THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Enhancing Privacy in the Advanced Metering Infrastructure: Efficient Methods, the Role of Data Characteristics and Applications

VALENTIN TUDOR

Division of Networks and Systems
Systems Security and Distributed Computing and Systems
Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2017

Enhancing Privacy in the Advanced Metering Infrastructure:
Efficient Methods, the Role of Data Characteristics and Applications

Valentin Tudor

ISBN 978-91-7597-632-7
Doktorsavhandlingar vid Chalmers tekniska högskola
Ny series nr. 4313

ISSN: 0346-718X
Technical report 140D
Department of Computer Science and Engineering
Systems Security and Distributed Computing and Systems
CHALMERS UNIVERSITY OF TECHNOLOGY
SE-412 96 GÖTEBORG, Sweden
Phone: +46 (0)31-772 10 00

Author e-mail: `tudor@chalmers.se`

Printed by Chalmers Reproservice
Göteborg, Sweden 2017

Enhancing Privacy in the Advanced Metering Infrastructure: Efficient Methods, the Role of Data Characteristics and Applications

Valentin Tudor

Department of Computer Science and Engineering, Chalmers University of Technology

ABSTRACT

Large quantities of data are produced and collected by computing and communication devices in cyber-physical systems. Information extracted from these data opens new possibilities but also raises privacy issues.

The characteristics of these data play an important role in the efficiency of privacy-enhancing technologies thus grasping the former's influence is a step forward in improving the latter. Privacy-enhanced data can be employed in cyber-physical systems' applications and their utility can be improved by fine-tuning the parameters of the privacy-enhancing technologies applied to the data. This can be coupled with an analysis of the efficiency of applications that employ privacy-enhanced preprocessed data for better insights on the trade-off between applications' utility and data privacy. Orthogonal to this, privacy-enhanced data originating from cyber-physical systems can be employed in monitoring solutions for cyber security. This is a step forward in fulfilling both the confidentiality and privacy requirements for these complex systems.

This thesis focuses on privacy in the context of the Advanced Metering Infrastructure (AMI) in the smart electrical grid and it has three primary objectives. The first is to study the characteristics of AMI datasets and how they influence the efficiency of privacy-enhancing technologies. The second objective is to identify methods and efficient algorithmic implementations, in connection to what can be deployed in contemporary hardware, as needed for Internet of Things-based systems. The third objective is to study the balance between confidentiality requirements and the requirement to monitor the communication network for intrusion detection, as an example.

This thesis advances the current research by showing (i) how different AMI privacy-enhancing techniques complement each other, (ii) how datasets' characteristics can be tuned in order to improve the efficiency of these techniques and (iii) how the need for privacy can be balanced with the need to monitor the AMI communication network.

Keywords: Advanced Metering Infrastructure, data privacy, data characteristics, applied differential privacy, communication security, system security, intrusion detection.

To my lovely daughter Smaranda and my dear wife Simona

Acknowledgments

First, I would like to express my gratitude to my supervisors, Dr. Marina Papatriantafilou and Dr. Magnus Almgren for their constant guidance, encouragement and understanding throughout my work leading to this thesis. My work has been partially supported by the European Comission Seventh Framework Programme (FP7/2007-2013) through the FP7-SEC-285477-CRISALIS project and the collaboration framework of Chalmers Energy Area of Advance.

Special thanks go to Iosif and Thomas, for all the discussions, fun and support in the five (looong :)) years we shared an office together. I also would like to give my appreciation to the current and former members of the Networks and Systems division for the wonderful and friendly working environment. Thank you Adones, Ali, Anders, Amir, Aras, Aljoscha, Andreas, Babis, Bapi, Bei, Beshr, Boel, Carlo, Christos, Daniel, Elad, Elena P., Erland, Farnaz, Georgios, Hannah, Ioannis, Ivan, Joris, Katerina, Mustafa, Nasser, Nhan, Olaf, Oscar, Paul, Peter L., Philippas, Pierre, Thomas R., Tomas, Valentin P., Viktor, Vincenzo, Wissam and Zhang. Thank you for all the discussions, support and feedback you have provided so far. I would also like to thank Anneli, Eva, Jonna, Marianne, Peter H., Rolf, Rebecca and Tiina for their great administrative support.

I would like also to thank all the people that made Göteborg a sunnier place: Alin, Amrish, Andy, Claudia, Costel, Cristina, Daniel Ch., Daniel Ci., Elena B., Irina, Laurenţiu, Lavinia, Maria, Mihai B., Mihai C., Oana, Rocio, Ruxandra, Vasiliki, Vlad. Thank you all for all the great time!

Last, but not least, I would like to thank my family for their constant encouragement, support and understanding over all these years — *Vã mulţumesc pentru tot!*
My greatest thanks go to my lovely daughter Smaranda and my lovely wife Simona, for their endless love and support. Thank you, I could not have accomplished this without you!

