator (NO) like load monitoring and prediction. Cryptographic computation methods typically employ homomorphic encryption methods, which have the special property that the product of the encrypted load values yields, after decryption, the *sum* of the load values. This property is exploited for the calculation of the average value [6], [7], [8]. Cryptographic methods can be combined with perturbation methods that add a well defined amount of noise to each single measurement in a distributed way such that the sum of these noise values is just sufficient for reaching differential privacy [9], [10], [11]. Differential privacy is a guarantee that the value of a single customer cannot influence the sum too much. As a consequence the sum cannot provide information about single customers.

In the PET literature, other use cases occur as evaluation of practical properties of the presented PET solution. To give an example, it has been studied that some PETs need extensive communication between smart meters and a communication structure that is organized as a tree. The high amount of connections needed in turn affects e.g. the resistance to failure of synchronization or failures of single smart meters [4]. On the other side, there are extensive collections of Smart Metering use cases [12], [13], [14], [15] that contain much more than two use cases. This observation motivates the question: are the use cases billing and aggregation enough or do other Smart Metering use cases need to be considered?

To best of the authors' knowledge no collection of use cases exists that is accurate enough to be used as a basis for the development of PET's. One contribution of this paper is a first step to the composition of a consistent set of Smart Metering use cases which is formulated in a way that is suitable for the development of privacy enhancing technologies for Smart Metering. During the composition of this set of Smart Metering use cases several topics remain open, which is another contribution.

The rest of this paper is organized as follows: Section II shows how the use case collection was derived from existing use case collections. The resulting use case collection is described in section III. Section IV contains a first attempt to estimate the privacy relevance of use cases, and thus also their relevance for PETs, based on the data items that are transferred. Finally, Section V concludes the paper.

II. Method

The approach here is to gather and combine as much information about Smart Metering use cases as possible from already existing documents [12], [13], [14], [15]. In section IV the privacy-sensitive use cases will then identified for further investigation.

A. Combination of Existing Use Cases

The resulting set of use cases should satisfy three main criteria. First, the set of use cases should be as complete as possible. This criterion is aimed to be fulfilled by combination of a number of different use case collections for Smart Metering. The use cases of the European Smart Metering Industry Group (ESMIG, [12]) is the most extensive set of use case collections and is therefore used as a starting point. Since it focuses on the business part ending at the Head End System (HES) as domain boundary, these use cases were enriched or specified in more detail by use cases from other use case collections that focus more on the customer site [13], [14].

B. Simplification of Involved Actors

As a second criterion, the use case description should be suitable for providing a quick overview over the use cases to people working on PETs. From a privacy point of view it is most important whether the user gives her data away or not. Together with the fact that this paper's focus lies in PETs that are applied near the customer's premise, this motivates the decision not to distinguish different parties outside the customer's premise. Thus, only two parties, called "customer" and "service provider/utility", respectively, occur. Note that the description of the billing and aggregation use cases for PETs each also only involve only a single service provider.

For privacy and security purposes a (trusted) third party (TTP) is typically needed. Such a third party could be an actor that distributes keys or acts as the trusted third party for trusted computation (TC) methods. Since the way a TTP comes into play and interacts with customers and the utility is *resulting* from the chosen privacy and security method employed and not directly from the use case, the TTP is omitted for the description of the use cases.

This rather crude simplification has two benefits. On one hand it considerably simplifies the description of the resulting use cases. On the other hand differences in architectures occurring for different use case collections vanish. This in turn enables a focus on a more concrete description of use cases. In fact, strictly speaking both the billing and the aggregation use cases are not use cases but even more specific *functions* that must be computed by the system. As a drawback, regarding all actors outside the household as equally trustworthy, PETs that describe how data are distributed outside the customer's premise cannot be treated.

C. Addition of Data Items

As a third criterion, the use case description should be detailed enough to enable a privacy analysis. This criterion especially implies a description of the collected data items. Since the data that need to be transmitted are not specified in detail and instead modeled by placeholders called, e.g., `MeterData`, data items needed were gathered from additional sources [16] containing so-called baseline data required for the delivery of benefits for network operators.

D. Visualization and Clustering

In order to consider possible dependencies between use cases, a single use case is visualized as a rectangle. If a use case leads to the call of a second use case, an arrow is drawn from the first to the second use case leading to a graph that has to be visualized and brought into an adequate layout: The *inner* use cases occurring inside the household were placed in the middle and separated from the others by a big, dashed rectangle (Figure 1). Then, the *outer* use cases were manually regrouped around the inner use cases such that the intersection of the use case arrows is minimized. It turned out that this worked particularly well, if use cases were clustered using clusters and sub-clusters similar to the ones described in [14]. These clusters were added leftmost and rightmost, respectively (Figure 1).

E. General Changes of Use Cases

In the aforementioned use case collections, the installation process is not described in detail, thus the corresponding use cases are mainly listed here. Although the `Asset-monitoring & error handling` module could have been sorted to the Monitoring sub-cluster, too, it remained at the Maintenance cluster due to its different scope. Here, the focus lies in detection of failures in assets, i.e., either in failures of meters, the communication line or the other endpoint, which is called Meter Data Management (MDM) for simplicity. `Customer move-in/out` and `Supplier Change` are use cases that are likely to call other use cases.

Many of the use cases need to change the configuration of the meter, for this reason a corresponding module `Meter configuration` is introduced and placed within the household. Similarly, many use cases need to inform the user via a `Customer information` module. Since the use cases `(Dis)connect energy supply` and `(Dis)connect devices` act at the household, they are also placed inside the household area. Energy consumption behavior can only change due to the tariff, if an additional module exists inside the household that changes the consumption behavior. The new `Local energy control` module represents either the inhabitant of the household or an automatic Customer Energy Management System (CEMS).

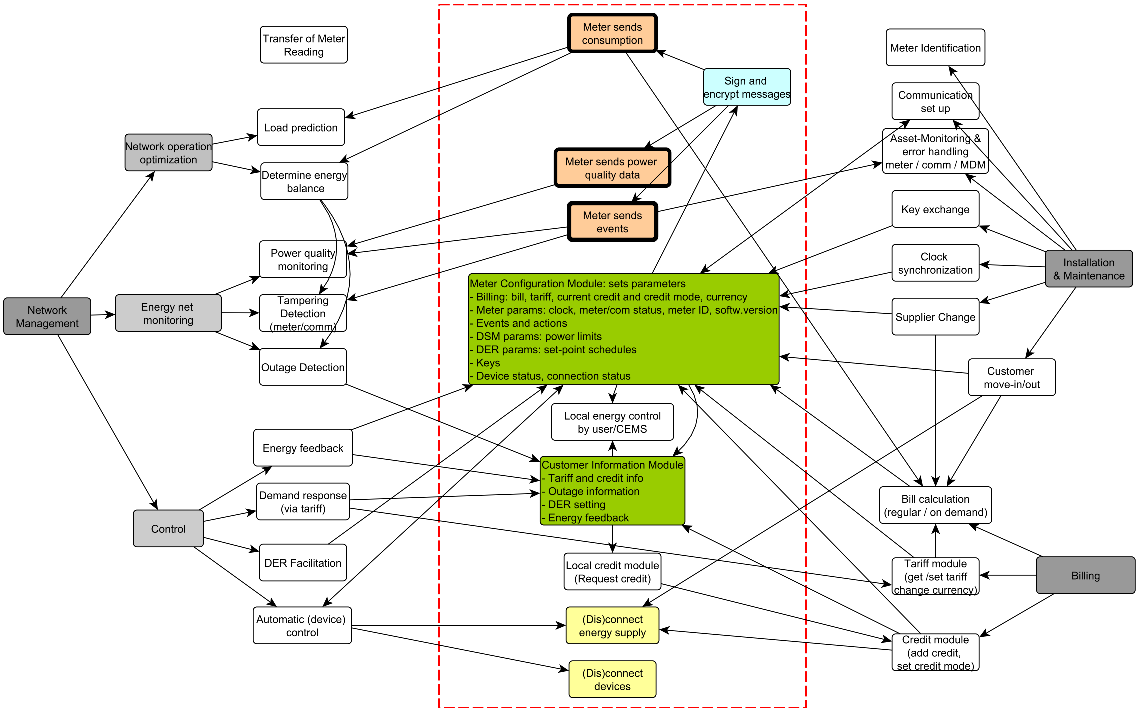

Transfer of Meter Reading
Meter sends consumption
Sign and encrypt messages
Meter Identification
Communication set up
Load prediction
Network operation optimization
Determine energy balance
Meter sends power quality data
Asset-Monitoring & error handling meter / comm / MDM
Meter sends events
Key exchange
Power quality monitoring
Clock synchronization
Network Management
Energy net monitoring
Tampering Detection (meter/comm)
Meter Configuration Module: sets parameters
- Billing: bill, tariff, current credit and credit mode, currency
- Meter params: clock, meter/com status, meter ID, softw.version
- Events and actions
- DSM params: power limits
- DER params: set-point schedules
- Keys
- Device status, connection status
Installation & Maintenance
Supplier Change
Outage Detection
Customer move-in/out
Local energy control by user/CEMS
Energy feedback
Customer Information Module
- Tariff and credit info
- Outage information
- DER setting
- Energy feedback
Bill calculation (regular / on demand)
Demand response (via tariff)
Control
DER Facilitation
Local credit module (Request credit)
Tariff module (get /set tariff change currency)
Billing
Automatic (device) control
(Dis)connect energy supply
Credit module (add credit, set credit mode)
(Dis)connect devices

Fig. 1. Description of use cases. Dashed line: boundary of the household

F. Privacy-Related Adaptions

In the last step it was tried to make the use cases more concrete in order to incorporate privacy-relevant behavior. For this purpose, some additional rectangles or modules were introduced.

ESMIG's original use case `Obtain Meter Reading` was renamed into `Transfer of Meter Reading`. This name better describes the original intention of this use case which is the distribution of the data to the different actors outside, including the specification which actor gets which granularity of the data. It also includes issues such as storage, re-use, deletion and correction of data and linkage to other data sets which is of special importance. Note that this use case has no connections to other use cases which can be explained by the fact that it describes topics that cannot be treated due to the simplifications of actors (Section II-B).

Three additional processes, with thicker outer boundaries emphasizing their importance for privacy, were introduced in order to explicitly describe the flow of data from inside the household to an external actor. The first is called `Meter sends consumption` and describes the highly privacy-relevant process where the meter sends its consumption data outside the household. In order to emphasize the flow of consumption data, the arrows point outwards the household. The meter also sends power quality events, which are generally viewed as not privacy relevant. The third rectangle is describing the sending of all other events.

In order to emphasize the necessity to perform cryptographic computations, an additional rectangle `Sign and encrypt messages` was introduced which offers the possibility to sign and encrypt messages.

The use case `Key exchange` could be seen as part of the use case `Communication setup`, however due to the high relevance for privacy and security it was held outside.

The use cases of the NO are only coarsely described in [12]. The original use case "network operation optimization" is seen here as a third sub-cluster because it is likely that it will contain use cases like `Load prediction` or `Determine energy balance` which are based on consumption data. Although in [12] the use case `Tampering detection` is described to be based on metering events only, it could be possible that also for this use case some form of energy balance based on consumption data is calculated. The same holds for `Outage Detection`. Therefore arrows coming from `Determine energy balance` are pointing to these use cases.

Since both of the two original use cases `Billing` and `Prepayment` need a tariff, the handling of the tariff was introduced as a new use case since it will likely be treated as a single `Tariff module`. As a second benefit, such a module shows more clearly how demand response can be achieved by changing the tariff. The original two use cases were named `Bill Calculation` and `Credit Module`. While most of the smaller billing and prepayment sub-use cases happen outside the household, the original sub-use case `Request credit` is triggered from within the household and was thus modeled by a `Local credit module`. Here, the information about the credit is modeled to be flowing via the use cases `Meter configuration` and `Customer information` to the `Local credit module`.

III. Discussion of the Use Case Description

The resulting description of use cases is shown in Figure 1. The use cases are aligned along the 3 columns in the middle. The big, dashed box denotes the outer boundary of the household, the middle column represents the respective use cases or modules inside the customer's premise. The modules corresponding to use cases running primarily outside the household are the two columns to the left and right of the dashed household boundary.

The figure gives a good impression about the high number of use cases, a PET must be compatible with. Especially the `Installation` and `Maintenance` use cases can turn out to be critical. If timestamps are used for cryptographic protocols, `Clock Synchronization` turns out to be a critical issue. The use case `Asset Monitoring & Error-handling` includes detection of failures (of meters, communication, data receiver) and remedies such as firmware updates, which, as `Key Exchange` is highly important for privacy due to its security relevance.

The figure immediately shows which use cases (the ones with thicker outer boundaries) lead to a flow of data items outside the household. As it is planned by ESMIG's use cases, here the calculation of the bill is performed outside the household via the module `Bill Calculation`. Thus, there is a need to send the consumption data outside the household. However, using verifiable computing methods, the `Bill Calculation` module could be placed inside the household which immediately removes the need to send consumption data outside. Instead of the consumption data, which would in this case even need to be attributable to a household, only the computed price and the corresponding zero knowledge proof would need to be sent outside.

Note that the description of the use cases shows data items that are transmitted outside but it does not indicate a specification in which form e.g. consumption data are sent. In order to study the impact of PETs such as aggregation using homomorphic encryption the data items need to be specified in more detail.

IV. Privacy Relevance of Data Items and Use Cases

In this paper use cases are considered as privacy-relevant, if *privacy-relevant data* of customers would be *transferred outside* the household if no PETs are employed. In this section, it is attempted to estimate the privacy relevance of use cases. Since there is a gap between the information needed and the information available in the use case collections, the result can only be regarded as a first estimation.

Data Item	Privacy relevance
Active energy per household [s]	high
Active energy per household [30min]	medium
Active energy per household [day]	low
Active energy per household [year]	negligible
Reactive energy [s]	medium (?)
Generated energy	low
Consumption data per household [s]	**high**
Consumption per household	high
Tariff	low
Credit	high (?)
Billing data	**high**
Mean voltage	no
Voltage sags and swells	no
Voltage alarms	no
Power quality data	**no**
Voltage events	no
Incoming supply failure detected/restored	no
Maximum demand in 30min > Threshold	low/medium (?)
Average energy > Threshold	low/medium (?)
Reactive average power > Threshold	low (?)
Energy consumption returned below limit	low/medium (?)
Meter events	**low/medium (?)**
Supply disabled/restored	low/high (?)
Device enabled/disabled (external)	low
(Device enabled/disabled (home automation))	(high)
Operating conditions	**low/high (?)**
Contracted power/flow	low
Meter status	no
Meter access log	no (?)
Device information	low
DER parameters	no
Keys	no
Configuration parameters	**low (?)**

TABLE II
Privacy relevance of data items without PETs applied.
(?): Further investigations needed

A. Privacy Relevance of Data Items

Due to its definition the privacy-relevance of use cases is based on the estimation of the privacy-relevance of *data* which is the topic of this subsection. The so-called baseline data of [16] are the basis for the list of data items. Meter configuration data were extended by items that are obviously necessary for use cases like e.g. cryptographic keys.

The classification of data items with respect to their privacy relevance is not a trivial task. For example, personal behavior and current circumstances determine activities in the household which in turn can lead to the use of appliances whose summed consumption is measured by smart meters. Thus, inferring personal behavior from consumption is not simple. However, several studies suggest that active energy data are highly privacy relevant [17], [18], [19]. The information gained from consumption data highly depends on the granularity of the measurements in *time*. This fact is reflected by the presence of several entries for active energy in Table II. The estimation of the privacy relevance for different measurement intervals is based on results of [20].

For reactive energy the situation is not so clear because in typical non-intrusive appliance load monitoring studies reactive energy is used together with active energy but not alone. Since motors have a reactive component [21], reactive energy could be used to identify a subset of appliances such as the garage door opener or the water pump. Since the garage door opener could deliver valuable information about leaving and arrival times, privacy relevance could be considered as medium. Generated energy has low privacy impact because it mainly depends on external factors like the weather.

Data items arising from billing scenarios are clearly personal and have therefore high potential of getting sensitive. Since the current tariff is more influenced by the service provider than by the customer, it is likely to have low privacy impact. However, the current credit could be considered as highly sensitive: the fact that the credit is zero could indicate that the customer is bankrupt. Dependent on the kind of tariff and on the frequency of its calculation, the bill itself could provide indirect information about the consumption.

Power Quality Data consist of mean voltage values, voltage sags and swells or voltage alarms. Since voltage does not depend on the household, all these data items are not privacy-relevant.

The privacy-relevance of meter events is likely to be low. However, this is not really clear for all data items. Data items that are not influenced by the customer such as *incoming supply failure detected* are not privacy-relevant. Many events are created when a physical quantity is compared with a threshold value. A comparison of a quantity like power with many different thresholds can be viewed as a quantization. The number and values of the threshold levels determine the information contained. As an example, a single comparison of the active power with the value of 1 kW can provide information about the usage of ohmic high-power appliances which are typically used for cooking. Through the timestamps of the events, the on/off pattern can be measured which could be used to determine the appliance more precisely.

Another class of data items are operating conditions. The status “disabled” or “restored” could indicate a consumer move-in, a consumer move-out, network control operations or a lack of credit in the prepayment scenario (Figure 1). Together with a credit zero information, the information about supply disablement could get very sensitive. Enablement or disablement of devices during device control is a direct information about appliances and highly privacy-relevant. If a device is automatically controlled from outside, as it is envisioned in Figure 1, privacy relevance could be considered as low since the usage of the appliances is not triggered by concrete actions of persons living in the household. However, this is not the case, if devices are controlled from within the household (home automation). Since home automation via smart metering is unlikely to happen it is not considered here.

Finally, meter configuration data are generally likely to be of low privacy relevance since they typically do not depend on the persons living in the household.

The results of the discussion above are also summarized in Table II where the privacy relevance of the data class is set as the maximum privacy relevance over its data items. Summarizing, consumption data and billing data are highly privacy-relevant. Power quality data and configuration parameters are rather privacy-safe, Meter events are likely to be of low or medium privacy relevance. By themselves, operating conditions have rather low privacy-relevance. However, in combination with credit information they could turn out to be a useful side-information. These results should be seen as a first, preliminary assessment of privacy-relevance needing further investigations.

Use case	PET method
Bill calculation	VC, TC
Load prediction (⋆)	Aggregation (⋆)
Determine energy balance (⋆)	Aggregation (⋆)
Tampering Detection (⋆)	Aggregation (⋆)
Outage Detection (⋆)	Aggregation (⋆)
Prepayment (Credit Modules)	(⋆)
Transfer of Meter Reading	(⋆)
Home Automation	Locality
Electric Vehicles	(⋆)

TABLE III
HIGHLY PRIVACY-RELEVANT USE CASES.
(⋆): USE CASE OR DATA ITEMS NEED TO BE SPECIFIED IN MORE DETAIL

Use case
Communication setup
Key exchange
Sign and encrypt messages
Clock synchronization
Asset Monitoring (failures)

TABLE IV
USE CASES THAT MUST BE COMPATIBLE WITH THE PET

B. Privacy Relevance of Use Cases

In principle, the combination of the results of sections III and IV-A leads to the privacy-relevance of use cases in a straightforward manner (left column of Table III). Note that in Figure 1 no PETs like Verifiable Computing are applied.

However, remaining uncertainties input lead to uncertainties in the classification. Here, a conservative approach is taken assuming supply disablement as highly privacy relevant and assuming that the data used for the calculation of the energy balance is also used for tampering and outage detection.

Therefore, as long as the precise intended treatment is unclear, the use cases for energy net monitoring and optimization remain privacy-relevant with a question mark. The uncertainty about the use case also leads to a question mark for the PET method: if the use cases can be handled using aggregated consumption data, then homomorphic aggregation would be a PET that could be applied. For `Tampering detection`, in the case of fraud alarm, it could still be necessary to check *single* consumption profiles, too.

Since privacy and some PETs such as homomorphic encryption rely on security the corresponding use cases must also be considered in the sense of constraints on the PET method (Table IV). Furthermore, for protocols that employ timestamps [8], proper clock synchronization is a constraint.

There are three use cases that are out of focus of this paper but clearly privacy-relevant. The use case `Transfer`

Data Item	Privacy relevance
Aggregated active energy [s], N≤5	high (?)
Aggregated active energy [s], N≥1000	low (?)
Monthly Bill and ZKP (with PET)	low

TABLE V
PRIVACY RELEVANCE OF DATA ITEMS WITH PETS APPLIED. (?): FURTHER INVESTIGATIONS NEEDED

of Meter Reading Data handles all topics arising after data have been collected. If locally controlled Home Automation data get outside, privacy is likely to be decreased. It is likely that through charging and the credit module the use cases concerning Electric Vehicles will be connected with the Smart Metering use cases considered here introducing information about the location of a person.

C. Tentative Application of PETs

Privacy impact of active energy not only depends on granularity in time but also depends on the *spatial* granularity, i.e. whether the data are available for each household or whether they are aggregated over households. There, the privacy relevance depends on the size of the aggregation set (Table V). In order to compute such aggregates, PETs need to be applied. If verifiable computing is used for the calculation of the bill, only the bill needs to be transferred instead of the consumption per household (Table V).

Privacy enhancing technologies are available for most of the use cases. The use case Home Automation can most easily be handled by locality which means that is likely that all tasks of this use case can be performed locally, based on parameters set from outside. Some of the use cases above have no privacy-preserving method assigned which does not mean that no methods exist. It rather seems plausible that methods from other fields such as the banking or the social network domains can be adapted.

V. CONCLUSION AND OUTLOOK

In the literature about privacy enhancing technologies for Smart Metering the two use cases billing and aggregation are considered. This paper constitutes a first step in answering the question if other use cases need to be considered, too. By combining and reorganizing use cases of existing use case collections a set of privacy relevant use cases was found. This set must be supplemented by another set of primarily security-relevant use cases.

While the results of the paper answered some questions, even more, new topics arised. There is a gap between the accuracy of the description of the use cases and the accuracy needed for the development of PETs. This holds especially for the data items that are transferred during the execution of use cases. It seems most important that this gap is bridged by the provision of more concrete systems and functions instead of use cases. An interesting topic for future research are possible privacy implications occurring for either new data items like e.g. the knowledge that the consumption is above or below a threshold or combinations of data items like e.g. the current credit together with the status of supply. A third topic for future research could be a generalization of the way the use cases were combined to general domains.

ACKNOWLEDGEMENTS

The financial support by the Austrian Federal Ministry of Economy, Family and Youth and the Austrian National Foundation for Research, Technology and Development is gratefully acknowledged.

REFERENCES

[1] M. A. Lisovich and S. B. Wicker, "Privacy concerns in upcoming residential and commercial demand-response systems," in *Clemson Power Systems Conference 2008*, Mar. 2008.

[2] E. L. Quinn, "Privacy and the new energy infrastructure," *Social Science Research Network (SSRN)*, Feb. 2009.

[3] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid — the new and improved power grid: A survey," *IEEE Communications Surveys & Tutorials*, vol. 99, no. 99, pp. 1–37, 2011. to appear.

[4] M. Jawurek, F. Kerschbaum, and G. Danezis, "Privacy technologies for smart grids - a survey of options," tech. rep., Microsoft Research, 2012.

[5] A. Rial and G. Danezis, "Privacy-preserving smart metering," in *Proceedings of the 10th annual ACM workshop on privacy in the electronic society*, WPES '11, (New York, NY, USA), pp. 49–60, ACM, 2011.

[6] Z. Erkin and G. Tsudik, "Private computation of spatial and temporal power consumption with smart meters," in *Proceedings of the 10th international conference on Applied Cryptography and Network Security*, ACNS'12, (Berlin, Heidelberg), pp. 561–577, Springer-Verlag, 2012.

[7] K. Kursawe, G. Danezis, and M. Kohlweiss, "Privacy-friendly aggregation for the smart grid," in *Privacy Enhanced Technology Symposium*, pp. 175–191, 2011.

[8] F. Li and B. Luo, "Preserving data integrity for smart grid data aggregation," in *Smart Grid Communications (SmartGridComm), 2012 IEEE Third International Conference on*, pp. 366–371, 2012.

[9] G. Acs and C. Castelluccia, "I have a dream! (differentially private smart metering)," in *Proc. Information Hiding Conference*, pp. 118–132, 2011.