Valentin Tudor
Göteborg, October 2017

List of Appended Papers

Paper I. **Valentin Tudor**, Magnus Almgren, Marina Papatriantafilou,
"The Influence of Dataset Characteristics on Privacy Preserving Methods in Advanced Metering Infrastructure", *under submission*, preliminary results presented in publications V and VI (please see next page).

Paper II. **Valentin Tudor**, Vincenzo Gulisano, Magnus Almgren, Marina Papatriantafilou,
"BES: Differentially Private Event Aggregation for large-scale IoT-based Systems", *under submission*, preliminary results presented in publication VII (please see next page).

Paper III. **Valentin Tudor**, Magnus Almgren, Marina Papatriantafilou,
"Employing Private Data in AMI Applications: Short Term Load Forecasting Using Differentially Private Aggregated Data", in *Advanced and Trusted Computing, 2016 Intl. IEEE Conference*, Toulouse, France, July 18—21, 2016 pp. 404—413.

Paper IV. **Valentin Tudor**, Magnus Almgren, Marina Papatriantafilou,
"Harnessing the Unknown in Advanced Metering Infrastructure Traffic", in *Proceedings of the 30th ACM/SIGAPP Symposium On Applied Computing (SAC 2015)*, Salamanca, Spain, April 13—17, 2015, pp. 2204—2211.

Other publications related to, but not included in the thesis:

Paper V. **Valentin Tudor**, Magnus Almgren, Marina Papatriantafilou,
"Analysis of the Impact of Data Granularity on Privacy for the Smart Grid", in *Proceedings of the 12th ACM workshop on privacy in the electronic society (WPES 2013)*, Berlin, Germany, November 4, 2013, pp. 61—70.

Paper VI. **Valentin Tudor**, Magnus Almgren, Marina Papatriantafilou,
"A Study on Data De-pseudonymization in the Smart Grid", in *Proceedings of the eighth European Workshop on Systems Security (EuroSec 2015)*, Bordeaux, France, April 21, 2015, Article No. 2.

Paper VII. Vincenzo Gulisano, **Valentin Tudor**, Magnus Almgren, Marina Papatriantafilou,
"BES: Differentially Private and Distributed Event Aggregation in Advanced Metering Infrastructures", in *Proceedings of the 2nd ACM International Workshop on Cyber-Physical System Security (CPSS 2016)*, Xi'an, China, May 30, 2016, pp. 59—69.

Paper VIII. Mihai Costache, **Valentin Tudor**, Magnus Almgren, Marina Papatriantafilou,
"Remote Control of Smart Meters: Friend or Foe?", in *Computer Network Defense (EC2ND), 2011 Seventh European Conference on*, Gothenburg, Sweden, September 6, pp. 49—56.

List of Acronyms

AMI	advanced metering infrastructure
COSEM	companion specification for energy metering
CPS	cyber-physical system
DLMS	device language message specification
DoS	denial of service
DDoS	distributed denial of service
DSO	distribution system operator
DP	differential privacy
ECR	encrypted command recognition
GDPR	general data protection regulation
HF	high-frequency
ID	identity
IDS	intrusion detection system
IoT	Internet of Things
kWh	kilowatt hour
LF	low-frequency
M-Bus	meter-bus
MAPE	mean absolute percentage error
NILM	nonintrusive load monitoring
PCA	principal component analysis
PET	privacy enhancing technology
PII	personally identifiable information
RQ	research question
SCADA	supervisory control and data acquisition
STLF	short term load forecasting
SG	smart grid
SM	smart meter

Contents

List of Figures

Part I

Introduction

Chapter 1

Overview

Nowadays, large quantities of data are produced every minute [47] with the help of devices connected to the Internet. Through their analysis, these large quantities of data can offer superior understanding of the sectors where they are collected from, but this comes at a cost: "with big data comes great responsibility" [49]. The information that can be extracted from these data raises privacy and ethical concerns [12] which need to be addressed and alleviated in the data collection process in order to minimize their impact on the data producers.

1.1 Motivation

Different sectors of our society are penetrated by interconnected lightweight devices which monitor and control different aspects of our lives. All these devices form what is called *Internet of Things (IoT)* [2], but the systems that the devices support are also commonly known as *smart environments* [21], *cyber-physical systems* [61] or *Industry 4.0* (Industrie 4.0) [50], depending on their capabilities and the specific locations where they operate. The estimation is that by 2020 there will be around 20 billion IoT devices installed, a big portion being comprised of consumer products.[1] Internet of Things devices slowly appear in locations and near equipment that were traditionally disconnected from the Internet. For example, the automotive industry is promoting interconnected vehicles, and it is envisioned that Internet connected cars will become standard in the future [41]. In the energy sector, cyber-physical systems bring advantages to the process of monitoring the production and delivery of energy, provide easier integration of

[1] `http://www.gartner.com/newsroom/id/3165317`

local renewable energy sources [81], and allow for a better management of the balance between energy production and consumption. Internet of Things devices transform residences into smart homes [19], improving comfort and making the dwellers more informed about their electrical energy consumption patterns and helping to reduce their environmental footprint.