[10] E. Shi, R. Chow, T. h. Hubert Chan, D. Song, and E. Rieffel, "Privacy-preserving aggregation of time-series data," in *Proc. NDSS Symposium 2011*, 2011.

[11] M. Jawurek and F. Kerschbaum, "Fault-tolerant privacy-preserving statistics," in *Privacy Enhancing Technologies*, Springer, 2012.

[12] European Business Systems Integration and Interoperability (EBSII) working group, "Innovative use cases, architecture and opportunities for the future european smart metering business systems," tech. rep., European Smart Metering Industry Group, Oct. 2012.

[13] Smart Meters Coordination Group, "Functional reference architecture for communications in smart metering systems," Tech. Rep. 50572, CEN/CLC/ETSI/TR, Dec. 2011.

[14] Smart Meters Coordination Group, "Smart metering use cases," tech. rep., CEN/CLC/ETSI/TR, July 2012.

[15] Engage Consulting Limited, "Smart metering system use cases," Tech. Rep. ENA-CR007-002 -1.1, ENA, Apr. 2010.

[16] Engage Consulting Limited, "Privacy impact assessment: Use of smart metering data by network operators," Tech. Rep. ENA-CF002-007-1.0, ENA, Oct. 2011.

[17] M. Lisovich, D. Mulligan, and S. Wicker, "Inferring personal information from demand-response systems," *IEEE Security & Privacy*, vol. 8, no. 1, pp. 11–20, 2010.

[18] 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), pp. 61–66, ACM, 2010.

[19] U. Greveler, B. Justus, and D. Löhr, "Multimedia content identification through smart meter power usage profiles," in *Proceedings of the 2012 International Conference on Information and Knowledge Engineering (IKE'12)*, (Las Vegas, USA), 2012.

[20] G. Eibl and D. Engel, "Influence of data granularity on nonintrusive appliance load monitoring," in *Proceedings of the Second ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec '14)*, (Salzburg, Austria), pp. 147–151, ACM, 2014.

[21] G. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, pp. 1870–1891, Dec. 1992.

3.14 KNIRSCH16A

- F. Knirsch, D. Engel, C. Neureiter, M. Frincu, and V. Prasanna. Privacy assessment of data flow graphs for an advanced recommender system in the smart grid. In O. Camp, E. Weippl, C. Bidan, and E. Aïmeur, editors, *Information Systems Security and Privacy – Revised and Selected Papers of ICISSP 2015*, volume 576 of *Communications in Computer and Information Science*, pages 89–106. Springer International Publishing, 2016. Best Paper Award.

Privacy Assessment of Data Flow Graphs for an Advanced Recommender System in the Smart Grid

Fabian Knirsch[1], Dominik Engel[1], Cristian Neureiter[1], Marc Frincu[2], and Viktor Prasanna[2]

[1] Josef Ressel Center for User-Centric Smart Grid Privacy, Security and Control,
Salzburg University of Applied Sciences,
Urstein Sued 1, A-5412 Puch/Salzburg, Austria
{fabian.knirsch, dominik.engel, christian.neureiter}@en-trust.at
[2] Ming-Hsieh Department of Electrical Engineering,
University of Southern California,
Los Angeles, USA
{frincu, prasanna}@usc.edu

Abstract. The smart grid paves the way to a number of novel applications that benefit a variety of stakeholders including network operators, utilities and customers as well as third party developers such as electric vehicle manufacturers. In order to roll out an integrated and connected grid that combines energy and information flows and that fosters bidirectional communications, data and information needs to exchanged and aggregated. However, collecting, transmitting and combining information from different sources has some severe privacy impacts on customers. Furthermore, customer acceptance and participation is the key to many smart grid applications such as demand response. In this paper we present (i) an approach for the model-based assessment of privacy in the smart grid that draws on a formal use case description (data flow graphs) and allows to asses the privacy impact of such use cases at early design time; and (ii) based on that assessment we introduce a recommender system for smart grid applications that allows users and vendors to make informed decisions on the deployment, use and active participation in smart grid use cases with respect to their individual privacy.

1 Introduction

In a smart grid a number of stakeholders (actors) have to cooperate effectively. Interoperability has to be assured on many layers, ranging from high level business cases to low level network communication. Data and information is sent from one actor to another in order to ensure effective communication. Furthermore, the exchange of vast amounts of data is crucial for many smart grid applications, such as demand response (DR) or electric vehicle charging [Cavoukian et al., 2010], [Langer et al., 2013]. However, this data is also related to individuals and privacy issues are an upcoming concern [McDaniel and McLaughlin, 2009], [Simmhan et al., 2011a]. Especially the combination of data, e.g., meter values and preferences for DR can exploit serious privacy threats such as the prediction of personal habits. In system engineering, privacy is a cross-cutting concern that has to be taken into account throughout the entire development life-cycle, which is also referred to as *privacy by design* [Cavoukian et al., 2010].

Model-driven privacy assessment is especially useful when applied in software engineering. In [Boehm, 2006], the author thoroughly investigates the phases in software engineering and the expected costs for error correction and change requests. Costs double with every phase and once an application or a service is delivered, the additional adding of crosscutting concerns such as privacy is tied to enormous costs. As a result, design time privacy assessment is preferred in early phases of the software engineering process. Therefore, a framework is needed to (i) model the system, including high-level use cases and concrete components and communication flows; and (ii) to assess the system's privacy impact using expert knowledge from the domain. Related work in the domain of automated assessments in the smart grid mainly focuses on security aspects and is not primarily concerned with privacy and the modeling in adherence to reference architectures.

In this paper we address these issues and present an approach for the model-driven assessment of privacy for smart grid applications. The framework proposed in this paper is designed to assist system engineers to evaluate use cases in the smart grid in an early design phase. For evaluation only meta-information is used and no concrete data is needed. We use Data Flow Graphs (DFG) to formally define use cases according to a standardized smart grid reference architecture. The assessment is based on an ontology driven approach taking into account expert knowledge from various domains, including customer views on privacy as well as system engineering concerns. The output is a set of threats and a quantitative analysis of risks, i.e., a number indicating the strength of that threat. To evaluate the system we draw on insights from the University of Southern California microgrid. The primary contributions of this paper are (i) the use of DFGs to model use cases in the smart grid; (ii) the usage of DFGs for a quantitative privacy assessment; and (iii) the use of an ontology driven approach to capture domain knowledge.

The remainder of this paper is structured as follows: In Section 2 related work in the area of smart grid reference architectures, privacy evaluation and automated assessment tools is presented. In Section 3 the architecture of the proposed framework and its components are described. This includes the concept of DFGs for modeling use cases in the smart grid, the principal design of the ontology and the mapping of data flow graphs to the ontology, the methodology for defining threat patterns and finally, how these patterns are matched to use cases. The framework is evaluated with a set of representative use cases in Section 4. Section 5 shows a practical application for the proposed framework as a recommender system for the potential privacy impact when using applications and services in the smart grid. Section 6 summarizes this paper and gives an outlook to further work in this area.

2 Related Work

In this section related work in the field of smart grid reference architectures, privacy evaluation and assessment as well as automated assessment tools are presented. Often, privacy and security are used interchangeably. For the purpose of this paper we refer to privacy as legally accessing data but not using it for the intended purpose. Security, by contrast, would involve the illegal acquisition of data. In both cases, the well established and widely understood terminology from security assessment is used, i.e., *threat*, *attacker*, *vulnerability* and *countermeasure*.

2.1 Reference Models

Stakeholders in the smart grid come from historically different areas, including electrical engineering, computer science and economics. To ensure interoperability and to foster a common understanding, standardization organizations are rolling out reference models and road maps. In the US the NIST Framework and Roadmap for Smart Grid Interoperability Standards [National Institute of Standards and Technology, 2012] and in the EU the Smart Grid Reference Architecture [CEN, Cenelec and ETSI, 2012b] were published. The European Smart Grid Architecture Model (SGAM) is based on the NIST Framework, but extends the model to better meet European requirements, such as distributed energy resources. In this paper we investigate use cases from the US. In particular we are focusing on use cases from the University of Southern California microgrid and we thoroughly discuss a typical DR use case. Investigations have, however, shown that for the purpose of this project all use cases from the US can be directly mapped to the European SGAM without the loss of information. Therefore we propose the utilization of the SGAM for two reasons: (i) the SGAM builds on the NIST model and allows to capture both, use cases from the US and the EU; and (ii) with the SGAM Toolbox [Dänekas et al., 2014] present a framework for modeling use cases based on the SGAM; in that way formally modeled use cases are the input for the evaluation.

2.2 Privacy

Privacy (and security) issues in the smart grid are addressed by standards in the US [National Institute of Standards and Technology, 2010] and the EU [CEN, Cenelec and ETSI, 2012a]. Privacy, in specific, has no clear definition. According to a thorough analysis in [Wicker and Schrader, 2011], privacy can be defined as the right of an individual's control over personal information. More formally this is defined by [Barker et al., 2009] in a four dimensional privacy taxonomy. The dimensions are *purpose*, *visibility*, *granularity* and *retention*. The *purpose* dimension refers to the intended use of data, i.e., what personal information is released for. The purpose ranges from single, a specific use only, to any. *Visibility* refers to who has permitted access. The range is from owner to all/world. *Granularity* describes to what extent information is detailed. The *retention* dimension finally is the period for storage of data. In any case, privacy is assured if all these dimensions are communicated clearly and fully disclosed to data owners and the compliance to the principles is governed. Hence, data is collected and processed for the intended purpose only, and the degree of visibility, granularity and retention is at the necessary minimum.

2.3 Assessment Tools

To measure the degree to which systems adhere to privacy requirements, approaches for automated qualitative assessments (resulting in statements of possible privacy impacts due to privacy critical actions or relationships) and quantitative assessments (resulting in a numeric value that determines the risk of privacy impacts) exist.

In [Ahmed et al., 2007], the authors present an approach towards ontology based risk assessment. The authors propose three ontologies, the *user environment ontology* capturing where users are working, i.e., software and hardware, the *project ontology* capturing concepts of project management, i.e., work packages and tasks and the *attack ontology* capturing possible attacks, e.g., non-authorized data access, virus distribution or spam emails. For a risk assessment, attacks (defined in the attack ontology) are matched with information available from the other ontologies. For a quantitative assessment, the annual loss expectancy is calculated by combining a set of harmful outcomes and the expected impact of such an outcome with the frequency of that outcome. The approach presented by Ahmed et al. is designed for security issues and does not explicitly cover privacy assessments.

In [Kost et al., 2011] and [Kost and Freytag, 2012] an ontology driven approach for privacy evaluation is presented. The aim of these papers is to integrate privacy in the design process. High-level privacy statements are matched to system specifications and implementation details. The proposed *privacy by design* process includes the following phases: identification of high-level privacy requirements, translation of abstract privacy requirements to formal privacy descriptions, realization of the requirements and modeling of the system and analyzing the system by matching formal privacy requirements to the formal system model. Contrary to our work this approach is not focused on use cases in the smart grid and therefore does not model systems based on a standardized reference architecture.

A workflow oriented security assessment is presented in [Chen et al., 2013]. This approach is not based on ontologies but on argument graphs. The presented framework uses *security goal*, *workflow and system description*, *attacker model* and *evidence* as an input. This information is aggregated in a discriminative set of argument graphs, each taking into account additional input. Nodes in the graph are aggregated using boolean expressions and the output is a quantitative assessment of the system. Instead of focusing on workflow analysis using graphs, we model systems as a whole in adherence to the standardized reference architecture using an ontology driven approach to integrate expert knowledge.

A considerably broader approach for an assessment tool that incorporates both, the balancing of privacy requirements and operational capabilities is presented in

[Knirsch et al., 2015]. This work presents a graph based approach that allows the modeling of systems with respect to the operational requirements of certain nodes (e.g. metering at a certain frequency) and the impact of privacy restrictions on subsequent nodes. The authors further present an optimum balancing algorithm, i.e. to what extent restrictions gained from privacy enhancing technologies and the necessary operational requirements can be combined. However, this needs sufficient information on how privacy is impacted by certain use cases which is provided by this work.

3 Architecture

This section is dedicated to an architectural overview as well as a detailed discussion of the components. Figure 1 shows the principal components of the proposed architecture, including input and output. For a privacy assessment, the framework accepts two inputs, a use case UC modeled as a DFG in adherence to the SGAM and a set of threat patterns T. In order to qualitatively analyze this input the use case is mapped to individuals – i.e., instances of classes – of an ontology (sometimes referred to as the *assertion box, ABox* [Shearer et al., 2008]). The corresponding class model (sometimes referred to as the *terminological box, TBox* [Shearer et al., 2008]) is based on the SGAM. This qualitative analysis provides explicit and implicit information about the elements from the DFG: actors, components, information objects and their interrelation. The results of the qualitative assessment are the input for the subsequent quantitative analysis. The output of that analysis is finally a class c from a set of classes C that the use case is assigned to. A threat pattern t is used to describe potential threats, where $t \in T$ and a class c represents a subset of threats T^*. A class c describes how threat patterns and the qualitative results are combined, which is presented as a threat matrix as an output. Note that the terminology *threat matrix* is borrowed from security analysis and that the output is not a matrix in the mathematical sense. A threat matrix compares a set of threats and the risk for these threats. Formally, the classifier is defined as Assign UC to c_i if $t \in T_i^*, \forall t \in T, 1 \leq i \leq \{C\}$. A threat exploits a set of vulnerabilities and is mitigated by a set of countermeasures. Each threat pattern can be evaluated for itself or multiple patterns are combined to classes of threats. A vulnerability is any kind of privacy impact for any kind of stakeholder or actor. Threats are evaluated using the attack vector model which is adapted from security analysis and defined in detail later in this paper. In general, an attack is feasible, if given (i) an attacker; (ii) a privacy asset; and (iii) the resources to perform the attack. Hence, a receiver or collector of privacy critical data items is potentially able to access these assets and to use them in a way not corresponding to the original purpose. This is formally represented as $\langle \text{data access}, \text{privacy asset}, \text{attack resources} \rangle$.

3.1 Data Flow Graphs

In order to qualitatively and quantitatively assess the privacy impact of a use case a formalization is crucial. In this section we introduce the concept of Data

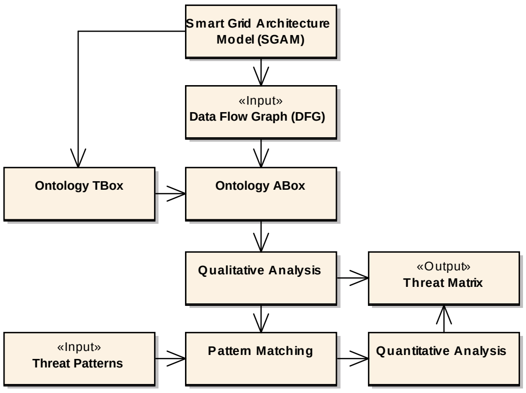


Fig. 1. Architecture overview showing input, output, components and principal information flows of the framework.

Flow Graphs (DFG) for the smart grid based on a model-driven design approach originally presented in [Dänekas et al., 2014] and [Neureiter et al., 2013]. DFGs formally capture all aspects of use cases in the smart grid in adherence to the SGAM. They contain high-level business cases as well as detailed views of a system's characteristics such as encryption and protocols. DFGs are a powerful tool as they allow both, easy modeling and full adherence to the reference architecture. Furthermore, in the graph relationships between actors, as well as the transported information objects (IO) are modeled. Nodes in a graph represent business actors, system actors or components and edges represent data flows annotated with IOs. In accordance to the standard [CEN, Cenelec and ETSI, 2012b], DFGs consist of the following five layers:

1. Business Layer. In a DFG this layer is a high level description of the business case. Business actors, their common business goal and their business requirements are modeled.
2. Function Layer. The function layer details the business case by mapping business actors to system actors and by dividing the high level business goals in use cases and steps.
3. Information Layer. This layer describes information flows in detail. System actors communicate to each other through IOs. IOs are characterized by describing information attributes on a meta-level. An IO is one of the key data used for classification and is discussed in greater detail below.
4. Communication Layer. The communication layer is a more detailed view on communication taking into account network and protocol specifications.
5. Component Layer. In a DFG this layer contains concrete components. Therefore system actors are mapped to components and devices.

Each layer is a directed graph. Both, nodes and edges can have attributes. The semantics, however, are varying. For instance, where attributed edges in the business layer describe a business case, in the information layer concrete metadata of communication flows are captured. Even though implicitly covered in the model presented above, for automated evaluation we introduce two additional layers: Between business and function layer we include the *Business Actor to System Actor Mapping* and between communication and component layer the *System Actor to Component Mapping*. This allows to capture the complexity of use cases on different levels while still maintaining the cross-layer relationship between high-level business actors and their representation as components. These layers are directed graphs as well, with edges indicating the mapping. The mapping defines a one to many relationship from business actors to system actors and from system actors to components. In the European Smart Grid Reference Architecture with the SGAM Methodology an approach for mapping use cases to the reference model is suggested. DFGs build on this methodology focusing on actors and their interrelation. An implementation for modeling DFGs in UML is available as the *SGAM Toolbox*[3]. Data Flow Graphs contain explicit information (what is modeled) and implicit information (what can be concluded). Conclusions are drawn using ontology reasoning.

3.2 Ontology Design

The ontology driven approach for classification has been chosen for two main reasons: (i) ontologies are powerful for capturing domain knowledge explicitly; and (ii) through logic reasoning [Shearer et al., 2008] ontologies are a source for implicit knowledge. The power of ontologies to formally capture knowledge and how to draw conclusions is discussed in [Guarino et al., 2009]. The power of reasoning for gaining additional, implicit knowledge can easily be outlined with two examples: In a DFG, information objects may be sent from an actor A to an actor B and from there to another actor C. This is explicitly modeled in the DFG. A reasoner in an appropriate ontology, however, may conclude directly the transitivity, hence that actor A in fact sends information to actor C. Another example is concerned with compositions of data. An information object I_1 may contain sensitive data and it may be used by an actor D to compose another information object I_2 that is sent to a collecting actor E. It is not explicitly modeled in the DFG, but it can be concluded by the reasoner, that E receives an information object which is of type sensitive data since I_2 is a composition of I_1. The ontology we propose here is designed to capture all aspects of a DFG. The ontology is modeled in OWL[4] and class expressions are stated in Manchester Syntax[5]. Therefore, all components available for modeling DFGs are represented either directly or as an abstraction in the ontology (referred to as the *TBox*). The DFG is represented in the ontology as a set of individuals (referred to as the

[3] http://www.en-trust.at/downloads/sgam-toolbox/

[4] http://www.w3.org/TR/owl-features/

[5] http://www.w3.org/TR/owl2-manchester-syntax/

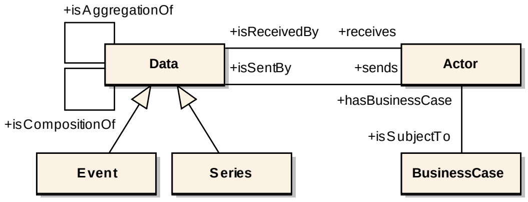


Fig. 2. Principal components of the ontology, showing a subset of the relationships between actor and data.

ABox). Figure 2 depicts the principal classes and relationships of the ontology and therefore the most relevant concepts for mapping a DFG to the ontology. This view shows the main classes and relationships for illustration purposes only; our current ontology comprises more than 60 classes, data properties and object properties. Crucial concepts represented immediately, include which actor sends or receives which data and IO and how these IOs are composed. Furthermore, a set of pre-classifiers is defined to determine implicit knowledge.

These classifiers are OWL classes using an equivalent class expression in Manchester Syntax. For instance, to determine if some aggregation consists of direct personal data, the following expression is used: `Data and isAggregationOf some DirectPersonalData`. To determine the multiplicity of the sending actor and if the data is a composition sent by many of such actors, more elaborate expressions can be phrased: `Data and isSentBy some Actor and Multiplicity value "n" and isCompositionOfMany some Data`.

3.3 Threat Patterns

In this paper we evaluate the privacy impact on customers, thus we identified the following list of typical high-level threats based on literature reviews [Cavoukian et al., 2010], [Langer et al., 2013], [Simmhan et al., 2011a]. These threats have been modified in order to be more representative for the use cases from the University of Southern California microgrid that are investigated in this paper. Subsequently, IOs that may cause these threats are determined.

Customer presence at home. This privacy concern is discussed in [Cavoukian et al., 2010]. To potentially determine a person's presence at home, some device in the customer premises is needed. This device collects data at a certain frequency, high enough to have a resolution that allows to draw conclusions on the energy usage of specific devices. Furthermore, data collected from that device needs to be sent to another actor (i.e., a utility). At the utility an individual or a system needs to have access to the data in an appropriate resolution. Since we always assume that data is accessed legally, we do not focus on unallowed data access. Additionally, the total delay of the data transmission is of relevance. If data is collected and transmitted in almost real time the presence at home can be determined immediately. If data is available with a delay only, the analysis of past events and predictions might be possible. If this information is published, an attacker might exploit this vulnerability in order to break in the house.

Tracking customer position. This threat is especially interesting for electric vehicle charging. Assuming the customer has some identification towards the charging station, at least the location, a timestamp and the amount of energy consumed will be recorded for billing. Depending on the design of the infrastructure only little information will be sent to the operator or a very detailed profile of the customer is maintained. Here, the multiplicity of the actors is crucial and the fact that different actors have access to the same data. Attacks for this threat are described in [Langer et al., 2013], e.g., using information for targeted ads, for tracking movements to certain places or to infer the income based on recharges.

3.4 Pattern Matching

Actual classification is done in the pattern matching process. For each actor in the DFG and the ontology, respectively, the attack vector is determined, i.e., to which resources does an actor have access and what is the effort. If that shows feasible matching this is seen as a threat. It can be retrieved immediately from the ontology if an actor has access to a certain IO. This is done by evaluating actor and data object properties and by incorporating information from the pre-classifiers. Furthermore, relationships on the business layer and data properties such as encryption are taken into account. The following, discriminative set of classifiers is used to determine potential threats: first, for each information object the data provider and the data collector are determined (according to the terminology defined in [Barker et al., 2009]) and it is assessed who has access to the data. This yields a list of three-tuples in the form ⟨information object (IO), data provider (DP), data collector (DC)⟩. Then it is determined if an information object either contains sensitive or direct personal data (according to the terminology defined in [The European Parliament and the Council, 1995]). This yields another three-tuple in the form ⟨information object (IO), sensitive (S), direct personal (DP)⟩. Finally it is determined if the attacker has actual data access, yielding one more three-tuples in the form ⟨information object (IO), data collector (DC), access (A)⟩. Data access depends on the relationship of actors, on data resolution, retention and encryption. Matching these tuples to each other results in the components of the attack vector, recalling ⟨data access, privacy asset, attack resources⟩ yields $\langle\langle \mathrm{IO}, \mathrm{DP}, \mathrm{DC}\rangle, \langle \mathrm{IO}, \mathrm{S}, \mathrm{DP}\rangle, \langle \mathrm{IO}, \mathrm{DC}, \mathrm{A}\rangle\rangle$. An exemplary attack vector for a DR use case where DR preferences are sent to the utility is $\langle\langle \text{DR preferences}, \text{customer}, \text{utility}\rangle$, $\langle \text{DR preferences}, \text{false}, \text{false}\rangle$, $\langle \text{DR preferences}, \text{utility}, \text{true}\rangle\rangle$. This already provides thorough qualitative analysis. It is possible to determine which actor can potentially threaten the privacy of another actor. It is even possible to conclude how and where this might happen. However, for a quantitative assessment the risk for a particular threat is calculated. While a qualitative assessment is useful in supporting detailed system design decisions and evaluation, for a very first outline of the overall system characteristics, a quantitative value is much more expressive. Further, providing a numeric value for the system's privacy impact helps to easily compare and contrast proposed designs.

Risk is calculated as the product of the *probability of occurrence* (PO) and the *expected loss* (EL). For the set T^* a number of patterns $t_{v,1} \dots t_{v,N}$ and $t_{c,1} \dots t_{c,M}$, respectively is defined. A pattern therefore contains a set of conditions for vulnerabilities $t_{v,i}$ and countermeasures $t_{c,i}$. Conditions are SPARQL ASK queries[6] that return either *true* or *false* if the pattern applies or not. For brevity, t'_v denotes the number of vulnerabilities that apply, t'_c the number of countermeasures that apply and t_v and t_c denote the total number of vulnerabilities and countermeasures, respectively. In this paper we propose the following approach for determining values for the probability of occurrence $PO(t'_v, t'_c)$ and the expected loss $EL(t'_v, t'_c)$: $PO(t'_v, t'_c)$ is determined by defining a plane that satisfies the following conditions: $PO(t'_v = t_v, t'_c = 0) = 1$, $PO(t'_v = 0, t'_c = t_c) = 0$ and $PO(t'_v = 0, t'_c = 0) = \frac{1}{2}$. This yields $PO(t'_v, t'_c) = \frac{1}{2}(\frac{t'_v}{t_v} - \frac{t'_c}{t_c} + 1)$. A linear model is chosen due to its simplicity and might be extended by more complex approaches in future. A condition that is of type *vulnerability* increases $EL(t'_v, t'_c)$, a condition of type *countermeasure* decreases $EL(t'_v, t'_c)$. The value of $EL(t'_v, t'_c)$ is defined in the pattern. Risk R is finally defined by $R = PO(t'_v, t'_c)EL(t'_v, t'_c)$.

To feed in the results gained from the qualitative analysis, certain variables in the query can be bound to instances. For example, given the following fraction of a query (where `usc` denotes the namespace prefix for actors and IOs in the University of Southern California microgrid)

```
$io usc:isSentBy ?systemactor . $io usc:isReceivedBy ?systemactor .
?systemactor usc:isRealizationOf ?businessactor .
?businessactor a usc:BusinessActor
```

to determine if *some* information object is sent by *some* business actor. It is now possible to bind the variable `$io` to a concrete value as determined in the qualitative assessment, e.g., `$io` ← `InformationObject.CustomerName`. This allows to assess a particular impact on a particular information object or component/actor based on the previously calculated attack vectors.

We developed generic patterns for *typical* threats, i.e., such as the ones mentioned above. The framework is, however, not limited to this set of patterns and allows the definition of an arbitrary number of additional patterns to meet the individual needs of the application scenario. The output of the framework is a threat matrix contrasting the results from the qualitative analysis and from the quantitative risk assessment. For a UC, a threat matrix contains the attack vector and the assigned risk for the determined class c.

For illustrative purposes, the following listing shows an example pattern for *customer presence at home*. This includes the vulnerability *device in customer*

[6] http://www.w3.org/TR/sparql11-query/

premises (exemplary assigned an EL of 4) and the countermeasure *aggregation of data from multiple customers* (exemplary assigned an EL of -6).

```
<Pattern name="customer presence at home">
  <Vulnerability
    name="device in customer premises">
    <EL>4</EL>
    <Condition>
      ?device x:isRealizationOf $ba .
      $ba a x:BusinessActor .
      ?device x:Zone
      "Customer Premises"^^xsd:string
    </Condition>
  </Vulnerability>
  <Countermeasure
    name="aggregation of data from multiple
      customers">
    <EL>-6</EL>
    <Condition>
      $io x:manyAreAggregatedBy ?io2 .
      ?io2 x:isReceivedBy ?ba1 .
      $io x:isRecevivedBy ?ba2
      FILTER (?ba1 != ?ba2)
    </Condition>
  </Countermeasure>
</Pattern>
```

4 Evaluation

For evaluating the framework new, previously unused use cases are applied. The set of threat patterns and their impact on privacy is based on the aforementioned literature reviews. We are therefore using a representative set of use cases describing typical applications in the smart grid. This includes, but is not limited to, smart metering, electric vehicle charging and DR. In this section a real-life use case from the University of Southern California microgrid, and a real-life use case from the Salzburg Smart Grid Model Region are evaluated as an example. These use cases have been chosen as they are (i) simple enough to verify results based on literature reviews; and (ii) complex enough to have an interesting combination of actors and information flows. Evaluation is performed with a prototypical implementation that uses DFGs and threat patterns as an input and produces a threat matrix as an output.

4.1 Smart Metering

For the Salzburg Smart Grid Model Region use case we investigate a typical smart metering scenario as shown in Figure 3. Smart metering is the basis for many advanced applications in the smart grid and therefore considered as a key

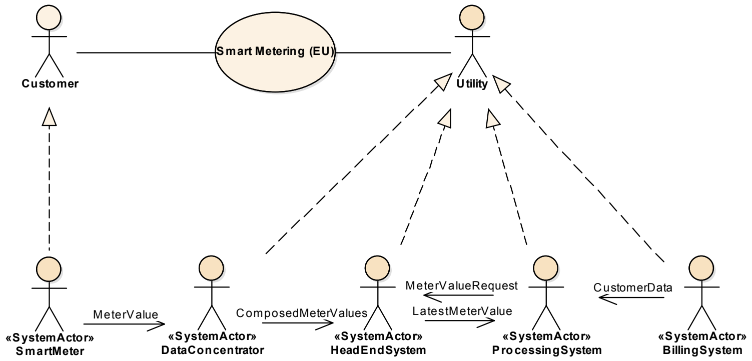


Fig. 3. Outline of the smart netering use case that is discussed for evaluation.

enabling technology [Arnold, 2011]. Today, smart metering is typically applied for network monitoring and billing. The use case is outlined as follows: once a smart meter is installed in a residential building meter values are collected at a fixed frequency. Due to regulatory provisions in this is (e.g., in Austria and Germany) limited to one value each 15 minutes and 96 values per day, respectively. Data for one day is summarized in the smart meter and forwarded to the utility on the previous day. Multiple smart meters are connected in a mash-like topology and data is sent to a data concentrator that (i) relays data from power line communication to IP; and (ii) collect data from the attached meter. Smart meter data is finally stored in a head-end system. For billing, meter data (energy consumption) is linked to additional data from the billing system, such as contract details, name, address and past payment behavior.

Actors. Business actors are the *user* and the *utility*. The user is mapped to the system actor *smart meter*. The utility is represented as *data concentrator*, *head-end system*, *billing system* and *processing system*. The latter is the component linking the data from the billing system to the data from the head-end system.

Information Objects. Meter values are sent at a fixed rate from the customer premises to the utility. The utility stores these values in the head end-system and the processing system finally combines both, data from the head-end system and data from the billing system.

Customer Presence at Home. When metering is done on a regular basis, it is easily detectable if a customer is present at home. The qualitative analysis shows that meter values are sent from the smart meter to the data concentrator and further to the head-end system. Storing in the head-end system is privacy critical, since data metered at a certain frequency is persisted. For this threat four vulnerabilities (device in customer premises, collecting data at a certain frequency, receiver has access to data, data retention is unlimited) and one countermeasure (aggregation of data from multiple customers) are identified, resulting in a *PO* of 0.9, and *EL* of 11.5 and a risk value of 10.35.

Identification of Customer Habits. While the intended use case for persisting meter data is billing, such data can be used to identify customer behavior, e.g., by running statistics and predicting future actions. For this threat eight vulnerabilities (device in customer premises, collecting data at a certain frequency, receiver has access to data, data retention is unlimited, composition of location and timestamp, different actors have access to the same data, location information with unlimited retention) and two countermeasures (aggregation of data from multiple customers, retention is for processing only) are identified, resulting in a *PO* of 0.75, an *EL* of 11.5 and a risk value of 8.63.

The model-driven assessment of the smart metering use case has shown that the risk of identifying customer habits is less than the risk of determining customer presence at home. This is due to the fact that determining presence is a yes/no decision whereas determining and predicting habits requires way more data and information.

4.2 Demand Response

For the University of Southern California microgrid use case, we are focusing on a DR scenario similar to the one described in [Simmhan et al., 2011b]. This scenario is outlined in Figure 4. A customer interested in DR creates an online profile stating on which DR actions the customer is interested to participate (e.g., turning down air condition). When the utilities want to curtail load with DR, a customer whose profile fits the current requirements is sent a text message to, e.g., turn down the air condition. This message is acknowledged by the customer and the utility further reads the meter values to track actual power reduction. Besides the data flows mentioned, this further involves the storing of the profile and the past behavior of the customer for a more accurate prediction. For modeling this use case as a DFG, the following actors and IOs are identified.

Actors. Business actors are the *user* and the *utility*. The user is mapped to the system actors *smart meter*, *device* and *portal*. DR requests are sent to the user device (e.g., a cell phone) and the user's DR preferences are set in the portal (e.g., a web service). The smart meter is used to measure actual curtailment. The utility is mapped to a *DR repository*, containing preferences for each user and past behavior, to a *prediction unit* predicting DR requests based on the preferences and a *control unit* to meter user feedback and actual curtailment.

Information Objects. Cross-domain/zone information flows include user preferences sent to the utilities, DR requests sent to the user from the utility and both, the user acknowledge/decline and the meter values sent back to the utility. Information flows within the utilities' premises are from the DR repository to the prediction unit and from the control unit to the DR repository. Given the threat patterns introduced in Section 3, we use our framework to determine the privacy impact of this use case which provides the following results.

Customer presence at home. The qualitative analysis shows that in the DR repository of the utility information about both, past customer behavior and

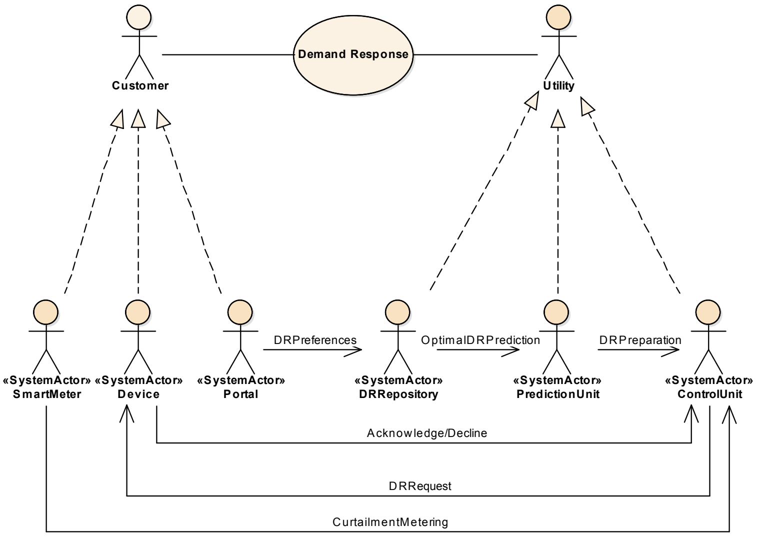


Fig. 4. Outline of the DR use case that is discussed for evaluation.

customer data is brought together, i.e., direct personal data is composed with a detailed history of a person's actions. Furthermore, the customer's acknowledge/decline and the measured curtailment reveal if a customer (i) responded to the DR request; and (ii) actually participated in DR; both is a indication for the presence at home. For this threat we identified four vulnerabilities (device in customer premises, collecting data at a certain frequency, receiver has access to data, data retention is unlimited) and one countermeasure (aggregation of data from multiple customers), resulting in a *PO* of 0.9, an *EL* of 11.5 and a risk value of 10.35.

Tracking customer position. In our case, this threat might apply in two different scenarios: First, this threat is immediate if the acknowledge/decline response to DR requests contains the customer position (e.g., if sent by a cell phone or other mobile device). This does not only show the customers past and present position, but also if the customer is able to remotely control devices in his premises. Second, when the customer is represented by an additional component *electric vehicle charging station*. Assuming that DR requests are also sent with respect to the charging behavior. Based on the amount of energy the customer is willing to DR it might be possible to estimate the consumption of the electric vehicle and subsequently the traveled distance. For this threat we identified two vulnerabilities (composition of location and timestamp, different actors have access to the same data) and one countermeasure (aggregation of data from multiple customers), resulting in a PO of 0.66, an EL of 5 and a risk value of 3.33.

The mode-driven assessment of the DR use case has shown that the risk of tracking customer position is low compared to the risk of determining customer presence at home. This result stems from the fact that there apply a number of vulnerabilities with high expected loss value, hence a device in the customer premises, data collected at a certain frequency, receiver has access to data and unlimited data retention. For the risk of detecting the customer presence at home, the same value applies as for the smart metering itself. This is due to the fact that smart meter data is used as a basis for demand response.

5 Recommender System

Having a framework for assessing the privacy impact of a use case in the smart grid is a powerful foundation for building a recommender system. The objective of such a recommender system is to provide users with the ability to decide on the usage of certain application and services in the smart grid based on the privacy impact of these applications and services. We therefore adapt the policy decision point (PDP) and policy enforcement point (PEP) patterns for a recommender system as originally presented in [Knirsch, 2014]. This is primarily targeting users in order to allow them having full control over information flows, but also the utilities and the vendors of third party applications.

The principal PDP-PEP architecture is standardized as Extensible Access Control Markup Language (XACML) in [Rissanen, 2013]. This architecture has already been applied to the smart grid by Jung et al. in [Jung et al., 2012]. The recommender system we present here enhances this approach by enabling an automated assessment of applications and use cases, respectively.

In general, a PDP is a component that evaluates access requests and issues some authorization. The PDP therefore provides some mechanisms to authenticate users, usually by prompting credentials such as username and password. The PDP then checks in a repository (policy store) if a certain user is granted access to a certain resource. The assessment framework presented in the previous sections is used as a PDP in order to allow privacy-aware data retrieval in the smart grid. The scenario at hand is as follows: a user wants to access a new application or service in the smart grid. This application or service has a certain privacy impact that has been assessed with this framework upon registration (*Registration of Application*). Additionally, the application or service is governed by a PEP. The PEP redirects the user to a PDP that displays to the user a list of privacy implications associated with this particular application or service. The user is then requested to confirm the intention to use the application or service. If the user accepts, the PEP grants access (*Accessing Applications*). For this system, we propose a traffic light-styled display of privacy implications (red: high risk, yellow: medium risk, green: no or low risk) with the option to show the full, detailed analysis.

Registration of Application:

1. A vendor submits a formal application description (DFG) to the recommender system.
2. The recommender system performs the model-based privacy assessment; this yields a set of qualitative metrics (attack vectors) that are stored in the PDP.

Accessing Applications:

1. A user request access to a new application registered at the recommender system.
2. The PEP of this application checks if the user has already allowed access.
3. If *no*, the user is redirected to the PDP and the qualitative assessment is performed based on the user's role (i.e., which business actor corresponds the user to for variably binding and business actor and information object) and the user allows or denies access.
4. If *yes*, the user is forwarded to the application.

In our prototypical implementation as presented in [Knirsch, 2014], Java 1.7 Servlets running on Apache Tomcat 7 represent PDP and PEP, respectively. A user request for an application is guarded by a PEP and forwarded to the PDP, including information about the intended application and the sending party. The PDP performs an ontology driven privacy assessment for the particular use case with a predefined set of threat patterns and displays the result to the user. The result shown includes (i) a summary for the overall privacy impact (traffic light: high, medium, no or low) in appropriate colors for immediate recognizability; and (ii) an optional detailed view showing the full threat matrix. The user is requested to either continue and allow access or cancel. If the user decides to continue, the browser is forwarded to the application. In case the user cancels, one is directed back to the PEP which displays that access will not be granted. For the prototypical implementation, the set of applications is given by the use cases defined above. As the focus is on demonstrating the PDP-PEP pattern for ontology-driven privacy assessment there is no actual implementation of the use cases, i.e., no application that actually performs demand response or the like. In practical use the formal use case description will be provided by either third-parties or the providers of the application themselves.

6 Conclusion and Future Work

In this paper we introduced both, a framework for the model-driven privacy assessment in the smart grid and an advanced recommender system based on that framework. The framework itself builds on an ontology driven approach matching threat patterns to use cases that are modeled in adherence to standardized reference architectures. The approach presented here builds on meta-information and high-level data flows. It has been shown how to utilize this framework to successfully assess the privacy impact on use cases in early design time. Exemplary threats and exemplary use cases draw on insights from the University of Southern California microgrid. Further we proposed a recommender system based on the PDP-PEP pattern. This system utilizes our privacy assessment framework in order to provide users the option to allow or deny access to applications and services based on their privacy impact. Future work will include the rolling out of our recommender system to a real-world setting.

Acknowledgment

The financial support of the Josef Ressel Center by the Austrian Federal Ministry of Economy, Family and Youth and the Austrian National Foundation for Research, Technology and Development is gratefully acknowledged. Funding by the Austrian Marshall Plan Foundation is gratefully acknowledged. The authors would like to thank Norbert Egger for his contribution to the prototypical implementation. Funding by the Federal State of Salzburg is gratefully acknowledged.

This material is based upon work supported by the United States Department of Energy under Award Number number DE-OE0000192, and the Los Angeles Department of Water and Power (LA DWP). The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, the LA DWP, nor any of their employees.

References

[Ahmed et al., 2007] Ahmed, M., Anjomshoaa, A., Nguyen, T., and Tjoa, A. (2007). Towards an ontology-based risk assessment in collaborative environment using the semanticlife. In *Proceedings of the The Second International Conference on Availability, Reliability and Security*, ARES 07, pages 400–407, Washington, DC, USA. IEEE Computer Society.

[Arnold, 2011] Arnold, G. (2011). *Green IT: Technologies and Applications*. Springer Berlin Heidelberg.

[Barker et al., 2009] Barker, K., Askari, M., Banerjee, M., Ghazinour, K., Mackas, B., Majedi, M., Pun, S., and Williams, A. (2009). A data privacy taxonomy. In *Proceedings of the 26th British National Conference on Databases: Dataspace: The Final Frontier*, BNCOD 26, pages 42–54, Berlin, Heidelberg. Springer.

[Boehm, 2006] Boehm, B. (2006). A view of 20th and 21st century software engineering. In *Proceedings of the 28th International Conference on Software Engineering*, ICSE 2006, pages 12–29, New York, NY, USA. ACM.

[Cavoukian et al., 2010] Cavoukian, A., Polonetsky, J., and Wolf, C. (2010). Smartprivacy for the smart grid: embedding privacy into the design of electricity conservation. *Identity in the Information Society*, 3(2):275–294.

[CEN, Cenelec and ETSI, 2012a] CEN, Cenelec and ETSI (2012a). Smart Grid Information Security. Technical report, CEN/Cenelec/ETSI Smart Grid Coordination Group Std.

[CEN, Cenelec and ETSI, 2012b] CEN, Cenelec and ETSI (2012b). Smart Grid Reference Architecture. Technical report, CEN/Cenelec/ETSI Smart Grid Coordination Group Std.

[Chen et al., 2013] Chen, B., Kalbarczyk, Z., Nicol, D., Sanders, W., Tan, R., Temple, W., Tippenhauer, N., Vu, A., and Yau, D. (2013). Go with the flow: Toward workflow-oriented security assessment. In *Proceedings of New Security Paradigm Workshop (NSPW)*, Banff, Canada.

[Dänekas et al., 2014] Dänekas, C., Neureiter, C., Rohjans, S., Uslar, M., and Engel, D. (2014). Towards a model-driven-architecture process for smart grid projects. In Benghozi, P.-J., Krob, D., Lonjon, A., and Panetto, H., editors, *Digital Enterprise Design & Management*, volume 261 of *Advances in Intelligent Systems and Computing*, pages 47–58. Springer International Publishing.

[Guarino et al., 2009] Guarino, N., Oberle, D., and Staab, S. (2009). *What Is an Ontology?* Handbook on Ontologies – International Handbooks on Information Systems. Springer, 2nd edition.

[Jung et al., 2012] Jung, M., Hofer, T., Dbelt, S., Kienesberger, G., Judex, F., and Kastner, W. (2012). Access control for a smart grid SOA. In *Proceedings of the 7th IEEE Conference for Internet Technology and Secured Transactions*, pages 281–287, London, UK. IEEE.

[Knirsch, 2014] Knirsch, F. (2014). Model-driven Privacy Assessment in the Smart Grid. Master's thesis, Salzburg University of Applied Sciences.

[Knirsch et al., 2015] Knirsch, F., Engel, D., Frincu, M., and Prasanna, V. (2015). Model-based assessment for balancing privacy requirements and operational capabilities in the smart grid. In *Proceedings of the 6th Conference on Innovative Smart Grid Technologies (ISGT2015)*. to appear.

[Kost and Freytag, 2012] Kost, M. and Freytag, J.-C. (2012). Privacy analysis using ontologies. In *CODASPY '12 Proceedings of the second ACM conference on Data and Application Security and Privacy*, pages 205–2016, San Antonio, Texas, USA. ACM.

[Kost et al., 2011] Kost, M., Freytag, J.-C., Kargl, F., and Kung, A. (2011). Privacy verification using ontologies. In *Proceedings of the 2011 Sixth International Conference on Availability, Reliability and Security*, ARES '11, pages 627–632, Washington, DC, USA. IEEE Computer Society.

[Langer et al., 2013] Langer, L., Skopik, F., Kienesberger, G., and Li, Q. (2013). Privacy issues of smart e-mobility. In *Industrial Electronics Society, IECON 2013 - 39th Annual Conference of the IEEE*, pages 6682–6687.

[McDaniel and McLaughlin, 2009] McDaniel, P. and McLaughlin, S. (2009). Security and privacy challenges in the smart grid. *Security Privacy, IEEE*, 7(3):75–77.

[National Institute of Standards and Technology, 2010] National Institute of Standards and Technology (2010). Guidelines for smart grid cyber security: Vol. 2, privacy and the smart grid. Technical report, The Smart Grid Interoperability Panel – Cyber Security Working Group.

[National Institute of Standards and Technology, 2012] National Institute of Standards and Technology (2012). NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 2.0. Technical Report NIST Special Publication 1108R2, National Institute of Standards and Technology.

[Neureiter et al., 2013] Neureiter, C., Eibl, G., Veichtlbauer, A., and Engel, D. (2013). Towards a framework for engineering smart-grid-speficic privacy requirements. In *Proc. IEEE IECON 2013, Special Session on Energy Informatics*, Vienna, Austria. IEEE.

[Rissanen, 2013] Rissanen, E. (2013). eXtensible Access Control Markup Language (XACML) Version 3.0. Online. http://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.pdf.

[Shearer et al., 2008] Shearer, R., Motik, B., and Horrocks, I. (2008). Hermit: A highly-efficient owl reasoner. In Dolbear, C., Ruttenberg, A., and Sattler, U., editors, *OWLED*, volume 432 of *CEUR Workshop Proceedings*. CEUR-WS.org.

[Simmhan et al., 2011a] Simmhan, Y., Kumbhare, A., Cao, B., and Prasanna, V. (2011a). An analysis of security and privacy issues in smart grid software architectures on clouds. In *IEEE International Conference on Cloud Computing (CLOUD), 2011*, pages 582–589. IEEE.

[Simmhan et al., 2011b] Simmhan, Y., Zhou, Q., and Prasanna, V. (2011b). Semantic information integration for smart grid applications. In Kim, J. H. and Lee, M. J., editors, *Green IT: Technologies and Applications*, pages 361–380. Springer, Berlin Heidelberg, Germany.

[The European Parliament and the Council, 1995] The European Parliament and the Council (1995). Official Journal L 281, 23/11/1995 P. 0031 - 0050 – Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995. Online.

[Wicker and Schrader, 2011] Wicker, S. and Schrader, D. (2011). Privacy-aware design principles for information networks. *Proceedings of the IEEE*, 99(2):330–350.

3.15 KNIRSCH15A

- F. Knirsch, D. Engel, M. Frincu, and V. Prasanna. Model-based assessment for balancing privacy requirements and operational capabilities in the smart grid. In *Proceedings of the 6th Conference on Innovative Smart Grid Technologies (ISGT)*, pages 1–5, Feb 2015.

Model-based Assessment for Balancing Privacy Requirements and Operational Capabilities in the Smart Grid

Fabian Knirsch*, Dominik Engel*, Marc Frincu† and Viktor Prasanna†
*Josef Ressel Center for User-Centric Smart Grid Privacy, Security and Control
Salzburg University of Applied Sciences
Urstein Sued 1, A–5412 Puch/Salzburg, Austria
Email: {fabian.knirsch, dominik.engel}@en-trust.at
†Ming-Hsieh Department of Electrical Engineering
University of Southern California
Los Angeles, USA
Email: {frincu, prasanna}@usc.edu

***Abstract*—The smart grid changes the way energy is produced and distributed. In addition both, energy and information is exchanged bidirectionally among participating parties. Therefore heterogeneous systems have to cooperate effectively in order to achieve a common high-level use case, such as smart metering for billing or demand response for load curtailment. Furthermore, a substantial amount of personal data is often needed for achieving that goal. Capturing and processing personal data in the smart grid increases customer concerns about privacy and in addition, certain statutory and operational requirements regarding privacy aware data processing and storage have to be met. An increase of privacy constraints, however, often limits the operational capabilities of the system. In this paper, we present an approach that automates the process of finding an optimal balance between privacy requirements and operational requirements in a smart grid use case and application scenario. This is achieved by formally describing use cases in an abstract model and by finding an algorithm that determines the optimum balance by forward mapping privacy and operational impacts. For this optimal balancing algorithm both, a numeric approximation and – if feasible – an analytic assessment are presented and investigated. The system is evaluated by applying the tool to a real-world use case from the University of Southern California (USC) microgrid.**

I. Introduction

In a smart grid a number of systems have to cooperate effectively. For instance, in a demand response (DR) use case, data is captured by a smart meter, stored in a database and finally used by a prediction unit to forecast customer energy usage. Data is captured, exchanged and processed in order to achieve this high-level use case. Other examples of such use cases include smart metering for billing or automated electric vehicle charging. As these use cases rely to a great extent on personal data, security and privacy are current issues and subject to ongoing research [1], [2], [3]. Privacy aware data retrieval and processing is therefore crucial in order to meet statutory and customer requirements. However, when adding to many privacy constraints, the system's ability to perform the intended task may degrade. In this paper, use cases are investigated that need an optimum trade-off between privacy and operational capabilities. There are use cases where both, privacy and operational capabilities can be achieved fully at the same time, this is, however, not subject of this paper.

As a motivating example, imagine a simple demand response use case where future energy consumption of a particular customer at a certain point in the day (e.g., around noon) is predicted based on past behavior. This requires to have smart meter data from that customer in a sufficient resolution (e.g., one meter value each fifteen minutes). On the other hand, when providing data in such a granularity the customer might be subject to privacy threats, such as predicting when the customer is present at home or the intended or inadvertent release of fine grained meter data to the public. One of the challenges in system engineering in the smart grid is thus to find a good trade-off between protecting an individual's privacy and being able to provide useful services. In Section II work is presented that performs privacy and security assessments based on an operational description of the system. There is, however, currently no approach that focuses on the evaluation of entire systems in the smart grid in order to find the optimum balance between privacy requirements and operational capabilities. This paper therefore contributes (i) a model that formally describes use cases in the smart grid; (ii) an algorithm to find the optimum balance between privacy and operational capabilities based on that model; and (iii) an approach to assess the impact of privacy constraints on the system. The algorithm presented in this paper involves the analytic solving of an equation. If this is not feasible, a numeric approximation can be applied. For evaluation a specific real-world DR use case drawing on insights from the USC microgrid is investigated closely.

The remainder of this paper is structured as follows: Section II provides an overview of related work in the domain of data flow analysis for security and privacy assessments. Further, state of the art assessment tools are discussed and it is shown how this work extends these tools with a holistic approach for optimization. Section III presents the abstract model for describing data flows and system dependencies by using graphs, transition functions and merging operators. Section IV discusses the two approaches for the optimal balancing algorithm, hence the analytic assessment and the numeric approximation. In Section V both approaches are evaluated by applying the tool to a real-world use case. Section VI

978-1-4799-1785-3/15/$31.00 ©2015 IEEE

summarizes this work and provides an outlook to future work.

II. Related Work

This section presents related work in the domain of data flow analysis and state of the art assessment tools. A workflow-oriented security assessment tool using graphs is presented in [4]. The framework proposed by the authors is based on the evaluation of argument graphs. The system's input are security goal, workflow description, system description, attacker model and evidence. The assessment itself applies a discriminative set of graphs, containing the workflow goal, the actors involved and the messages exchanged. The result of the assessment process is quantitatively presented as an availability score and a confidentiality score. Both are plugged into the system by the evidence, which is based on (statistical) data about the devices. This tool is comprehensive for security analysis, however does not deal with the impact of security constraints on the operational capabilities. In the domain of the smart grid, McKenna et al. [5] discuss the issue of finding the optimum trade-off for smart metering frequencies between customer privacy and application feasibility. The authors illustrate some of the privacy impacts that are becoming evident with certain frequency intervals and investigate typical use cases, such as DR, and the need of data for the successful operation of these systems. The issue of balancing privacy requirements and operational capabilities is also addressed in other fields apart from the smart grid: Oliveira and Zaiane [6] present algorithms for balancing privacy constraints in data mining applications. Massaguer et al. [7] discuss a middleware for pervasive spaces. Their focus is on finding the trade-off between privacy and utility of such a middleware. While these approaches deal with balancing for data retrieval and processing, they do not propose a mathematical model to formally address the issue of balancing privacy and operational requirements.

III. Data Flow Model

An approach towards the modeling of use cases in the smart grid based on the European Smart Grid Reference Architecture [8] are *Data Flow Graphs* (DFG). Neureiter et al. [3], Dänekas et al. [9] and Knirsch et al. [10] thoroughly discuss the application of such directed graphs to privacy assessments in the smart grid. DFGs provide a detailed view of a system on multiple layers, ranging from high-level business goals to low level interactions of components. DFGs capture actors and information objects and support a wide range of attributes. These graphs provide a holistic view of a use case and are a powerful tool for interdisciplinary communication and detailed assessments. Based on the concept of representing data flows in the smart grid as directed graphs, we propose an abstraction of DFGs to a simplified *Data Flow Model* that only consists of nodes and directed edges and a minimum set of attributes, hence transition functions and a privacy requirements/operational requirement for each node. Reduced complexity makes numeric and analytic calculations feasible to be performed on this model.

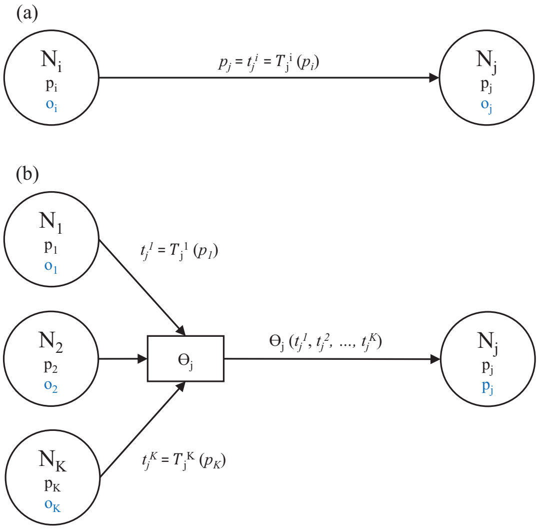


Fig. 1. Abstract data flow graph describing the model with nodes and edges.

Each use case is characterized by a set of actors, i.e., units (smart meter, DR prediction unit, ...), and by a set of information flows from one actor to another, i.e., data items. The model presented here is not limited to physical units, but also allows to be applied to more high-level concepts such as *goals*, e.g., effective DR prediction. In an abstract notation this can be represented as a directed graph with a set of nodes N representing units or goals and a set of transitions T representing flows from a source node to a target node. A node N_k is described by a value p_k that defines the privacy requirement for that node and by a value o_k that defines the operational requirements for that node, so that $\alpha p_k + (1-\alpha) o_k = 1$ and therefore $\alpha \in [0,1]$. Thus requirements are represented as a numeric value in the range 0 to 1. The higher the value the more the requirement weighs. The above condition is introduced for normalization purposes.

An edge is described by a transition function T and a merging operator Θ, described in detail in Section III-A and Section III-B, respectively. The model for describing systems and information flows is in its simplest form as shown in Figure 1 (a). An edge connects a source node N_i with a target node N_j by $(p_j, o_j) = t_j^i = T_j^i\,(p_i, o_i)$; hence a function that maps the privacy and operational requirements of the source to the target. The other, general, case where a target node has more than one incoming edge is shown in Figure1 (b). In addition to the transition function a merge operation needs to be defined and the general form of an edge is given in Equation 1. The merging operator Θ maps a set of one or more input values p_k (the transition vector), each in the domain [0,1], from parent nodes N_k with $k = 1 \dots K$ to one single output value in the same range. Note that this notation can be simplified in practice as the sum of privacy requirements and operational requirements in each node is defined to be 1 and thus only one of the parameters (either p or o) needs to be passed. Therefore, in the following only p is taken into account. For the sake of simplicity, recursive edges are not defined in the data flow model. Recursive edges would represent a system that sends data to itself and for the transition only the identity function would be feasible, since such a system has no practical privacy

$$p_i = \Theta_i \left(t_i^1 = T_i^1(p_1), \dots, t_i^k = T_i^k(p_k)\right) \quad (1)$$

or operational impact on itself.

A. Transition Function

The transition function T is crucial as it immediately defines to what extent the destination system is able to perform its operations. The transition function is a case specific function that needs to satisfy the condition $T : [0,1] \rightarrow [0,1]$, so that the sum of the privacy requirements and operational requirements is one; and hence must not have a singularity in that interval, so that the model is not running in an undefined state. The transition function can be determined by practical observations or models, depending on the particular use case.

As an example for determining the transition function a (sub-)graph with two nodes is given, *smart meter* (N_a) and *DR prediction unit* (N_b). The transition function should represent the fact that the accuracy of DR prediction degrades if (for the particular use case) only data in low resolution is available, e.g., if DR prediction is used to forecast customer energy consumption on a hourly basis, one meter value per day is not sufficient. If we are further assuming that accuracy is following exponential behavior, the following transition function could be used: $p_b = T_b^a(p_a) = \frac{e^{p_a}-1}{e-1}$. The more the privacy is tuned up (thus lower frequency for metering), the less capable (thus less accurate) is the prediction unit.

B. Merging Operator

The merging operator Θ maps the transition vector which described incoming transitions to one single output value in the range 0 to 1. This operator can be determined by practical observations or models, depending on the particular use case or a generic approach can be found that equally incorporates each input value, e.g., by calculating the arithmetic mean.

C. Interpreting Results

Once proper values for p and o for the node of interest are found, these results must be interpreted accordingly to be applied to the system's characteristics in reality. The objective of interpreting results is therefore to map these normalized values to a property that impacts the privacy awareness or the operational capability of the system. This mapping is heavily dependent on individual characteristics and generic approaches provide only limited applicability. In our motivating example we discussed the impact of metering frequency on privacy and operation for subsequent systems. Hence, here we need to find a mapping from p_i with N_i = "Smart Meter" to the meter frequency f_s. In the evaluation we discuss this issue thoroughly and we present such a mapping for the DR use case and in particular for the metering frequency in that scenario.

IV. Optimal Balancing

Once the model is constructed and all nodes and edges including the transition functions and the merging operators are defined, it is possible to calculate the optimal balancing between privacy and operational requirements. The optimal balancing is given by the solution of an equation. If solving

$$\bar{p} = \sum_{i=1}^{N} p_i \quad (2)$$

$$p_1 = \bar{p}N - \sum_{i=2}^{N} p_i \quad (3)$$

this equation is not feasible, a numeric approach for approximating the result can be applied. This section discussed both approaches in detail. The objective of the optimal balancing is to perform the following: (i) automatically find the best trade-off between privacy requirements and operational capabilities for a system that is under development; or (ii) assess to what extent an existing system meets given privacy or operational requirements.

A. Analytic Assessment

The optimal balancing algorithm is performed on the entire system. The analytic assessment thus involves the solving of Equation 3, given an arbitrary $\bar{p}$ in the interval $[0,1]$, e.g., $\frac{1}{2}$ for the optimal balance. Again, the equality condition can be replaced by a greater equal or less equal condition. The equation yields a solution for each p_i for each node. In practice it is sufficient to specify the solution for p_1 or, in case each function T and each operator Θ has a well defined inverse function, to find the solution for an arbitrary p_i and then apply the given functions or the inverse functions in order to calculate the values for the node of interest. By doing so the model can be used to assess the impact of privacy/operational requirements in a particular node for other nodes. For complex systems consisting of many nodes, transitions and merging operations, solving the equation might not be possible or feasible. In the following section we therefore present a numeric algorithm.

B. Numeric Approximation

For approximating the result, a numeric approach can be applied. The algorithm for this approach is given as follows: (1) vary the values for p_1 and o_1, respectively, in the very first node and in the allowed interval, hence from 0 to 1 in a given step size Δ (e.g., 0.01); (2) compute T and Θ for each subsequent transition to get according values for each node; (3) for each variation, summarize and normalize the values for p and o for each node by $\bar{p} = \frac{1}{N}\sum_{i=1}^{N} p_i$ where N is the total number of nodes; and (4) find the variation where $\bar{p} = \frac{1}{2}$. It can be shown that the variation that satisfies the above condition yields the optimal balance between privacy requirements and operational requirements for the system as a whole.

If not a balanced system is intended, but a system that is either privacy aware to a certain extend or able to perform operations to a certain extent, the equality condition in $\bar{p} = \frac{1}{2}$ can be replaced by a more general condition involving a threshold s, such as $\bar{p} \geq s$ or $\bar{p} \leq s$. The remaining variations that satisfy this condition may then be subject to closer investigation.

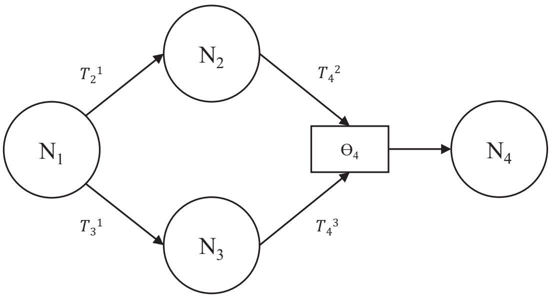


Fig. 2. Data flow model of the demand response use case.

V. Evaluation

For evaluating the system we apply privacy constraints and operational capabilities to a real-world use case that draws on insights from the USC microgrid[1]. First, a data flow graph for this use case is defined and all transition functions and merge operations are set in accordance to practical experiences. Second, we compare the results of the numeric approximation and the analytic assessment in order to find an optimal balance for that use case. The resulting value is finally validated with experiences for that use case gained in practical applications.

A. Use Case Outline

The system that is modeled as a data flow graph is a typical DR use case as described by Simmhan et al. in [11]. The purpose of DR is to curtail load during peak periods by requesting customers to reduce their energy demand for a certain period of time of a certain amount, e.g., by turning off or adjusting the HVAC units. In order to determine which action at which customer is most effective, a prediction model based on current and past energy usage is applied. This model, however, needs meter data of a customers' energy usage at a certain frequency, high enough for accurate predictions. Currently data granularity is one value each fifteen minutes, however, if necessary resolutions up to one value each minute are feasible. In practice the former is used in order to avoid fluctuations in data.

This setting implies two major privacy issues for customers, also addressed by Wicker and Schrader in [12]: (i) if the metering frequency is to high, information about the customer is revealed in (almost) real-time, e.g., if the customer presence at home can be predicted with high accuracy or even which devices are turned on; and (ii) metered data is stored in a database and it is therefore possible to maintain detailed profiles over time. Such information can be released to the public and may immediately affect the customer.

A graph representing this use case is depicted in Figure 2. N_1 represents a smart meter capturing data at a certain frequency, N_2 represents a database storing that data, N_3 represents a DR prediction unit and N_4 represents the goal *effective load curtailment*. The transitions and merging operations are defined as follows:

$$\bar{p} = \frac{1}{4}\left(2p_1 + \frac{e^{p_1}-1}{e-1} + \frac{1}{2}\left(\left(\frac{e^{p_1}-1}{e-1}\right)^3 + p_1\right)\right) \qquad (4)$$

- T_2^1, the metering frequency has no operational impact for data storage in the data base. We are assuming a scalable database which can handle an arbitrary number of streams from meters at any frequency. This transition is therefore the identity function $p_2 = T_2^1(p_1) = p_1$.
- T_3^1, effective DR prediction heavily relies on a metering frequency that is close to real-time. A low frequency therefore reduces the operational capabilities of the prediction unit. This transition is therefore defined as $p_3 = T_3^1(p_1) = \frac{e^{p_1}-1}{e-1}$.
- T_4^2, there are no operational impacts for the overall goal of load curtailment on this path. This transition is therefore again the identity function $t_4^2 = T_4^2(p_2) = p_2$.
- T_4^3; if the operational capabilities of the DR prediction unit are low, the goal of load curtailment can not be achieved sufficiently. This transition therefore reduces the operational capabilities or increases the privacy: $t_4^3 = T_4^3(p_3) = (p_3)^3$.
- $\Theta_{2_{1,2},3}$, for the sake of simplicity the merging operation is defined as the arithmetic mean by $\Theta_4(T_4^2, T_4^3) = \frac{1}{2}(T_4^2 + T_4^3)$.

All functions are bound in the interval $[0, 1]$, hence any value lower than 0 is mapped to 0 and any value greater 1 is mapped to 1.

B. Assessment

For the analytic assessment Equation 3 is applied to the above definitions. This yields Equation 4. Solving this equation for p_1 gives $p_1 \approx 0.59$.

Fig. 3 shows the results for the numeric approximation. Evaluation is performed with a step size $\Delta = 0.01$ for $\bar{p} \geq \frac{1}{2}$ and implemented in Matlab R2010b. The top plot shows the sum of the privacy requirements for each step, the middle plot shows the sum of the operational requirements for each step and the bottom plot shows the overlap of figures, indicating the intersection of the curves where the condition for S is first met. The greater equal condition is preferred over an equality condition in order to deal with numeric inaccuracies (the exact value of $\bar{p}$ might not be reached). Values for p_1 where the condition is met are indicated with a dotted line. The condition is first met at $p_1 \approx 0.59$ and therefore identical to the expected analytic result.

C. Interpretation

Once the assessment is performed, the resulting value, hence $p_1 \approx 0.4$, needs to be mapped to practical meaning. While this is heavily depending on the use case at hand, we propose the following approach for this scenario.

Electricity usage is continuous and digital (smart) metering is sampling that continuous signal at a certain frequency f_s. Following the Nyquist-Shannon sampling theorem [13], f_s needs to be at least twice as high as the highest frequency f_{max} in the signal in order to keep all the information of the original

[1] http://smartgrid.usc.edu/

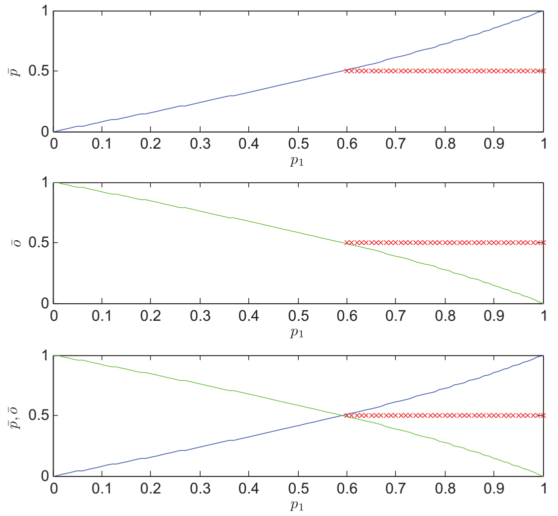


Fig. 3. Plot of the results of the numeric approximation for the use case applied for evaluation.

signal and to restore it losslessly. In practice a metering signal will consist of high frequencies due to peaks in the time domain, e.g., when switching on the light. Full operational capability is therefore given with f_s close to infinity and hence not feasible. If f_s approaches zero, by contrast, privacy is at its maximum. In practice, the upper bound for a meter frequency is given by physical limitations in data capturing and processing by e.g., $f_s = \frac{1}{5}$, hence one value each five seconds.

By describing this with a linear function yielding the privacy impact dependent on the frequency, with $p_1 = -5x + 1$ the intended outcome is achieved. Solving this equation for f_s and by replacing p_1 with 0.59 we get $\frac{0.59-1}{-5} = 0.082$ and thus a meter value approximately every 12.2 seconds. This metering frequency is the one – that based on the model – describes the optimum trade-off between the privacy requirements of the user and the designated goal *effective DR prediction*. Optionally, for a given metering frequency the impact on the goal can be determined, e.g., if f_s is given by $\frac{1}{10}$, this yields $p_1 = 0.5$ and by applying the transition functions and the merging operation $p_4 \approx 0.27$.

VI. Conclusion and Future Work

In this paper an approach has been presented that allows to assess the trade-off between privacy requirements and operational capabilities. Therefore a use case in the smart grid is modeled as a directed graph with nodes and edges. For edges transition functions and merging operations are defined. Based on that graph, an algorithm can be applied for finding the optimum balancing. This can be achieved by either solving an equation or – if this is not feasible – by using a numeric approximation. Finally, we proposed a mapping of the resulting values back to real-world applicability. For evaluation, a demand response use case from the USC microgrid was assessed and discussed.

Future work will focus on integrating this model into existing privacy assessment tools. This allows such systems to provide a more holistic assessment also taking into account the operational capabilities.

Acknowledgment

The financial support of the Josef Ressel Center by the Austrian Federal Ministry of Economy, Family and Youth and the Austrian National Foundation for Research, Technology and Development is gratefully acknowledged. Funding by the Austrian Marshall Plan Foundation is gratefully acknowledged. This material is based upon work supported by the United States Department of Energy under Award Number number DE-OE0000192, and the Los Angeles Department of Water and Power (LA DWP). The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, the LA DWP, nor any of their employees.

References

[1] P. McDaniel and S. McLaughlin, “Security and privacy challenges in the smart grid,” *Security Privacy, IEEE*, vol. 7, no. 3, pp. 75–77, May 2009.

[2] A. Cavoukian, J. Polonetsky, and C. Wolf, “Smartprivacy for the smart grid: embedding privacy into the design of electricity conservation,” *Identity in the Information Society*, vol. 3, no. 2, pp. 275–294, 2010.

[3] C. Neureiter, G. Eibl, A. Veichtlbauer, and D. Engel, “Towards a framework for engineering smart-grid-speficic privacy requirements,” in *Proc. IEEE IECON 2013, Special Session on Energy Informatics*. Vienna, Austria: IEEE, November 2013.

[4] B. Chen, Z. Kalbarczyk, D. Nicol, W. Sanders, R. Tan, W. Temple, N. Tippenhauer, A. Vu, and D. Yau, “Go with the flow: Toward workflow-oriented security assessment,” in *Proceedings of New Security Paradigm Workshop (NSPW)*, Banff, Canada, September 2013.

[5] E. McKenna, I. Richardson, and M. Thomson, “Smart meter data: Balancing consumer privacy concerns with legitimate applications,” *Energy Policy*, vol. 41, pp. 807–814, 2012, modeling Transport (Energy) Demand and Policies.

[6] S. Oliveira and O. Zaiane, “Algorithms for balancing privacy and knowledge discovery in association rule mining,” in *Database Engineering and Applications Symposium, 2003. Proceedings. Seventh International*, July 2003, pp. 54–63.

[7] D. Massaguer, B. Hore, M. Diallo, S. Mehrotra, and N. Venkatasubramanian, “Middleware for pervasive spaces: Balancing privacy and utility,” in *Middleware 2009*, ser. Lecture Notes in Computer Science, J. Bacon and B. Cooper, Eds. Springer Berlin Heidelberg, 2009, vol. 5896, pp. 247–267.

[8] CEN, Cenelec, and ETSI, “Smart Grid Reference Architecture,” CEN/Cenelec/ETSI Smart Grid Coordination Group Std., Tech. Rep., November 2012.

[9] C. Dänekas, C. Neureiter, S. Rohjans, M. Uslar, and D. Engel, “Towards a model-driven-architecture process for smart grid projects,” in *Digital Enterprise Design & Management*, ser. Advances in Intelligent Systems and Computing, P.-J. Benghozi, D. Krob, A. Lonjon, and H. Panetto, Eds. Springer International Publishing, 2014, vol. 261, pp. 47–58.

[10] F. Knirsch, D. Engel, C. Neureiter, M. Frincu, and V. Prasanna, “Model-driven Privacy Assessment in the Smart Grid,” Josef Ressel Center for User-Centric Smart Grid Privacy, Security and Control, Tech. Rep., January 2014.

[11] Y. Simmhan, Q. Zhou, and V. Prasanna, “Semantic information integration for smart grid applications,” in *Green IT: Technologies and Applications*, J. H. Kim and M. J. Lee, Eds. Berlin Heidelberg, Germany: Springer, 2011, pp. 361–380.

[12] S. Wicker and D. Schrader, “Privacy-aware design principles for information networks,” *Proceedings of the IEEE*, vol. 99, no. 2, pp. 330–350, Feb 2011.

[13] H. Nyquist, “Certain topics in telegraph transmission theory,” *Proceedings of the IEEE*, vol. 90, no. 2, pp. 280–305, Feb 2002.

3.16 EIBL14A

- G. Eibl and D. Engel. Influence of data granularity on nonintrusive appliance load monitoring. In *Proceedings of the Second ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec '14)*, pages 147–151, Salzburg, Austria, 2014. ACM.

Influence of Data Granularity on Nonintrusive Appliance Load Monitoring

Günther Eibl
Josef Ressel Center for User-Centric Smart Grid
Privacy, Security and Control
Salzburg University of Applied Sciences
Urstein Sued 1, Puch/Salzburg, Austria
guenther.eibl@en-trust.at

Dominik Engel
Josef Ressel Center for User-Centric Smart Grid
Privacy, Security and Control
Salzburg University of Applied Sciences
Urstein Sued 1, Puch/Salzburg, Austria
dominik.engel@en-trust.at

ABSTRACT

Decreasing time resolution is the simplest possible privacy enhancing technique for energy consumption data. However, its impact on privacy analyses of load signals has never been studied systematically. Non-intrusive appliance load monitoring algorithms (NIALM) have originally been designed for energy disaggregation for subsequent energy feedback. However, the information on appliance use may also be misused for the extraction of personal information. In this work, the effect of decreasing the time resolution in the usual first step, namely edge detection, is studied. It is shown that event values can be estimated rather reliably, but the detection rate of events significantly decreases with increasing measurement time interval.

Categories and Subject Descriptors

I.5 [**Pattern Recognition**]: Design Methodology—*Pattern analysis*

General Terms

Privacy

Keywords

Privacy enhancement; smart metering; data representation; load disaggregation; edge detection

1. INTRODUCTION

There is a lot of public concern and discussions on the privacy impact of smart metering. However, the discussion is led without knowing the extent of personal information that can be read out of smart meter load profiles. Even more so, there is nearly a complete lack of knowledge about how the amount of personal information relates to the measured time interval. For example, in many European countries, it is planned, that people can opt-in for delivering their load data in 15 minute time intervals. To our knowledge, no one has tried to assess the amount of personal information that can be extracted on 15 minute time interval load profiles.

Note that the decrease in time resolution can be viewed as the most straightforward and simplest privacy enhancing technology (PET), cf. [3]. The goal of this work is making a first step towards the study of its actual impact. This work is a first step, because we focus on determining appliances. The main reasoning behind this approach is that activities of persons in the house trigger appliances that sum up to the total load. The activities themselves are already personal information of which some general habits could be deduced. However, such an analysis of general habits is out of scope of this work.

Information on the appliances are usually extracted from the load profiles by means of so-called 'non-intrusive appliance load monitoring analysis" (NIALM). There is a lot of literature on NIALM algorithms ([5, 15, 2, 1, 14, 8, 6, 13]). The goal of these algorithms is the disaggregation of the total load into the individual appliances loads, e.g., for sake of providing energy feedback to the end-user. From the privacy viewpoint, such NIALM analyses can be seen as a first step of attacking methods, which aim at the unauthorized extraction of personal information.

There are only a few papers treating the technical details of privacy implications of smart metering. In [9], load data were recorded with parallel video data which were processed into activity logs. A NIALM analysis was done yielding the input for subsequent behavior-extraction routines. Extracted behaviors include, e.g., presence, sleep cycles or meal times. In [12] the load profile is divided into so-called power segments using a density based clustering technique. These power segments are described by features such as start time, average power and duration. It is illustrated how such power events could be used for answering several privacy questions. In [4], it is shown that under ideal conditions load curves can be used to identify the currently viewed TV-program.

In this work, the impact of reducing the time granularity on the first part of typical low-frequency NIALM algorithms, namely edge detection ([5, 9, 1, 2, 8, 13]) is studied. In Section 2.1, event detection is described as part of low frequency NIALM analyses. In Section 2.2 the investigated edge detection methods are reviewed. After describing the experimental setup in Section 3, the performance of different edge detection methods is compared in Section 4.1. The core Section 4.2 of this work describes the effect of the time reso-

IHMMSEC 2014 Salzburg, Austria
Copyright 2014 ACM 978-1-4503-2647-6/14/06 ...$15.00.
http://dx.doi.org/10.1145/2600918.2600920.

Time	1s	3s	15s	20s	1min
Paper	[12, 11, 2, 1]	[8, 6]	[9]	[14]	[13]

Table 1: Time Granularities of low-frequency NIALM-studies

lution on the detection of events. Finally, Section 5 contains conclusion and outlook.

2. EVENT DETECTION METHODS

Event detection methods are the typical first analysis step in low frequency NIALM algorithms. Decreasing the performance of event detection is a countermeasure against a possible NIALM-privacy-attack and increase privacy. After discussing why event detection is such a useful first step of NIALM analysis, the event detection methods that are investigated in the experimental part are described.

2.1 Event Detection as Part of NIALM

NIALM approaches are divided broadly into two kinds of methods: high frequency methods look at the waveform of appliances or study transients or higher order harmonics. While high-frequency methods usually need a sampling in the range of kHz, low-frequency methods typically analyze load profiles which are sampled using time intervals in the order of seconds (see Table 1).

Since in this work the time granularity is decreased for privacy purposes, the focus is laid on low-frequency instead of high frequency NIALM methods. Supervised low frequency methods usually consist of several blocks: edge detection, cluster analysis and finding pairs of on-and off clusters for the determination of the duration of an appliance. Edges are sharp increases or decreases of the load signal due to turning on or off an appliance. More generally, edges arise due to the change from one state to another state of an appliance when modeled as a finite state machines (FSM). NIALM algorithms commonly use edges instead of the absolute values for two reasons: First, using absolute values in the presence of unknown appliances, these unknown appliances could be described as a combination of other known appliances. Second, there are adverse cases where a small change in the measured power would result in a big change in the configuration of used appliances, which is an implausible result [5]. Since edge detection is a common first step of a NIALM algorithm, if a decrease of time resolution is able to negatively influence edge detection, the subsequent part of the NIALM algorithm is expected to suffer significantly as well. Considering a possible abuse of NIALM algorithms the diminished disaggregation ability is beneficial from the privacy perspective. For sake of completeness it is noted that the use of edges is common but not mandatory, e.g., in [12] shape features are used instead of edges.

2.2 Investigated Event Detection Methods

In this section event detection methods used in the experimental part are reviewed. A main assumption of this work is the modeling of appliances as finite state machines (FSMs) having different power values for different states. In this work an event $e = (t_e, \Delta P_e)$ is a transition between two such states which is represented by its onset time t_e and the difference between the two power levels of the states ΔP_e. Many appliances have only two states and can simply only be turned on or off. Correspondingly, events for which the signal increases ($\Delta P > 0$) are called on-events because they should typically arise from turning on such an onoff-appliance. Analogously, events for which the signal decreases are called off-events.

The most straightforward method detects an edge, if the backward difference $\Delta P_i = P_i - P_{i-1}$ between consecutive points exceeds a threshold. Each detected edge is considered to be an event $e = (t_i, \Delta P_i)$. This method can be classified as one that focuses on the transition between two levels of a signal [10]. If the transition needs several time intervals, this method divides the transition between two levels in several edges having smaller values than the transition which is usually an unwanted behavior.

The drawback of the backward difference method can be accommodated by merging of subsequent occurring edges stemming from backward differences into a single event [1]. The value of the event is the sum of the individual edge values which can be both positive and negative. The time where the event occurs is defined as the onset time, i.e., the time of the first edge contributing to the event.

Another method proposed in [5] is called 'transient passing edge detection." As its names suggests it is a method focusing on the power levels of the two transition states instead of the transition itself. A transition is defined as being not steady. In the first step the method finds the steady subsequences of the signal. This is done using a sliding window approach where a point is considered part of a steady subsequence, if the range of it and the next $n - 1$ does not exceed a given threshold. The whole signal is thus divided into consecutive steady parts st and transitions tr. For the description of the event, all subsequences $(st_i, tr_{i+1}, st_{i+2})$ are considered. The onset-time t_{e_i} for the description of the event is the last time point of the first steady part st_i. The transition value ΔP_i is the difference between the median of the values of the first steady part st_i and the median of the values of the consecutive steady part st_{i+1}. Taking the median value over the whole steady part leads to a greater robustness in the determination of the event value ΔP_i.

3. EXPERIMENTAL SETUP

The experiments were done using a so-called low frequency dataset of the publicly available REDD-dataset [7]. This dataset consists of measurements of the apparent power for 6 different houses. Measurements are available for the main circuits `mains1` and `mains2`, and for subcircuits like for example kitchen outlets and measurements of individual appliances.

Although the decrease of the time granularity seems straightforward (integrating over the period), it is in fact not. There are several possibilities. First, considering a time interval, different statistics could be computed for this interval. The most straightforward statistic is the average load value which should be enough for most practical solutions such as normal billing or time-of-use billing. However, for some reasons, e.g., pricing based on the maximum load or for control reasons, the maximum load needed during the time interval, could be another useful number. Other statistics as for example the standard deviation of the load values, are also possible but will not be considered further. Finally, there is still the possibility of simple sampling, i.e., taking the load value at the specific point in time.

In the subsequent experiments, three variants are considered: (i) taking the average load in a time interval, (ii) taking the maximum load in a time interval and (iii) sampling at time points.

In order to account for noise, for all methods, events e with a value ΔP smaller than a threshold of 20 Watt are discarded. The same threshold was used for the detection of the stable parts of transient edge detection. The minimal required number of steady points n in transient passing was set to 3 which has good detection properties at reasonable stability.

4. RESULTS

In this section, different edge detection methods are compared with respect to their ability to detect events in smart meter load profiles. Then the effect of the decrease of time resolution on the events found is described.

4.1 Event detection

Since the results are based on the events found, the performance of the event detection methods is assessed for the highest available time resolution of 3 seconds first. A value of 20W was used as threshold for the removal of events occurring due to noise. If the threshold is set too low, additional edges can occur which tends to happen for high-power devices. For low-power devices such as lighting, a noise threshold that is in turn too high can lead to a loss of events. Therefore, the tradeoff between noise removal and the detection of events from low-power devices has to be considered.

The form of the load consumption of appliances can be quite complex. As an example, the load consumed for a full run of the dishwasher is shown in Figure 1. Since the dishwasher's load profile has such a rich structure with long and short on-durations at different power levels and power levels that are decreasing, it was chosen for demonstration of effects of different edge detection settings and of the change of time granularity. Simpler devices for heating are usually purely ohmic and show high power values. These are the appliances whose load profiles have the highest similarity to a rectangular profile.

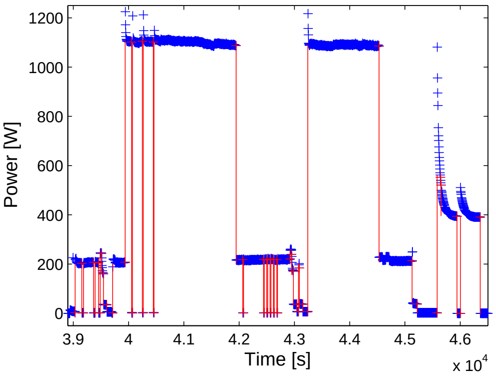


Figure 1: Dishwasher events, marked as "+", detected using transient detection at the highest time resolution

As expected, the simple backward difference yields more, but disturbing, events and can therefore not be recommended (compare Figures 1 and 2).

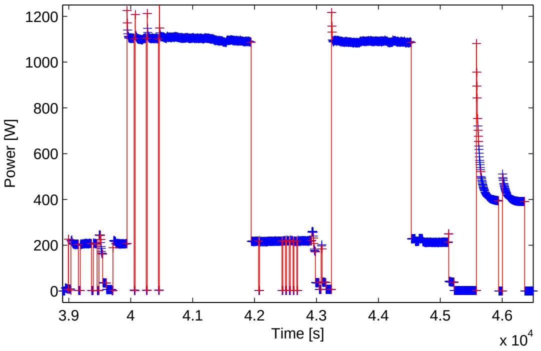


Figure 2: Dishwasher events, marked as "+", detected using backward differences at the highest time resolution. Too many events are detected (compare with Figure 1).

Generally, in terms of detecting appliances, both transient passing and edge merging give good and very similar results. There are also only tiny differences due to the use of the different variants of decreasing the time resolution. The correctness of the edges found was visually verified for all appliances. Additionally, the edge values of all appliances are shown in Figure 3. It can be seen that for all appliances rather distinct edge values can be found. The expected strong similarity of the absolute values of the on-events and the off-events leads to the symmetric look of Figure 3. More importantly, this figure suggests that some appliances such as washerDryer3 should be easily distinguishable from others. Other appliances such as kitchen outlets 2 and 4 are expected to be hardly distinguishable from others. For another class of appliances such as the dishwasher only some levels are distinguishable from the events of other appliances. The fact that the result of edge detection enables to formulate such an expected behavior shows the value of edge detection for a possibly privacy invading analysis of load profiles.

4.2 Effect of Decrease of Time Resolution

In this section, the influence of time granularity Δt on the events found above is studied. First, transient passing using the averaging statistic is studied. As can be seen in Figure 4 with increasing the time interval fewer edges are detected. Especially short-lived states cannot be detected anymore. The edges that are still detected have surprisingly stable heights ΔP.

Another remarkable point is that already with a time interval of 5 minutes, nearly the whole finer structure cannot be seen any more. These results can also be seen for the `mains` signals which was calculated as the sum of the

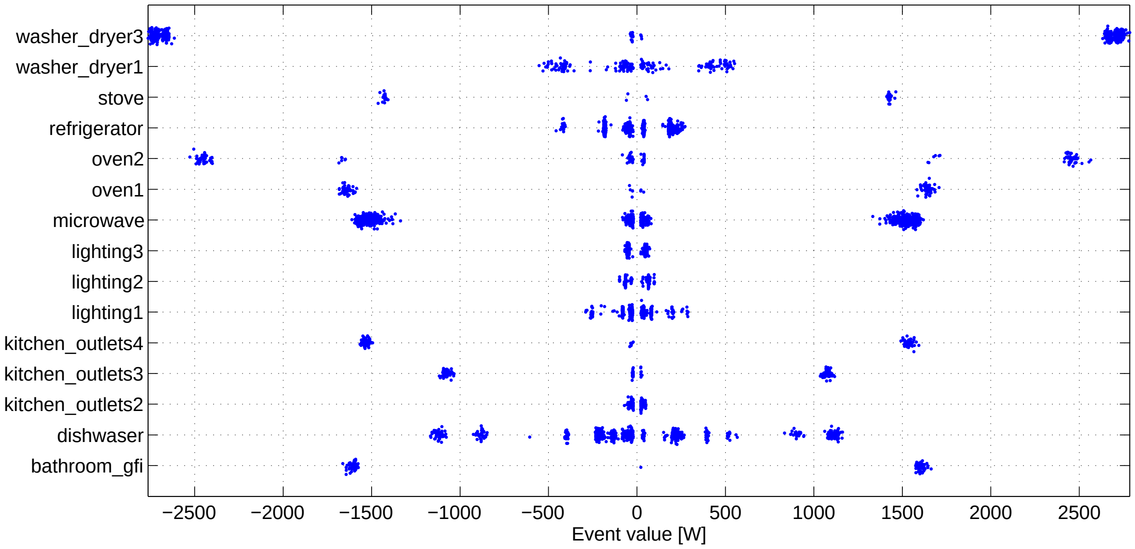


Figure 3: All events found at the highest time resolution, detected using transient detection. The symmetry of the figure stems from the strong similarity of the absolute values of on-events and their corresponding off-events.

`mains1` and the `mains2` signals. Using a 5 minute interval mostly privacy-irrelevant refrigerator events remain.

Possible effects on the decrease in privacy due to the decrease in time resolution can already be estimated. Since the edge heights are rather stable it seems reasonable that the edges of different appliances can still be distinguished at higher time intervals. However, the detection rate of appliances is diminished. In summary, the effect of a decrease in time resolution means that single events cannot be detected reliably. However, for the identification of habits, the detection of each single event is not necessary.

Comparing the different edge detection and time decrease variants, the following behavior could be seen: For high time resolution, edge merging and transient passing lead to nearly identical results, however, for lower time resolutions transient passing seems to better preserve the edge values. The results of both transient passing and edge merging are quite insensitive to the kind of statistic. Although still leading worse results, it should be noted that the performance of the backward difference method is better with taking the max statistic or with sampling than with taking the average statistic where extensive smearing of edge values occurs.

5. CONCLUSION AND OUTLOOK

The impact of decreasing the time resolution on privacy analysis of load signals obtained from smart metering to date has not been studied systematically. Based on the reasoning that knowledge about appliance use can be used as a first step in a privacy attack, the influence of the time interval on edge detection methods has been studied.

Three edge detection methods were investigated: the transient passing method [5], merging of backward differences and simple backward differences. Based on experiments with the REDD-data [7] the simple backward difference cannot be recommended as an edge detection tool in this setting leading to too many edges.

The decrease of the measurement time interval as a privacy enhancing operation has the effect that edge detection still works in the sense that edge heights can be detected in a stable manner. Privacy is enhanced in a way that not every edge is detected. The longer the time interval the fewer edges can be detected. Already with 5 minute intervals, for most of the appliances, the number of detected edges is significantly decreased. A potential privacy consequence would state that not every single event but rather regular habits can be detected.

This work constitutes the first, descriptive assessment of the effect of a decrease of data granularity on smart meter privacy focusing on the detection of appliance use. Next logical steps include the development of quantifiable performance indicators, e.g., based on the result of subsequent pattern recognition algorithms. Using these performance indicators the difference of the effect on different appliances should be described and visualized in a way that is also understandable for non-experts. Furthermore, when appropriate datasets are available, personal information such as activities or habits should be considered in addition to appliance usage.

6. ACKNOWLEDGMENTS

The financial support by the Austrian Federal Ministry of Economy, Family and Youth and the Austrian National Foundation for Research, Technology and Development is gratefully acknowledged. Funding by the Austrian Research Promotion Agency (FFG) under Bridge Project 832082 is gratefully acknowledged.

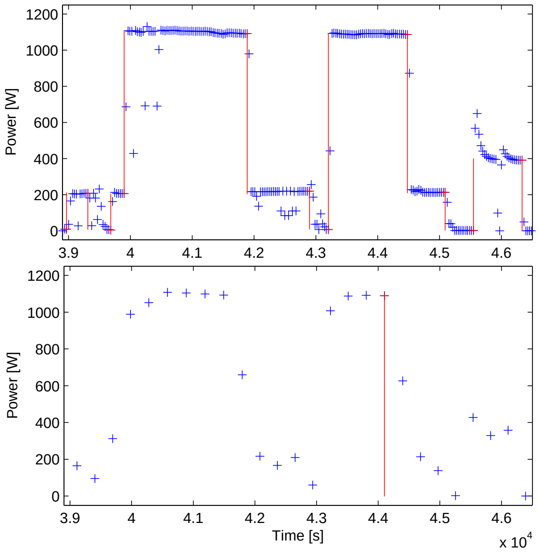


Figure 4: Dishwasher events, marked as "+", detected for Δt =30s (top) and 5 minutes (bottom).

7. REFERENCES

[1] M. Baranski and J. Voss. Genetic algorithm for pattern detection in nialm systems. In *IEEE International Conference on Systems, Man and Cybernetics*, 2004.

[2] D. C. Bergman, D. Jin, J. Juen, N. Tanaka, C. Gunter, and A. Wright. Distributed non-intrusive load monitoring. In *Proceedings of the IEEE/PES Conference on Innovative Smart Grid Technologies (ISGT 2011), Anaheim, CA, USA*, 2011.

[3] D. Engel. Wavelet-based load profile representation for smart meter privacy. In *Proc. IEEE PES Innovative Smart Grid Technologies (ISGT'13)*, pages 1–6, Washington, D.C., USA, Feb. 2013.

[4] U. Greveler, B. Justus, and D. Löhr. Multimedia content identification through smart meter power usage profiles. In *Proceedings of the 2012 International Conference on Information and Knowledge Engineering (IKE'12)*, Las Vegas, USA, 2012.

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

[6] H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han. Unsupervised disaggregation of low frequency power measurements. In *The 11th SIAM International Conference on Data Mining*, pages 747–758, 2011.

[7] J. Kolter and M. Johnson. Redd: A public data set for energy disaggregation research. In *Workshop on Data Mining Applications in Sustainability (SIGKDD)*, pages 1–6, 2011.

[8] J. Z. Kolter and T. Jaakkola. Approximate inference in additive factorial hmms with application to energy disaggregation. *Journal of Machine Learning Research - Proceedings Track*, 22:1472–1482, 2012.

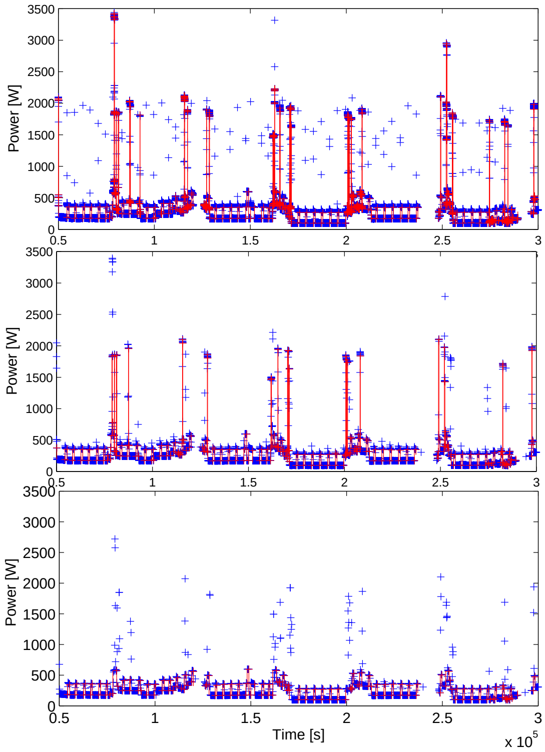


Figure 5: Mains events, marked as "+", detected with Δt=3s (top), 60s (middle) and 5min (bottom).

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

[10] M. A. Little and N. S. Jones. Generalized methods and solvers for noise removal from piecewise constant signals, background theory. In *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, volume 467, pages 3088–3114, 2011.

[11] A. Marchiori, D. Hakkarinen, Q. Han, and L. Earle. Circuit-level load monitoring for household energy management. *IEEE Pervasive Comp.*, 10:40–48, 2011.

[12] 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, pages 61–66, New York, NY, USA, 2010.

[13] O. Parson, S. Ghosh, M. Weal, and A. Rogers. Non-intrusive load monitoring using prior models of general appliance types. In *Twenty-Sixth Conference on Artificial Intelligence (AAAI-12).*, 2012.

[14] E. Vogiatzis, G. Kalogridis, and S. Z. Denic. Real-time and low cost energy disaggregation of coarse meter data. In *4th IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, 2013.

[15] M. Zeifman and K. Roth. Nonintrusive appliance load monitoring: Review and outlook. *IEEE Transactions on Consumer Electronics*, 57:76–84, 2011.

3.17 UNTERWEGER15B

- A. Unterweger, D. Engel, and M. Ringwelski. The effect of data granularity on load data compression. *Springer Lecture Notes in Computer Science – Energy Informatics 2015*, 9424:69–80, 2015.

The Effect of Data Granularity on Load Data Compression

Andreas Unterweger[1], Dominik Engel[1], and Martin Ringwelski[2]

[1] Salzburg University of Applied Sciences, Josef Ressel Center for User-Centric Smart Grid Privacy, Security and Control, Urstein Süd 1, 5412 Puch/Salzburg, Austria. firstname.lastname@en-trust.at
[2] Technische Universität Hamburg-Harburg, Institut für Telematik, Am Schwarzenberg-Campus 1, 21073 Hamburg, Germany

Abstract A vast volume of data is generated through smart metering. Suitable compression mechanisms for this kind of data are highly desirable to better utilize low-bandwidth links and to save costs and energy. To date, the important factor of data resolution has been neglected in the compression of smart meter data. In this paper, we review and evaluate compression methods for smart metering in the context of different resolutions. We show that state-of-the-art compression methods are well suited for high resolution, but not for low resolution data. Furthermore, we elaborate on the compression performance differences between appliance-level and household-level load data. We conclude that the latter are practically incompressible at most resolutions.

1 Introduction

In smart grids, the volume of data to be processed, transmitted and stored is considerable. In the distribution grid, smart meters are a source of high data volume. Depending on the use case and regulatory restrictions, different measurements are collected by a smart meter in different granularities, typically in measurement intervals of 60 seconds up to 15 minutes (cf. Table 10 in [7]). Smart meters are also capable of collecting measurements related to power quality. All measurements can technically be done in smaller intervals (i.e., seconds).

It is evident that compressing the data generated in smart metering is highly desirable. Smart meters are typically connected via low-bandwidth links, such as PLC. Through compression, the bandwidth of these links can be utilized more efficiently. The increase in efficiency, of course, depends on the measurement interval and will increase with smaller intervals. Furthermore, transmitting data in compressed form is more energy-efficient than transmitting data in uncompressed form – given that an appropriately light-weight compression scheme is used, the power needed for compression is significantly lower than the power needed for transmission. Finally, at the receiving end, where the data needs to be stored, compression can help to save costs.

It comes at no surprise that a number of proposals have been made for the compression of smart metering data. However, none of these contribution explicitly addresses the issue of resolution and its impact on compression performance.

While there are many benefits to compression, it has to be evaluated how well the raw data is suited for compression at different measurement intervals, i.e., varying data granularity. An overview of standard compression methods applied to smart metering data is given by Ringwelski et al. [11]. Furthermore, the authors propose their own method. Unterweger and Engel [13] propose a compression method that allows resumability. Two contributions that implicitly address resolution, because they both employ the wavelet transform for compression are Ning et al. [10] and Khan et al. [8]. However, neither takes the impact of resolution of the *source* data into account.

In general, there is little research that addresses the resolution of smart metering data, mostly in the area of smart meter privacy. Eibl and Engel [5] give an account on the influence of data granularity on privacy in smart metering. Approaches for privacy-preserving smart metering are presented by Efthymious and Kalogridis [4] and Engel [6]. Sankar et al. [12] introduce an information-theoretic framework for smart meter privacy, which implicitly addresses data resolution as part of the proposed privacy measure.

In this paper, we evaluate the compression algorithms proposed by Ringwelski et al. [11] and Unterweger and Engel [13] in the context of source data resolution. This is an important perspective, as different use cases in the smart grid will require different measuring intervals and therefore different resolutions of load data. An appraisal on how this resolution impacts compression performance gives an important guideline on what amount of data needs to be transmitted for the individual smart metering use cases.

This paper is structured as follows: In Section 2, we describe the compression algorithms that we evaluate in Section 3. Section 4 concludes.

2 Compression algorithms

Several algorithms for compressing load data have been studied in the literature. We focus on those algorithms which have been specifically designed for load data in the context of smart metering, where resources are typically sparse, i.e., execution time and memory consumption have to be minimized.

For reference, we use two standardized encodings for load data which do not compress the data. For our measurements in Section 3, we use two tailored compression algorithms. All four approaches are described in the following sections. Although some encodings specify the use of units (e.g., watts), we focus on the value encoding only. Unit signaling can be amended if necessary, but is out of scope of this work.

2.1 Reference algorithms

Two standards for transmitting load data are commonly used: IEC 62056-21 [3] and IEC 61334-6 [2], also referred to as A-XDR. Both specify value encodings which do not perform any compression whatsoever, minimizing computational complexity. In the following, we describe both algorithms briefly since we use them for reference measurements.

IEC 62056-21 Values are encoded in their base 10 representation with a decimal point and encoded as ASCII [1] bytes. The value *123.45*, for example, is encoded as `00110001 00110010 00110011 00101110 00110100 00110101`, requiring six bytes – five digits and the decimal point.

Since the length of each encoded value depends on its magnitude, an additional delimiter between subsequent values is required so that they can be separated during decoding. Without additional signaling information, an underscore (ASCII character 137), for example, can be used as a delimiter. This way, the values *123.45* and *123.56*, for example, are concatenated to *123.45_123.56* before encoding, requiring a total of $6 + 1 + 6 = 13$ bytes.

A-XDR Unsigned integer values are encoded in their base 2 representation with a fixed length, e.g., 16 bits. The value *12345*, for example, is encoded as `00110000 00111001`, requiring 2 bytes. Although floating-point values are not supported, multiplying the floating-point value by 10^n, where n is the number of decimal places after the decimal point, yields an integer value which can be encoded using A-XDR.

Since the number of decimal places does typically not change within a load data time series, no additional signaling for n is required. However, the number of bits required for representation may have to be increased to accommodate for the increased value range due to the multiplication by 10^n. For example, encoding the value *123.45* (as *12345*, see above) requires at least 14 bits, as opposed to the value *123*, which only requires 7 bits.

As stated above, A-XDR coding uses a fixed bit length for representing values. Thus, all values can be decoded without the need for any additional delimiters as opposed to the IEC 62056-21 value coding described above.

2.2 DEGA coding

Unterweger and Engel [13] have proposed a compression algorithm for load data which exploits the data characteristics of load profile data. Their encoding algorithm, which we refer to as DEGA (Differential Exponential Golomb and Arithmetic) coding due to its main elements, is illustrated in Fig. 1 and consists of five steps (labeled A-E).

First, the floating-point input values are normalized (A) to make them integer, as explained for A-XDR in Section 2.1. Second, the differences between consecutive values are calculated (B), since they are typically smaller than the values themselves. Third, the differences are encoded as Signed Exponential Golomb code words of order zero (C) for variable-length coding. Fourth, the code words are concatenated (D) and finally compressed using an adaptive binary arithmetic coder (E).

During processing, the code word concatenation step (D) is usually implicitly contained in the code word generation step (C). A detailed explanation of each step as well as a description of the decoding process can be found in [13].

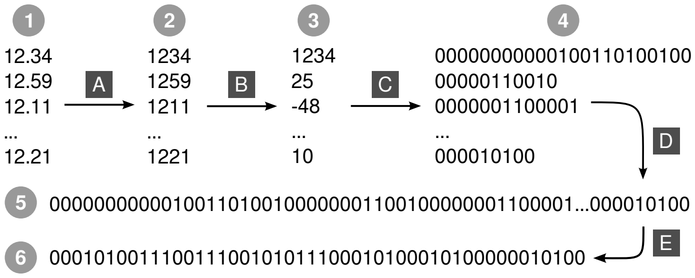


Figure 1. Overview of DEGA coding [13]: Input values (1) are normalized and their differences (3) are represented as Signed Exponential-Golomb code words (4) which are concatenated (5) and arithmetically coded.

2.3 LZMH coding

Ringwelski et al. [11] have proposed a compression algorithm for load data with low memory requirements. The algorithm is referred to as Lempel Ziv Markov Chain Huffman (LZMH) coding and combines ideas of the Lempel Ziv Markov Chain Algorithm (LZMA) and a variant of Adaptive Trimmed Huffman (AHT) coding as described below and illustrated in Fig. 2. It is designed to process ASCII-coded IEC 62056-21 data as described in Section 2.1 as input.

If at least three of the following characters are found in the history of the last m characters, a reference to it is coded (LZMA-like), consisting of a byte offset and the length, using an optimal prefix code. Conversely, when no sufficient reference is found, it is encoded as a Huffman code word (AHT-like). This code word originates from an adaptive Huffman tree which represents the symbol probabilities that are updated for each encoded character.

To keep memory requirements low, a history buffer of $m = 128$ characters is used and the size of the Huffman tree is limited to the size of the input alphabet which may be reduced to the ten decimal digits and the decimal point. A more detailed description of the algorithm can be found in [11].

3 Evaluation

We analyze the compression performance and execution times of A-XDR, DEGA and LZMH coding for IEC 62056-21 input data. The used data sets are described in detail in Section 3.1. As opposed to prior work, we study the effect of different data granularity on the results.

We evaluate different data granularity levels by summing up c consecutive input data values with inter-value temporal distance t, for example, 5 minute (300 seconds) granularity for $t = 3$ (seconds) with $c = 100$. We use the same

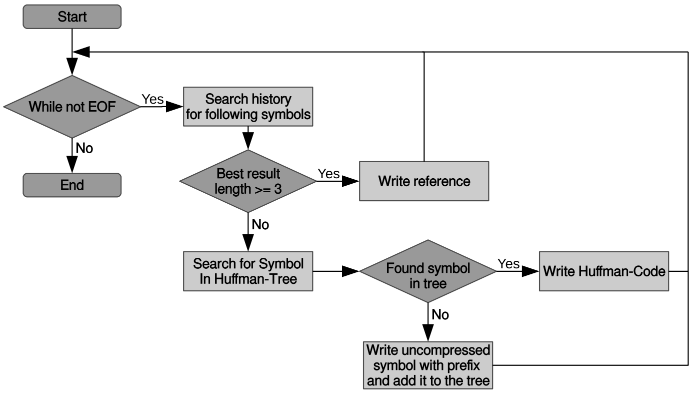

Start
While not EOF
Yes
No
End
Search history for following symbols
Best result length >= 3
Yes
Write reference
No
Search for Symbol In Huffman-Tree
Found symbol in tree
Yes
Write Huffman-Code
No
Write uncompressed symbol with prefix and add it to the tree

Figure 2. Overview of LZMH coding: Each input symbol is either encoded as a reference to an already processed symbol or as a Huffman code word based on its probability.

granularity levels as Eibl and Engel [5], i.e., 3 s, 9 s, 30 s, 1 min., 5 min., 15 min. and 1 h, if available.

To achieve comparable results, we have reimplemented the A-XDR and DEGA coding algorithms in the C programming language. LZMH is already implemented in C and has only been modified slightly so that it uses the same input/output functions. These changes do not affect its compression performance.

3.1 Load data sets

We use two load data sets for our evaluation: the low-frequency MIT REDD data set [9] and a data set from a local energy provider, referred to as the SAG data set henceforth. Both data sets are described briefly below.

REDD The low-frequency MIT REDD data set is a collection of load data from between 11 and 26 channels of 6 different houses. In total, there are 116 channels. Each channel containing load data is available separately.

The load data values are average apparent power readings in Watts with two decimal places, i.e., they are effectively stored with an accuracy of one hundredth of a Watt. They have an inter-value temporal distance of $t = 3$ (seconds) for all channels but the mains, which have $t = 1$. The values cover measurement intervals of between 2.7 and 25.8 days.

SAG The SAG data set is a collection of load data from 508 households and industrial plants. As opposed to the REDD data set, only the mains of each

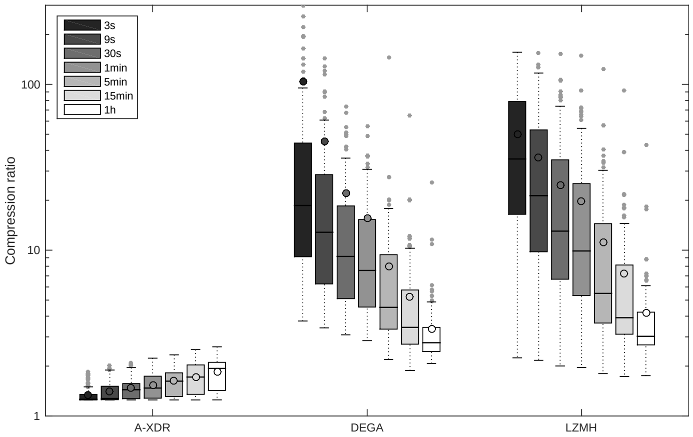


Figure 3. Compression performance of different algorithms compared to IEC 62056-21 value encoding for the REDD load data set at different data granularity levels.

household are available. They are summed up in one single value, i.e., they are not available as separate channels.

The load data values are accumulated energy readings in kWh with three decimal places, i.e., they are effectively stored with Wh accuracy. They have an inter-value temporal distance of $t = 300$, i.e., 15 minutes. The values cover measurement intervals of exactly one year.

3.2 Compression performance

Due to the different characteristics of the two load data sets described in Section 3.1, we analyze the compression performance for each data set separately. All input data is encoded in the form of IEC 62056-21 values as described in Section 2.1, which we use as reference. The results are described in the following sections.

REDD data set Figure 3 shows an overview of the compression performance of the A-XDR, DEGA and LZMH algorithms for the REDD data set. Each channel is compressed separately and its compressed size is expressed relative to the input data size as a ratio. A compression ratio of 5, for example, means that the compressed data requires only 20% of the size of IEC 62056-21 value encoding.

The compression ratio distribution for all channels is depicted as a box plot with added mean compression ratios (filled circles with black borders) and outliers (gray circles without borders). The y axis is logarithmic and capped at 300. Thus, four outliers representing all-zero valued channels are not depicted.

Obviously, DEGA and LZMH exhibit significantly better compression performance than A-XDR, which does not compress by design. Still, it achieves compression ratios greater than 1 compared to IEC 62056-21 value encoding. This is due to the fact that all input values are at least five bytes long (one decimal digit before the decimal point, two thereafter and one delimiter), but typically longer, whereas A-XDR values are always four bytes in size.

In general, LZMH outperforms DEGA at all granularity levels, where the performance difference increases with data granularity. At the finest granularity level (3 s, dark gray boxes), DEGA and LZMH achieve compression ratios of 18.59 and 35.48, respectively. They drop to 2.77 and 3.02, respectively, at the coarsest granularity level (1 h, white boxes).

Compared to A-XDR coding with a median compression ratio of 1.94 at this granularity level, it becomes clear that both, DEGA and LZMH, are practically ineffective at compressing load data with high (1 h) inter-value temporal distances. A-XDR is expected to outperform both compression algorithms at even coarser granularity levels, e.g., at inter-value temporal distances of 24 h.

In general, increased inter-value temporal distances yield larger input values, i.e., they have more decimal digits and therefore yield longer IEC 62056-21 values. Since A-XDR values are of constant size, their compression ratio increases relatively at coarser granularity levels, whereas DEGA and LZMH coding become less efficient in terms of compression performance. This is mainly due to the increased input entropy.

Coarser data granularity impacts compression performance due to the summing of values. Thus, the mains (channels 1 and 2) of all houses from the REDD data set deserve special attention. They, too, are effectively sums of multiple other channels and therefore likely to behave differently than the other channels. Figure 4 shows the compression performance of only the mains.

As expected, the compression performance of DEGA and LZMH coding for the mains is significantly poorer than the respective performance for all channels depicted in Figure 3. Although the best median compression ratio for fine-grain data (3 s, dark gray in Figure 4) is 4.90, double-digit compression performance is not achievable for the mains.

Interestingly, when compressing only the mains, DEGA outperforms LZMH at all granularity levels. The reverse is true when looking at the compression performance of all channels in Figure 3. Still, at medium granularity levels (1 min., medium gray in Figure 4), compression becomes ineffective when compared to uncompressed A-XDR coding.

Even more surprisingly, at coarser granularity levels (15 min., light gray), A-XDR actually outperforms DEGA coding with a median compression ratio of 2.39 vs. 2.25 despite the fact that A-XDR does not compress by design. This means that the mains are effectively incompressible at this resolution.

SAG data set Figure 5 shows the compression results of the A-XDR, DEGA and LZMH algorithms for the SAG load data set. Since the latter only has 15-

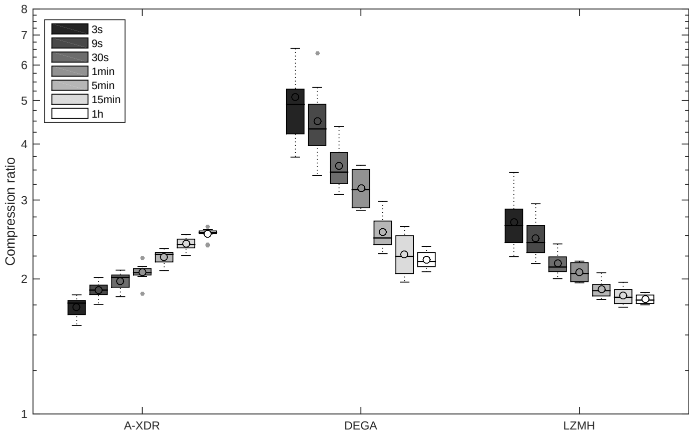


Figure 4. Compression performance of different algorithms compared to IEC 62056-21 value encoding for only the mains from the REDD load data set.

minute resolution, finer granularity levels cannot be evaluated. The visualization is identical to the one in Figure 3 for the REDD data set.

Since the SAG data set only contains measurements from the mains and not from individual channels, the results are similar to the results of the mains from the REDD data set illustrated in Figure 4. Again, DEGA outperforms LZMH coding, but the compression ratios of both are higher when compared to A-XDR, i.e., the data can be compressed to some extent, even at a temporal inter-value distance of 1h.

The main reason for this, considering that the mains from the REDD data set are incompressible as explained above, is the different accuracy of the data. While the REDD mains data has an accuracy of one hundredth of a Watt, the SAG data has an accuracy of only one Watt-hour. This significantly reduces the entropy since the highly volatile least significant digits are missing.

Apart from the lower accuracy, the value range is also reduced, i.e., the kWh readings (SAG) are significantly smaller than Watt readings (REDD). This also explains the high number of outliers (gray circles without borders) in Figure 5 for DEGA and LZMH coding: Households with a lower power consumption yield smaller values which can be more compressed more easily.

3.3 Execution time

The DEGA and LZMH compression algorithms reduce the data rate in a number of cases as described above. However, they are computationally more complex

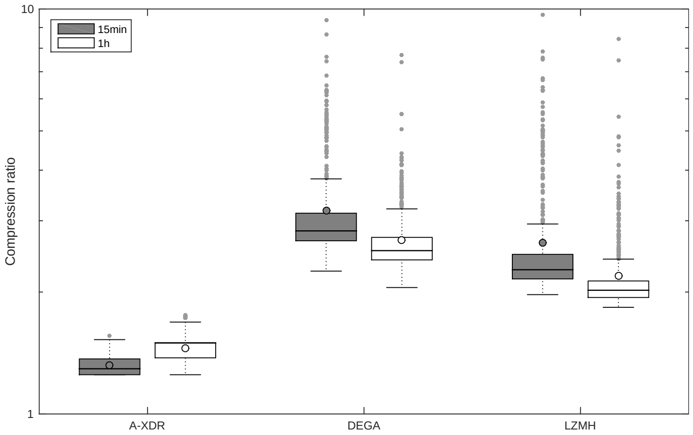


Figure 5. Compression performance of different algorithms compared to IEC 62056-21 value encoding for the SAG load data set at different data granularity levels.

than uncompressed data transmission. Thus, the additional code execution time has to be analyzed.

We measure the execution time similar to Unterweger and Engel [13]. Each channel (REDD data set) or household (SAG data set) is processed, as a whole, three times for cache warming and then five times for the actual time measurements. The five time results are averaged and divided by the number of data values in the processed channel/household to yield the average processing time per data value.

Again, the REDD and SAG data sets are evaluated separately due to their different data characteristics. A-XDR encoding is used as reference for uncompressed processing. All results have been obtained on a virtualized 64-bit *Ubuntu* 14.04 machine with *gcc* 4.8.2 running on an Intel Xeon W3503 CPU.

REDD data set Figure 6 shows an overview of the execution time per data value required by the A-XDR, DEGA and LZMH algorithms for the REDD data set. Despite the powerful CPU used for benchmarking, the processing time is in the microsecond range, i.e., most likely in the 10- or 100-microsecond range on less powerful hardware, e.g., smart meters. This can be considered feasible.

LZMH coding is clearly faster than DEGA coding. Surprisingly, it is, in the majority of cases, even faster than uncompressed A-XDR coding. This can be explained by the fact that both algorithms process data on a per-character basis, but A-XDR requires a conversion to floating point values which involves

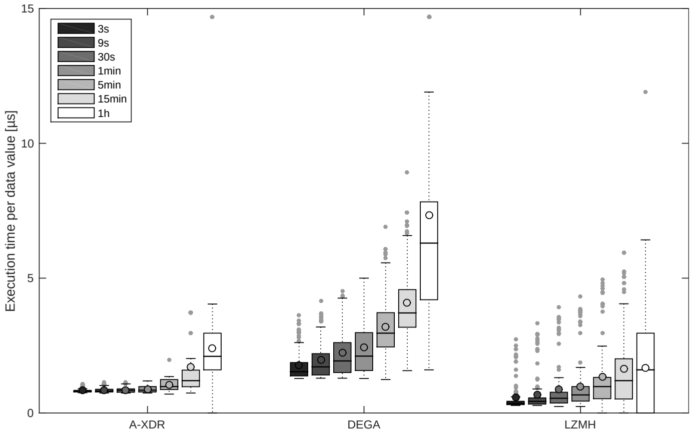


Figure 6. Execution times per value of different algorithms when compressing channels of the REDD load data set at different data granularity levels.

expensive floating point operations. These require about as much time as the whole compression step of LZMH coding, which is very compact.

Execution times increase at finer data granularity levels for all algorithms due to the relative increase in size of the input data. Since the (summed) values are larger in terms of magnitude, their IEC 62056-21 representations are longer. This explains the increased slopes of the median execution times for A-XDR and DEGA coding at coarser granularity levels. As LZMH coding does not convert the representation of the values, its slope is not affected by their magnitude, but by their redundancy, resulting in a smaller slope.

SAG data set Figure 7 shows an overview of the execution time per data value required by the A-XDR, DEGA and LZMH algorithms for the SAG data set. The visualization is identical to the one in Figure 6 for the REDD data set, with the exception of the data granularity range due to the 15-minute inter-value temporal distance of the original data.

The order of magnitude of the execution times is the same as for the REDD data set. However, the absolute values are lower for all algorithms due to the smaller (and therefore shorter) input values. Interestingly, also the differences between 15 min. and 1 h granularity are significantly smaller for the SAG data set than for the REDD data set. Again, this is due to the range (and therefore the length) of the input values.

The slope between the 15 min. and 1 h granularity levels for the SAG data set in Fig. 7 is comparable to the slope for 3 s to 5 min. granularity levels for the

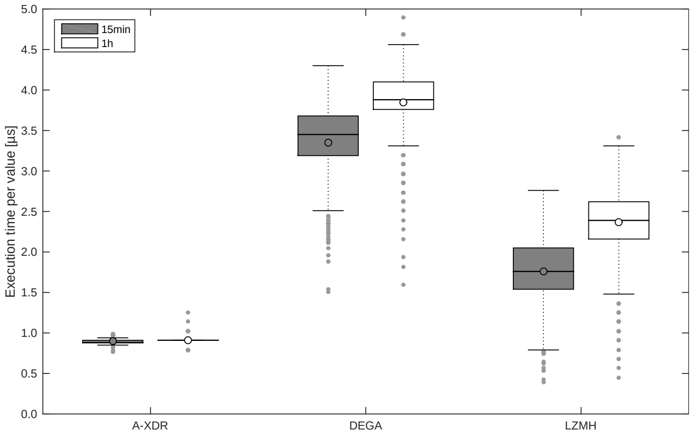


Figure 7. Execution times per value of different algorithms when compressing households of the SAG load data set at different data granularity levels.

REDD data set in Fig. 6. This shows that both, execution times and execution time differences are highly dependent on the input data length.

In addition, the differences in terms of execution time between DEGA and LZMH coding are smaller. This is due the lower compression efficiency of LZMH. This also explains why, in contrast to the execution times for the REDD data set (see Fig. 6), LZMH is slower than A-XDR coding for the SAG data set.

4 Conclusion

Load data from the evaluated data sets is compressible, but only at fine data granularity levels, e.g., 3 second intervals. At coarser granularity levels, compression becomes less effective or even futile, i.e., the reduction in data rate is practically insignificant compared to uncompressed encoding. This effect is stronger for the mains of a household than for per-room or per-device channels which have lower entropy and are therefore easier to compress. When compressing the tested load data sets, LZMH coding by Ringwelski et al. is recommended for the latter type of channels at fine data granularity levels. For coarser granularity levels as well as the mains, DEGA coding by Unterweger and Engel offers higher compression ratios at the cost of longer execution time.

5 Acknowledgements

The authors would like to thank Günther Eibl for his help in visualizing the compression ratio results. They would also like to thank their partner Salzburg AG for providing additional real-world load data.

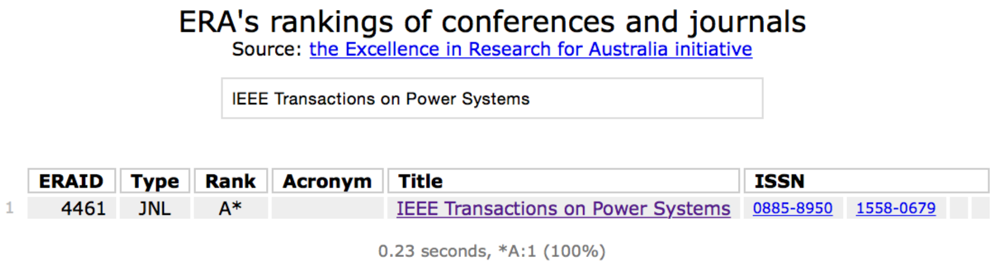
The financial support by the Austrian Federal Ministry of Economy, Family and Youth and the Austrian National Foundation for Research, Technology and Development is gratefully acknowledged.

References

1. Coded Character Sets – 7-Bit American National Standard Code for Information Interchange (7-Bit ASCII) (1986)
2. Distribution Automation Using Distribution Line Carrier Systems – Part 6: A-XDR Encoding Rule (2000)
3. Electricity Metering – Data Exchange for Meter Reading, Tariff and Load Control – Part 21: Direct Local Data Exchange (2002)
4. Efthymiou, C., Kalogridis, G.: Smart Grid Privacy via Anonymization of Smart Metering Data. In: Proceedings of First IEEE International Conference on Smart Grid Communications. pp. 238–243. Gaithersburg, Maryland, USA (2010)
5. Eibl, G., Engel, D.: Influence of Data Granularity on Smart Meter Privacy. IEEE Transactions on Smart Grid 6(2), 930–939 (2015)
6. Engel, D.: Wavelet-based Load Profile Representation for Smart Meter Privacy. In: Proc. IEEE PES Innovative Smart Grid Technologies (ISGT'13). pp. 1–6. Washington, D.C., USA (2013), http://dx.doi.org/10.1109/ISGT.2013.6497835
7. European Commission: Cost-benefit analyses & state of play of smart metering deployment in the EU-27. Tech. rep., European Commission Report (2014), http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52014SC0189&from=EN
8. Khan, J., Bhuiyan, S., Murphy, G., Arline, M.: Embedded zerotree wavelet based data compression for smart grid. In: Industry Applications Society Annual Meeting, 2013 IEEE. pp. 1–8 (2013)
9. Kolter, J., Johnson, M.J.: Redd: A Public Data Set for Energy Disaggregation Research. In: Workshop on Data Mining Applications in Sustainability (SIGKDD). pp. 1–6 (Aug 2011)
10. Ning, J., Wang, J., Gao, W., Liu, C.: A Wavelet-Based Data Compression Technique for Smart Grid. Smart Grid, IEEE Transactions on 2(1), 212–218 (2011)
11. Ringwelski, M., Renner, C., Reinhardt, A., Weigel, A., Turau, V.: The Hitchhiker's guide to choosing the compression algorithm for your smart meter data. In: 2012 IEEE International Energy Conference and Exhibition (ENERGYCON). pp. 935–940 (Sep 2012)
12. Sankar, L., Raj Rajagopalan, S., Mohajer, S., Vincent Poor, H.: Smart meter privacy: A theoretical framework. IEEE Transactions on Smart Grid 4(2), 837–846 (2013)
13. Unterweger, A., Engel, D.: Resumable Load Data Compression in Smart Grids. IEEE Transactions on Smart Grid 6(2), 919–929 (2015), http://dx.doi.org/10.1109/TSG.2014.2364686

APPENDIX A

A.1 MATERIAL FOR ASSIGNMENT OF PUBLICATION CATEGORY



ERA's rankings of conferences and journals

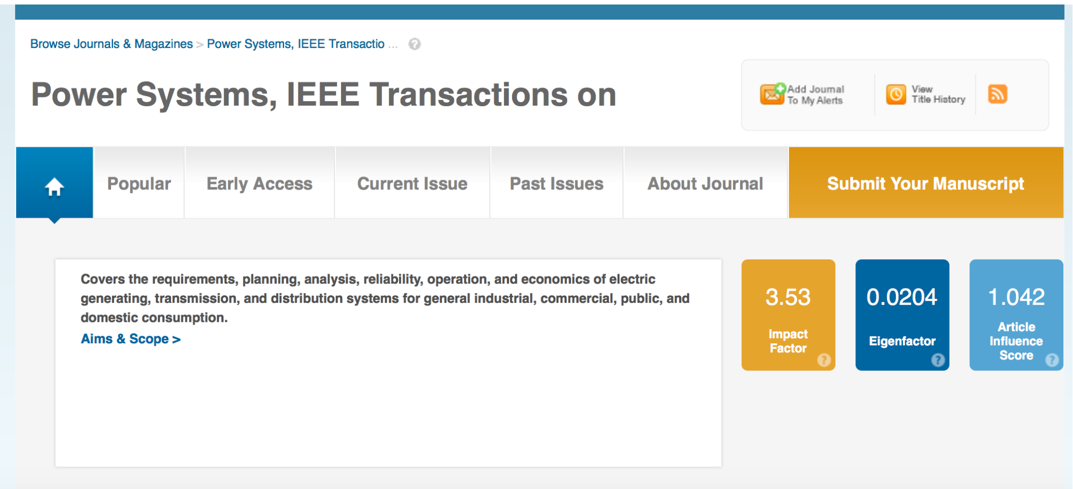
Source: the Excellence in Research for Australia initiative

IEEE Transactions on Power Systems

	ERAID	Type	Rank	Acronym	Title	ISSN
1	4461	JNL	A*		IEEE Transactions on Power Systems	0885-8950 1558-0679

0.23 seconds, *A:1 (100%)

(a) ERA-Ranking *IEEE Transactions on Power Systems* (retrieved February 2015)



Browse Journals & Magazines > Power Systems, IEEE Transactio ...

Power Systems, IEEE Transactions on

Covers the requirements, planning, analysis, reliability, operation, and economics of electric generating, transmission, and distribution systems for general industrial, commercial, public, and domestic consumption.

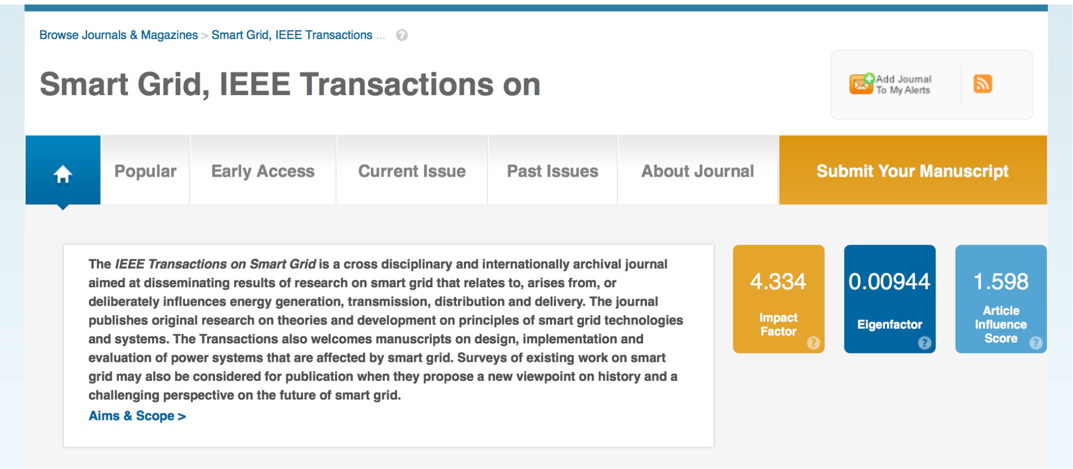
Aims & Scope >

3.53 Impact Factor

0.0204 Eigenfactor

1.042 Article Influence Score

(b) Impact Factor *IEEE Transactions on Power Systems* (retrieved February 2015)



Browse Journals & Magazines > Smart Grid, IEEE Transactions ...

Smart Grid, IEEE Transactions on

The *IEEE Transactions on Smart Grid* is a cross disciplinary and internationally archival journal aimed at disseminating results of research on smart grid that relates to, arises from, or deliberately influences energy generation, transmission, distribution and delivery. The journal publishes original research on theories and development on principles of smart grid technologies and systems. The Transactions also welcomes manuscripts on design, implementation and evaluation of power systems that are affected by smart grid. Surveys of existing work on smart grid may also be considered for publication when they propose a new viewpoint on history and a challenging perspective on the future of smart grid.

Aims & Scope >

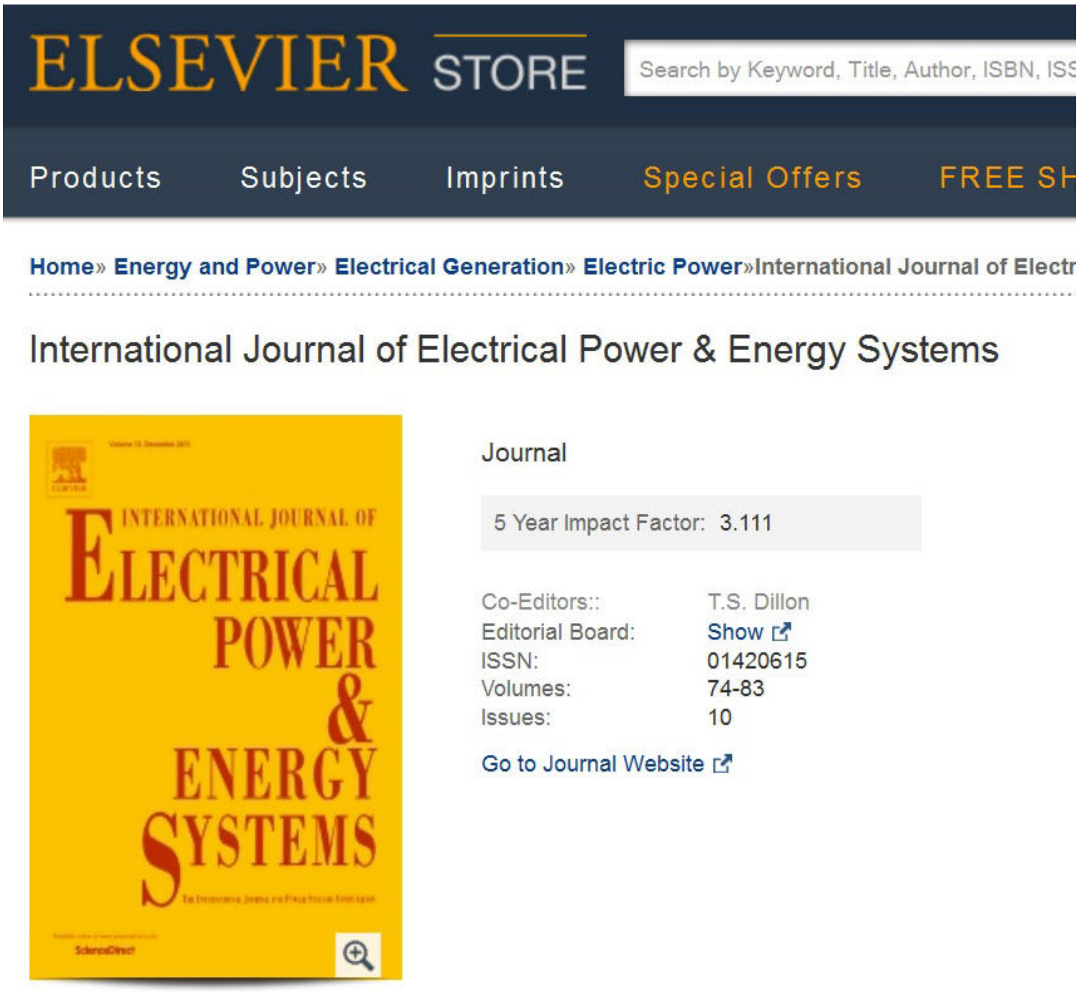
4.334 Impact Factor

0.00944 Eigenfactor

1.598 Article Influence Score

(c) Impact Factor *IEEE Transactions on Power Systems* (retrieved February 2015)

Figure A.1.: Derivation of category A* for *IEEE Transactions on Smart Grids* by Impact Factor


ELSEVIER STORE
Search by Keyword, Title, Author, ISBN, ISS
Products
Subjects
Imprints
Special Offers
FREE SH

Home» Energy and Power» Electrical Generation» Electric Power»International Journal of Electr

International Journal of Electrical Power & Energy Systems

INTERNATIONAL JOURNAL OF ELECTRICAL POWER & ENERGY SYSTEMS
ScienceDirect

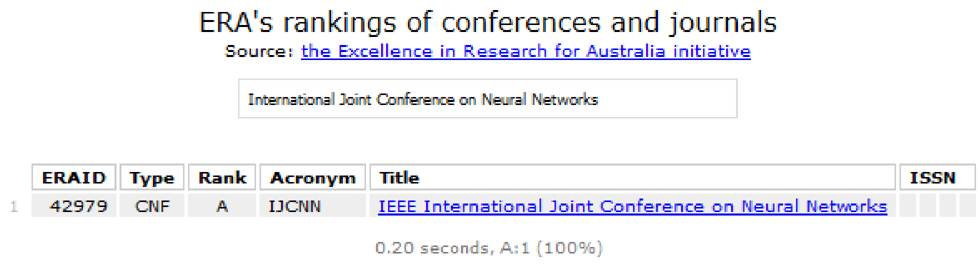
Journal

5 Year Impact Factor: 3.111

Co-Editors::	T.S. Dillon
Editorial Board:	Show
ISSN:	01420615
Volumes:	74-83
Issues:	10

Go to Journal Website

Figure A.2.: Impact Factor of *International Journal of Electrical Power and Energy Systems* (retrieved September 2015)


ERA's rankings of conferences and journals

Source: the Excellence in Research for Australia initiative

International Joint Conference on Neural Networks

	ERAID	Type	Rank	Acronym	Title	ISSN
1	42979	CNF	A	IJCNN	IEEE International Joint Conference on Neural Networks	

0.20 seconds, A:1 (100%)

Figure A.3.: ERA Ranking of *International Joint Conference on Neural Networks*

A.2 CURRICULUM VITAE

Curriculum Vitae
Dr. Dominik Engel

Personal Details

Name	Dominik Engel
Academic Degree	Dipl.-Ing. Mag. Dr.
Contact Details	Salzburg University of Applied Sciences Urstein Sued 1, 5412 Puch/Salzburg, Austria E-Mail: dominik.engel@en-trust.at Phone: +43 (0) 50 2211 1305

Education

Sep. 2008	Doctoral Degree in Technical Sciences (passed with distinction), University of Salzburg PhD Thesis on "Media Encryption for Still Visual Data" Supervisor: Univ.-Prof. Dr. Andreas Uhl
Apr. 2004	Master's Degree in English and American Literature and Language Studies (passed with distinction), University of Salzburg
Dec. 2002	Master's Degree in Applied Informatics with application domain "Artificial Intelligence" (passed with distinction), University of Salzburg
1998/99	Year abroad at University of East Anglia (UK)
June 1996	A-Levels/Matura (passed with distinction), Bundesrealgymnasium Innsbruck, Sillgasse

Extracurricular Activities

2002	Member of Department Council for Scientific Computing at the University of Salzburg
2001	Member of Department Council for Computer Science at the University of Salzburg
2001–2002	Student representative for Applied Informatics at the University of Salzburg
2000–2002	Member of the Curriculum Council for Applied Informatics at the University of Salzburg
1998	Founding member of non-profit organization "subnet – platform for media culture and experimental technologies"

Academic and Professional Career

since Sep. 2016	Member of the Board of Salzburg Wohnbau Group
since Sep. 2015	Head of Department *Network Technologies and Security* at the Salzburg University of Applied Sciences
since Jan. 2013	Director Josef Ressel Center for User-Centric Smart Grid Privacy, Security and Control
since Dec. 2011	FH-Professor at the Salzburg University of Applied Sciences
Sep. 2010–Nov. 2011	Senior Researcher and Lecturer at the Salzburg University of Applied Sciences
Oct. 2008–Sep. 2010	International Product Manager Content Security at Sony DADC
March 2006–Sep. 2008	Researcher at the department of Computer Sciences at the University of Salzburg
May 2005–Feb. 2006	Austrian Academy of Sciences scholar (DOC Dissertation Grant)
Nov. 2004–April 2005	Research Associate in the research group “Multimedia Signal Processing and Security” at the University of Salzburg
March 2003–Sep. 2004	Research Associate in the DFG Collaborative Research Centre SFB/TR 8 “Spatial Cognition” at the University of Bremen
Nov. 2000–Dez. 2002	Research Assistant in the research group “Multimedia Signal Processing and Security” at the University of Salzburg
2000–2004	Development and support of a forecasting software for “Wind River”
2000–2001	Development and suppport of a CRM software for student bank accounts for “Bank Austria”
1999–2002	Board member and project manager for “subnet”

Research Projects

Jan. 2013–Dec. 2017	Josef Ressel Center for User-Centric Smart Grid Privacy, Security and Control Director, 14 researchers
Jun. 2015–Nov. 2017	RASSA-Architektur – “Reference Architecture for a Secure Smart Grid in Austria” (FFG project no. 848811), Lead Work Package “Secure Consumer Integration”, 3 researchers
Sept. 2015–Aug. 2016	PROMISE – “Process Mining for Intrusion Detection in Smart Energy Grids” (FFG project no. 849914) Consortium Lead, 3 researchers
Apr. 2013–Sep. 2015	INTEGRA – ‘Integrated Smart Grid Reference Architecture of Local Intelligent Distribution Grids and Transregional Virtual Power Plants” (FFG Project No. 838793) Project Lead for University of Applied Sciences, one researcher
Jan. 2012–Dec. 201 4	‘Privacy-protected Video Surveillance on Scalable Bitstreams” (FFG Project No. 832082) Project Lead for University of Applied Sciences
Aug. 2011–Jan. 2012	“Security in Industrial Process Control Systems”, (Collaborative project with company partner Copa-data) Principal Project Lead, two research assistants

Nov. 2006–Sep. 2008	“Adaptive Streaming of Secure & Scalable Wavelet Videos” (FWF project no. P19159) Researcher
Nov. 2004–Feb. 2008	EU Network of Excellence ECRYPT (IST-2002-507932) Researcher
Nov. 2004–Dec. 2005	“Adaptive Security Techniques for Visual Data in Wavelet-based Representation” (FWF project no. P15170) Researcher
Mar. 2003–Sep. 2004	“R1-[ImageSpace]” in the DFG Collaborative Research Centre SFB TR/8 “Spatial Cognition” Researcher
Nov. 2000–Dec. 2002	“Object-based Image and Video Compression with Adaptive and Hybrid Wavelet Techniques” (FWF Project No. P13732) Researcher

Professional Trainings

2012	Leadership (Salzburg University of Applied Sciences)
2011	Cisco Certified Academy Instructor (CCAI)
2010	The Storyboard Approach – Advanced techniques of creating powerful presentations (BCD Business Communication Design, Switzerland)
2009	Leadership Competence (P.E.P. Pleiner, Evers and Partner)
2009	Functional Competence (P.E.P. Pleiner, Evers and Partner)
2009	Optical Storage Media for Engineers (Sony DADC)

Contributions to Standardization

since 2013	CEN/CENELEC/ETSI Smart Grid Co-ordination Group (European Mandate M/490), Working Group ”‘Smart Grid Information Security (SGIS)’”, Work Package 3: Privacy
since 2013	German Commission for Electrical, Electronic & Information Technologies of DIN and VDE (DKE), Steering Committee STD_1911 “Normung E-Energy / Smart Grids”, Working Group STD_1911.11 “Smart Grid Informationssicherheit”’
since 2013	German Association for Electrical, Electronic & Information Technologies (VDE), Working Group “Energy Information Networks” (in German: “Energieinformationsnetze”)
since 2013	Austrian Computer Society (OCG), Working Group “Energy Informatics” (in German: “Energieinformatik”)
since 2014	Austrian Technology Platform for Smart Grids
since 2014	Austrian Electrotechnical Association (ÖVE), Working Group “Smart Grids”

Grants and Awards

2009	Nominated by the University of Salzburg for the "Gesellschaft für Informatik (GI)" dissertation award
2007	ECRYPT BOWS-2 Watermarking Challenge, Ep. 1: 2. Place
2006	EU Network of Excellence in Cryptology ECRYPT research grant
2005	Austrian Academy of Sciences (ÖAW) DOC dissertation grant (acceptance rate 2005: 17%)
2003	German Academic Exchange Service (DAAD) grant for Cognitive Science Summer School at the New Bulgarian University, Sofia, Bulgaria
2003	University of Salzburg merit grant for passing the Master's degree in Applied Informatics with distinction
2001–2002	Austrian Science Fund (FWF) research grant

Talks (Selection)

- *The Interplay of Data Resolution and Privacy in Smart Metering*, Dagstuhl Seminar 16032 "Privacy and Security in Smart Energy Grids", 2016
- *Privacy-preserving Smart Metering: Methods and Applicability*, Keynote – Communications for Energy Workshop, Vienna, 2013
- *Privacy and Security Challenges in the Privacy and Security Challenges in the Smart Grid User Domain*, Keynote – 1st ACM Workshop on Information Hiding and Multimedia Security, Montpellier, France, 2013
- *Privacy Challenges in Smart Grids*, Panel Session on Smart Grid Security, IEEE ISGT EU 2014, Istanbul, Turkey
- Panelist Round Table *Sichere IKT Architektur im Smart Grid* (in German), Session "Sicherheit, Systemkontrolle und Versorgungssicherheit", Smart Grids Week, Salzburg, Austria, 2013
- *Datenschutz im Smart Metering: Herausforderungen und Lösungsansätze* (in German), VDE Smart Grid Forum,
  Hannover Messe (Industry trade show on industrial automation, energy, industrial supply and more), Germany, 2014
- Panelist Round Table *Smart Metering – hemmen Privacy Bedenken den technischen Fortschritt?* (in German), Session "Kunden und Märkte", Smart Grids Week, Graz, Austria, 2014
- *Status der europäischen Standardisierung für IT-Security und Privacy im Smart Grid* (in German), Österreichs Energie, Vienna, Austria, 2014
- *Sichere IKT-Architektur im Smart Grid* (in German), Österreichs Energie, Vienna, Austria, 2013
- *Datenschutz und -sicherheit im intelligenten Stromnetz* (in German), Lecture series "Anwendungen in Wirtschaft und Technik", University of Salzburg, Austria, 2013
- *Video Processing Activities and Applied Research at Sony DADC*, with M. Aster, Invited Talk – 6th International Symposium on Image and Signal Processing and Analysis (ISPA '09), Salzburg, Austria, 2009

Board Memberships and Reviewing Activities

- Associate Editor for Springer EURASIP Journal on Information Security
- General Chair and Program Committee Chair for Conference on Availability, Reliability and Security (ARES) 2016 (together with Stephen Wicker, Cornell University)
- Chair for Special Session on "Security and Privacy Technologies for Intelligent Energy Networks" at ACM IHMMSEC 2014 (together with Zekeriya Erkin, TU Delft)
- Reviewer for URSES+ Research Program for the Netherlands Organisation for Scientific Research (NWO)
- Reviewing for Journals
  Communications of the ACM, Wiley Journal of Software: Evolution and Process, IEEE Transactions on Emerging Topics in Computational Intelligence, IEEE Transactions on Dependable and Secure Computing, Springer EURASIP Journal on Information Security (Associate Editor), IEEE Transactions on Smart Grid, Elsevier Signal Processing: Image Communication, Elsevier Journal of Information Security and Applications, IEEE Transactions on Industrial Electronics, IEEE Transactions on Industrial Informatics, MDPI Energies, IEEE Transactions on Image Processing, Hindawi International Journal of Distributed Sensor Networks, IEEE Transactions on Circuits and Systems for Video Technology, Journal of Computing and Information Technology, IEEE Transactions on Information Forensics and Security, IET Journal on Image Processing, EURASIP Journal on Image and Video Processing, IET Journal on Information Security, International Journal of Image and Graphics
- Conference Technical Program Committees
  "DACH Energieinformatik" 2017, IEEE Innovative Smart Grid Technologies (ISGT) 2017, Conference on Availability, Reliability and Security (ARES) 2017, IEEE International Conference on Industrial Informatics (INDIN) 2016, Conference on Availability, Reliability and Security (ARES) 2016, "DACH Energieinformatik" 2016, IEEE Symposium on Industrial Electronics (ISIE) 2016, IEEE Emerging Technologies and Factory Automation (ETFA) 2016, Conference on Availability, Reliability and Security (ARES) 2015, International Workshop on Multimedia Forensics and Security (MFSec) 2015, Smart Energy Grid Security Workshop (SEGS) 2014, IEEE IECON 2014, ACM IHMMSEC 2014, "GI Sicherheit" 2014, "DACH Security" 2014, "DACH Security" 2013, IEEE IECON 2013, IEEE IWIES 2013, Workshop in Information Security Theory and Practice (WISTP) 2004, European Conference on Artificial Intelligence (ECAI) 2004, Member of the organizing committee und Session Chair "Media Encryption" of "International Conference on Communications and Multimedia Security" (CMS) 2005, Member of the organizing committee of ECRYPT Summer School on Multimedia Security 2005, Member of the review and organisation committee of International Conference on Spatial Cognition 2003, Member of the organizing committee of International Workshop on Spatial and Visual Components in Mental Reasoning About Large-Scale Spaces 2003

Teaching

- "Mobile Networks and Security" (Lecture Part on IT-Security), Salzburg University of Applied Sciences (since 2015)
- "Energy Informatics Fundamentels: Network and Communication Technologies", Salzburg University of Applied Sciences (since 2015)
- "Network Reliability and Virtualization", Salzburg University of Applied Sciences (since 2013)
- "Internet Infrastructure and Security", Salzburg University of Applied Sciences (since 2013)
- "Cryptology", Salzburg University of Applied Sciences (since 2006)
- "Master Seminar", Salzburg University of Applied Sciences (2011)
- "Network Reliability and Security", Salzburg University of Applied Sciences (2011–2012)
- "Mobile & Distribution Networks", Salzburg University of Applied Sciences (2010-2012)
- "Multimedia Technologies", Salzburg University of Applied Sciences (2010–2013)

- "Distributed and Autonomous Systems", Salzburg University of Applied Sciences (2009)
- "Advanced Topics in Databases", University of Salzburg (2008)
- "Database Systems", University of Salzburg (2007–2008)
- "Introduction to Unix Systems", University of Salzburg (2006–2008)
- "Software Project", University of Bremen (2003–2004)
- "Software Development", University of Bremen (2003)

Supervision of Student Theses (Selection)

- Joris Lückenga (Salzburg University of Applied Sciences, MSc., 2015), *Reduction of False Positives in Smart Grid Intrusion Detection*, joint supervision with Robert C. Green, Bowling Green State University, USA
- Christian Peer (Salzburg University of Applied Sciences, MSc., 2014), *Secure Signal Processing for Smart Grid Privacy*, joint supervision with Stephen Wicker, Cornell University, USA
- Fabian Knirsch (Salzburg University of Applied Sciences, MSc., 2014), *Generic Data Models and Semantic Retrieval in Smart Grid IT Infrastructures*, joint supervision with Victor Prasanna, University of Southern California, USA,
- Wolfgang Lausenhammer (Salzburg University of Applied Sciences, MSc., 2014), *User-Centric Simulation of Demand Response Optimization*, joint supervision with Robert C. Green, Bowling Green State University, USA
- Kaibin Bao (Karlsruhe Institute of Technology), *Measuring Information Disclosure in Load Monitoring Data by Disaggregation of Sum Load Profiles*, Mentor ("'Shepherd"') for the Energy Informatics PhD Workshops 2014, Dissertation Supervisor: Hartmut Schmeck
- René Blaschke (Salzburg University of Applied Sciences, MSc., 2012), *Entwurf und Implementierung einer sicheren IKT-Architektur für Smart Grids*
- Christian Peuker (Salzburg University of Applied Sciences, MSc., 2012), *Kommunikationssicherheit in einer Smart Metering Infrastructure*
- Markus Schober (Salzburg University of Applied Sciences, MSc., 2012), *Evaluation of privacy protection methods with JPEG2000 ROI encryption in video surveillance scenarios*
- Adnan Srna (Salzburg University of Applied Sciences, MSc., 2012), *Selective Encryption Methods for Securing Multi-Resolution Smart Meter Data*
- Michael Lechner (University of Salzburg, MSc., 2009), *Object persistence and object relational mapping*, joint supervision with Helge Hagenauer, Univ. of Salzburg

Publications

Journal Publications

[1] F. Knirsch, G. Eibl, and D. Engel. Multi-resolution privacy-enhancing technologies for smart metering. *EURASIP Journal on Information Security*, 2017(1):6, 2017.

[2] C. Neureiter, D. Engel, and M. Uslar. Domain specific and model based systems engineering in the smart grid as prerequesite for security by design. *Electronics*, 5(2):24, 2016.

[3] A. Unterweger, F. Knirsch, G. Eibl, and D. Engel. Privacy-preserving load profile matching for tariff decisions in smart grids. *EURASIP Journal on Information Security*, 2016(1):1–17, 2016.

[4] D. Engel and G. Eibl. Wavelet-based multiresolution smart meter privacy. *IEEE Transactions on Smart Grid*, PP(99):1–12, 2016. preprint.

[5] W. Lausenhammer, D. Engel, and R. Green. Utilizing capabilities of plug in electric vehicles with a new demand response optimization software framework: Okeanos. *International Journal of Electrical Power and Energy Systems*, 75:1–7, 2016.

[6] G. Eibl and D. Engel. Influence of data granularity on smart meter privacy. *IEEE Transactions on Smart Grid*, 6(2):930–939, March 2015.

[7] A. Unterweger and D. Engel. Resumable load data compression in smart grids. *IEEE Transactions on Smart Grid*, 6(2):919–929, March 2015.

[8] C. Neureiter, G. Eibl, D. Engel, S. Schlegel, and M. Uslar. A concept for engineering smart grid security requirements based on SGAM models. *Computer Science - Research and Development*, pages 1–7, 2014.

[9] S. Auer, A. Bliem, D. Engel, A. Uhl, and A. Unterweger. Bitstream-based JPEG encryption in real-time. *International Journal of Digital Crime and Forensics*, 5(3):1–14, 2013.

[10] D. Engel, T. Stütz, and A. Uhl. Assessing JPEG2000 encryption with key-dependent wavelet packets. *EURASIP Journal on Information Security*, 2012(1):1–16, 2012.

[11] D. Engel, T. Stütz, and A. Uhl. A survey on JPEG2000 encryption. *Multimedia Systems*, 15(4):243–270, 2009. Springer.

[12] R. Kutil and D. Engel. Methods for the anisotropic wavelet packet transform. *Applied and Computational Harmonic Analysis*, 25(3):295–314, 2008.

[13] D. Engel, E. Pschernig, and A. Uhl. An analysis of lightweight encryption schemes for fingerprint images. *IEEE Transactions on Information Forensics and Security*, 3(2):173–182, June 2008.

[14] D. Engel, T. Stütz, and A. Uhl. Format-compliant JPEG2000 encryption in jpsec: Security, applicability and the impact of compression parameters. *EURASIP Journal on Information Security*, (Article ID 94565):doi:10.1155/2007/94565, 20 pages, 2007.

Book Chapters and Contributions to Collections

[15] F. Knirsch, A. Unterweger, G. Eibl, and D. Engel. Privacy-preserving smart grid tariff decisions with block-chain-based smart contracts. In W. Rivera, editor, *Sustainable Cloud and Energy Services: Principles and Practices*. Springer International Publishing, 2017. to appear.

[16] G. Eibl and D. Engel. Differential privacy for real smart metering data. *Computer Science – Research and Development*, 32(1):173–182, 2017.

[17] F. Knirsch, D. Engel, C. Neureiter, M. Frincu, and V. Prasanna. Privacy assessment of data flow graphs for an advanced recommender system in the smart grid. In O. Camp, E. Weippl, C. Bidan, and E. Aïmeur, editors, *Information Systems Security and Privacy – Revised and Selected Papers of ICISSP 2015*, volume 576 of *Communications in Computer and Information Science*, pages 89–106. Springer International Publishing, 2016. Best Paper Award.

[18] C. Dänekas, C. Neureiter, S. Rohjans, M. Uslar, and D. Engel. Towards a model-driven-architecture process for smart grid projects. In P. Benghozi, D. Krob, A. Lonjon, and H. Panetto, editors, *Digital Enterprise Design & Management*, volume 261 of *Advances in Intelligent Systems and Computing*, pages 47–58. Springer International Publishing, 2014.

[19] R. Blaschke, S. Suhrer, and D. Engel. Serviceorientierte Architekturen für Smart Grids. In J. Hofmann and C. Felden, editors, *IT für Smart Grids*, volume 50 of *Praxis der Wirtschaftsinformatik*, pages 16–25. HMD, 2013. In German.

[20] D. Engel. Media encryption for still visual data. In D. Wagner, A. Bernstein, T. Dreier, S. Hoelldobler, G. Hotz, K.-P. Loehr, P. Molitor, G. Neumann, R. Reischuk, D. Saupe, M. Spiliopoulou, and H. Stoerrle, editors, *Ausgezeichnete Informatikdissertationen 2008*, Lecture Notes in Informatics, pages 81–90. Berlin: Springer, 2009. ISBN 978-88579-413-4.

[21] D. Engel, T. Stütz, and A. Uhl. Efficient transparent JPEG2000 encryption. In C.-T. Li, editor, *Multimedia Forensics and Security*, chapter 16, pages 336–359. Idea, 2007. ISBN 978-159904869-7.

Refereed Conference Proceedings

[22] K. Böhmer, F. Stertz, T. Hildebrandt, S. Rinderle-Ma, G. Eibl, C. Ferner, S. Burkhart, and D. Engel. Application and Testing of Business Processes in the Energy Domain. In *Fachtagung Datenbanksysteme für Business, Technologie und Web (BTW)*, pages 25–32, 2017.

[23] G. Eibl, C. Ferner, T. Hildebrandt, F. Stertz, S. Burkhart, S. Rinderle-Ma, and D. Engel. Exploration of the potential of process mining for intrusion detection in smart metering. In *3rd International Conference on Information Systems Security and Privacy*, 2017. to appear.

[24] A. Veichtlbauer, O. Langthaler, D. Engel, C. Kasberger, F. P. Andrén, and T. Strasser. Towards Applied Security-by-Design for DER Units. In *Proceedings of the 21st IEEE International Conference on Emerging Technologies and Factory Automation (ETFA 2016)*, 2016. to appear.

[25] C. Neureiter, M. Uslar, D. Engel, and G. Lastro. A standards-based approach for domain specific modelling of smart grid system architectures. In *Proceedings of International Conference on System of Systems Engineering (SoSE) 2016*, pages 1–6, Kongsberg, Norway, June 2016. Best Paper Award.

[26] J. Lückenga, D. Engel, and R. Green. Weighted vote algorithm combination technique for anomaly based smart grid intrusion detection systems. In *Proceedings of International Joint Conference on Neural Networks (IJCNN) 2016*, pages 2738–2742, Vancouver, Canada, July 2016.

[27] A. Unterweger, D. Engel, and M. Ringwelski. The effect of data granularity on load data compression. *Energy Informatics 2015*, 9424:69–80, 2015.

[28] M. Uslar and D. Engel. Towards generic domain reference designation: How to learn from smart grid interoperability. In *Poster Proceedings of DACH Energy Informatics 2015*, pages 1–12, 2015.

[29] J. Schwarzer and D. Engel. Evaluation of data communication requirements for common demand response models. In *Proceedings of IEEE International Conference on Industrial Technology (ICIT) 2015*, pages 1311–1316, Seville, Spain, 2015. IEEE.

[30] M. Pichler, A. Veichtlbauer, and D. Engel. Evaluation of OSGi-based architectures for customer energy management systems. In *Proceedings of IEEE International Conference on Industrial Technology (ICIT) 2015*, pages 2455–2460, Seville, Spain, 2015. IEEE.

[31] G. Eibl, D. Engel, and C. Neureiter. Privacy-relevant smart metering use cases. In *Proceedings of IEEE International Conference on Industrial Technology (ICIT) 2015*, pages 1387–1392, Seville, Spain, 2015. IEEE.

[32] W. Lausenhammer, D. Engel, and R. Green. A game theoretic software framework for optimizing demand response. In *Proceedings of the 6th Conference on Innovative Smart Grid Technologies (ISGT)*, pages 1–5, Feb 2015.

[33] F. Knirsch, D. Engel, M. Frincu, and V. Prasanna. Model-based assessment for balancing privacy requirements and operational capabilities in the smart grid. In *Proceedings of the 6th Conference on Innovative Smart Grid Technologies (ISGT)*, pages 1–5, Feb 2015.

[34] F. Knirsch, D. Engel, C. Neureiter, M. Frincu, and V. Prasanna. Model-driven privacy assessment in the smart grid. In *Proceedings of the 1st International Conference on Information Systems Security and Privacy (ICISSP)*, pages 173–181, Feb 2015. Best Paper Award.

[35] C. Peer, D. Engel, and S. Wicker. Hierarchical key management for multi-resolution load data representation. In *Proceedings of 5th IEEE International Conference on Smart Grid Communications (SmartGridComm 2014)*, pages 926–932, Venice, Italy, Nov. 2014. IEEE.

[36] C. Neureiter, S. Rohjans, D. Engel, C. Dänekas, and M. Uslar. Addressing the complexity of distributed smart city systems by utilization of model driven engineering concepts. In *Proceedings VDE Kongress 2014*, pages 1–6, Oct. 2014.

[37] C. Neureiter, D. Engel, J. Trefke, R. Santodomingo, S. Rohjans, and M. Uslar. Towards consistent smart grid architecture tool support: From use cases to visualization. In *Proceedings of IEEE Innovative Smart Grid Technologies (ISGT) 2014*, Istanbul, Turkey, Oct. 2014. IEEE.

[38] N. Egger, C. Neureiter, and D. Engel. Adopting an SGAM based demand side management architecture for the realization of ambient assisted living. In *Proceedings VDE Kongress 2014*, pages 1–6, Oct. 2014.

[39] C. Dänekas, D. Engel, C. Neureiter, S. Rohjans, J. Trefke, and M. Uslar. Durchgängige Werkzeugunterstützung für das EU Mandat M/490: Vom Anwendungsfall bis zur Visualisierung. In *Proceedings VDE Kongress 2014*, pages 1–6, Oct. 2014. In German.

[40] C. Peuker and D. Engel. Praxistaugliche Kommunikationssicherheit in einer Smart Metering Infrastruktur. In J. Lüthi and H.-P. Steinebacher, editors, *Proc. 8. Forschungsforum der Österreichischen Fachhochschulen*, pages 72–76, 2014. In German.

[41] R. Blaschke and D. Engel. Flexible und sichere Integration von Smart Grid Use-Cases in einer serviceorientierten IKT-Architektur. In J. Lüthi and H.-P. Steinebacher, editors, *Proc. 8. Forschungsforum der Österreichischen Fachhochschulen*, pages 42–46, 2014. In German.

[42] G. Eibl and D. Engel. Influence of data granularity on nonintrusive appliance load monitoring. In *Proceedings of the Second ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec '14)*, pages 147–151, Salzburg, Austria, 2014. ACM.

[43] J. Schwarzer, A. Kiefel, and D. Engel. The role of user interaction and acceptance in a cloud-based demand response model. In *Proc. IEEE IECON 2013, Special Session on Energy Informatics*, pages 4797–4802, Vienna, Austria, Nov. 2013. IEEE.

[44] C. Neureiter, G. Eibl, A. Veichtlbauer, and D. Engel. Towards a framework for engineering smart-grid-specific privacy requirements. In *Proc. IEEE IECON 2013, Special Session on Energy Informatics*, pages 4803 – 4808, Vienna, Austria, Nov. 2013. IEEE.

[45] D. Engel and G. Eibl. Multi-resolution load curve representation with privacy-preserving aggregation. In *Proceedings of IEEE Innovative Smart Grid Technologies (ISGT) 2013*, pages 1–5, Copenhagen, Denmark, Oct. 2013. IEEE.

[46] A. Veichtlbauer, D. Engel, F. Knirsch, O. Langthaler, and F. Moser. Advanced metering and data access infrastructures in smart grid environments. In *Proc. 7th International Conference on Sensor Technologies and Applications*, pages 63–68, Barcelona, Spain, Aug. 2013.

[47] D. Engel. Wavelet-based load profile representation for smart meter privacy. In *Proceedings IEEE PES Innovative Smart Grid Technologies (ISGT'13)*, pages 1–6, Washington, D.C., USA, Feb. 2013. IEEE.

[48] D. Engel, A. Uhl, and A. Unterweger. Region of interest signalling for encrypted jpeg images. In *Proceedings of the first ACM workshop on Information hiding and multimedia security (IHMMSEC '13)*, pages 165–174, Montpellier, France, 2013. ACM.

[49] S. Schinwald, D. Engel, and M. Seidler. Efficient automated liquid detection in microplates. In *Proceedings of 25th IEEE International Computer-Based Medical Systems (CBMS) Symposium*, pages 1–4, Rome, Italy, June 2012.

[50] D. Engel. Conditional access smart meter privacy based on multi-resolution wavelet analysis. In *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies*, pages 45:1–45:5, New York, NY, USA, 2011. ACM.

[51] D. Engel, T. Stütz, and A. Uhl. Evaluation of JPEG2000 hashing for efficient authentication. In *Proceedings of International Conference on Multimedia & Expo, ICME '09*, pages 1728–1731, New York, NY, USA, June 2009.

[52] D. Engel, T. Stütz, and A. Uhl. Efficient transparent JPEG2000 encryption with format-compliant header protection. In *Proceedings of IEEE International Conference on Signal Processing and Communications, ICSPC '07*, Dubai, UAE, Nov. 2007.

[53] D. Engel, T. Stütz, and A. Uhl. Format-compliant JPEG2000 encryption with combined packet header and packet body protection. In *Proceedings of ACM Multimedia and Security Workshop, MM-SEC '07*, pages 87–95, Dallas, TX, USA, Sept. 2007.

[54] D. Engel and A. Uhl. An attack against image-based selective bitplane encryption. In *Proceedings of the IEEE International Conference on Image Processing, ICIP '07*, volume II, pages 141–144, San Antonio, TX, USA, Sept. 2007. IEEE.

[55] D. Engel and A. Uhl. An evaluation of lightweight JPEG2000 encryption with anisotropic wavelet packets. In E. J. Delp and P. W. Wong, editors, *Security, Steganography, and Watermarking of Multimedia Contents IX*, Proceedings of SPIE, pages 65051S1–65051S10, San Jose, CA, USA, Jan. 2007. SPIE.

[56] D. Engel, R. Kutil, and A. Uhl. A symbolic transform attack on lightweight encryption based on wavelet filter parameterization. In *Proceedings of ACM Multimedia and Security Workshop, MM-SEC '06*, pages 202–207, Geneva, Switzerland, Sept. 2006.

[57] D. Engel and A. Uhl. Lightweight JPEG2000 encryption with anisotropic wavelet packets. In *Proceedings of International Conference on Multimedia & Expo, ICME '06*, pages 2177–2180, Toronto, Canada, July 2006.

[58] S. Bertel, T. Barkowsky, D. Engel, and C. Freksa. Computational modeling of reasoning with mental images: basic requirements. In *Proceedings of the 7th International Conference on Cognitive Modeling (ICCM 2006)*, pages 50–55, Trieste, Italy, Apr. 2006.

[59] D. Engel and A. Uhl. Secret wavelet packet decompositions for JPEG 2000 lightweight encryption. In *Proceedings of the 31st International Conference on Acoustics, Speech, and Signal Processing, ICASSP '06*, volume V, pages 465–468, Toulouse, France, May 2006.

[60] D. Engel and A. Uhl. Security enhancement for lightweight JPEG 2000 transparent encryption. In *Proceedings of Fifth International Conference on Information, Communication and Signal Processing, ICICS '05*, pages 1102–1106, Bangkok, Thailand, Dec. 2005.

[61] D. Engel and A. Uhl. Parameterized biorthogonal wavelet lifting for lightweight JPEG 2000 transparent encryption. In *Proceedings of ACM Multimedia and Security Workshop, MM-SEC '05*, pages 63–70, New York, NY, USA, Aug. 2005.

[62] D. Engel, S. Bertel, and T. Barkowsky. Spatial principles in control of focus in reasoning with mental representations, images, and diagrams. In *Spatial Cognition IV. Reasoning, Action, and Interaction*, Lecture Notes in Computer Science, pages 181–203. Berlin: Springer, 2005.

[63] T. Barkowsky, S. Bertel, D. Engel, and C. Freksa. Design of an architecture for reasoning with mental images. In *Proceedings of the International Workshop on Spatial and Visual Components in Mental Reasoning about Large-scale Spaces*, Bad Zwischenahn, Germany, Sept. 2003.

[64] D. Engel and A. Uhl. Adaptive image compression of arbitrarily shaped objects using wavelet packets. In *Proceedings of the 23rd International Picture Coding Symposium 2003 (PCS 2003)*, pages 283–288, St. Malo, France, Apr. 2003.

[65] D. Engel and A. Uhl. Adaptive object-based image compression using wavelet packets. In *Proceedings of the 4th International Symposium on Video/Image Processing and Multimedia Communications (VIPromCom 2002)*, pages 183–187, Zadar, Croatia, June 2002.

Invited Papers (Non-refereed) and Other Non-refereed Contributions

[66] D. Engel. Privacy and security challenges in the smart grid user domain (invited talk). In *Proceedings of the first ACM workshop on Information hiding and multimedia security (IH&MMSec '13)*, pages 85–86, Montpellier, France, 2013. ACM.

[67] D. Engel. Privacy-preserving smart metering: Methods and applicability (invited talk). In *Proceedings of the fourth Workshop on Communications for Energy Systems*, pages 9–16, Vienna, Austria, Sept. 2013. Austrian Electrotechnical Association.

[68] F. Fredersdorf, J. Schwarzer, and D. Engel. Die Sicht der Endanwender im Smart Meter Datenschutz. *Datenschutz und Datensicherheit - DuD*, 39(10):682–686, 2015.

Patents

[69] M. Spitzlinger, A. Winter, W. Hinterhölzl, P. Eisenmann, K. Holzapfel, D. Engel, and J. Grünberger. Method for copy protection, Sony DADC, EP2010/003073, 2010.

Other Formats

[70] D. Engel. *Media Encryption for Still Visual Data – An Analysis of Selected Techniques for Natural Images and Fingerprint Data in the Spatial and Wavelet Domain*. PhD thesis, Department of Computer Sciences, University of Salzburg, Austria, June 2008.

[71] D. Engel. Modelling self and other – a hybrid approach to the analysis of images of self and other in the radio addresses delivered by the american president before and after 9/11. Master's thesis, Department of English and American Studies, University of Salzburg, Austria, 2004.

[72] D. Engel. Adaptive object-based image compression with wavelet methods. Master's thesis, Department of Scientific Computing, University of Salzburg, Austria, 2002.

[73] F. Knirsch, D. Engel, C. Neureiter, M. Frincu, and V. Prasanna. Model-driven privacy assessment in the smart grid. Technical Report 2014-01, Josef Ressel Center for User-Centric Smart Grid Privacy, Security and Control, July 2014.

[74] P. Eder, D. Engel, and A. Uhl. JPEG2000-based scalable video coding with MCTF. Technical report, Department of Computer Sciences, University of Salzburg, Austria, 2007.

[75] S. Bertel, T. Barkowsky, and D. Engel. The specification of the casimir architecture. Technical report, R1-[ImageSpace], SFB/TR8 Spatial Cognition; http://www.sfbtr8.uni-bremen.de/project/r1/, Bremen, Germany, 2004.

Salzburg, May 15, 2017

COLOPHON

This document was typeset with LaTeX 2$_\varepsilon$. The *TeX Live* system was used with the collection of *KOMA-Script* packages and numerous other valuable packages from various authors, acquired through the CTAN network. The layout is inspired by and based on the layout of the *ClassicThesis* package by André Miede (available through CTAN).

The main typeface is Adobe *Minion Pro*, designed by Robert Slimbach. The mathematical symbols are set with *MnSymbol* by Achim Blumensath. The chapter numbers use the *Euler* typeface, designed by Hermann Zapf (assisted by Donald Knuth).