All these improvements and benefits come at a cost, as the large quantities of data collected by IoT devices raise privacy concerns due to the sensitive information that can be inferred from them. Historically, the medical domain is one of the most privacy sensitive, as patients' data need to be stored and processed following strict procedures [74]. With the introduction of medical IoT devices and sensors, patients can be monitored in real-time, improving the health care quality [27], but resulting data need to be collected, transmitted, stored and processed in a secure and privacy-preserving fashion [80]. Similar concerns exist also in the vehicular sector, as data produced and collected in vehicles can be used to infer sensitive information regarding the users' driving style and their whereabouts [93]. Data collected in the electrical energy network also raises privacy concerns, as information regarding the customers' lifestyles can be extracted from the energy consumption patterns [69]. These are only a few examples, but they stress the importance of understanding, studying and overcoming the privacy challenges [38] raised by large data generated in the Internet of Things. Closely connected to the need for privacy is the one to secure and monitor the IoT devices and ensure their correct behavior as their Internet connectivity makes them vulnerable and susceptible to malicious utilization.[2] Devices deployed in cyber-physical systems control infrastructures which are critical for the functioning of our society and their failure or malicious utilization may affect the environment and human lives [37, 68, 77]. This emphasizes the need for monitoring solutions tailored to the special characteristics of the IoT environment in order to benefit the most from what the IoT devices have to offer [44, 72, 102] and to detect possible misbehavior [101] in the communication network [14]. Besides the individual privacy and security challenges [38] of IoT environments, additional challenges arise when these two requirements are put together. One of these challenges relates to the need to monitor devices that employ encryption in their communication as a confidentiality and privacy measure. In this case the monitoring solutions need to attune the security and the privacy requirements, and this becomes possible by harnessing the information extracted from IoT environments [7].

[2] https://krebsonsecurity.com/2016/10/source-code-for-iot-botnet-mirai-released/

The work presented in this thesis is motivated by:

- The privacy concerns raised by the large quantities of data collected in different Internet of Things environments.
- The need and the benefit of employing these data in practical applications while preserving the privacy of the data producers.
- The need to balance the privacy of the data producers with monitoring the behavior of the devices that collect these data.

We focus on a specific Internet of Things environment, a section of the electrical grid called the *Advanced Metering Infrastructure (AMI)*, which exhibits the aforementioned privacy and security challenges. We formulate research questions, identify problems and propose and evaluate solutions for the AMI. Due to the similarities in the data producing process, *monitoring equipment* and *privacy related* solutions for the Advanced Metering Infrastructure are many times applicable to other IoT environments.

The rest of this introductory chapter is structured as follows: Section 1.2 describes general aspects of the Advanced Metering Infrastructure, outlines privacy issues of large scale AMI data collection together with challenges of employing them in practical AMI applications and also the requirement of balancing the need for privacy with the need to monitor the behavior of the AMI devices. In Section 1.3 we formulate our research questions, Section 1.4 contains an overview of our methodology and we present the main contributions of this thesis in Section 1.5. Section 1.6 contains the summary of the appended publications, and we present our conclusions, followed by future research directions in Section 1.7. Parts II, III and IV contain the research articles which present in detail the contributions and results of this thesis.

1.2 Background, challenges and related work

1.2.1 The smart electrical grid and the Advanced Metering Infrastructure (AMI) - the challenges of large data

With the deployment of the Internet of Things, the *electrical grid* is transitioning to the so-called *smart [electrical] grid* [97]. The European Commission Directorate General for Research defines smart grids as "electricity networks that can intelligently integrate the behavior and actions of all users connected to it – generators, consumers and those that do both – in order to efficiently deliver sustainable, economic and secure electricity supplies" [97]. The objective of the smart grid is to enhance the classical electric network through IoT devices installed in key locations.

The European Commission portal[3] offers an overview of the undergoing smart grid projects in the European Union.

The electrical grid is divided in three main sections: the generation section (where electricity is produced in power plants), the transmission section (electricity is transported over high-voltage lines) and the distribution section (electricity is delivered to the consumers). The first two sections already benefited from automated monitoring and control systems, while the last one is currently undergoing the upgrade process. The generation and transmission sections are monitored and managed by the *Supervisory Control and Data Aquisition (SCADA)* system, a type of industrial control system. The concept of SCADA predates the smart grid one, but it is continuously upgraded with the help of new IoT devices and by transitioning from the traditional legacy systems to new ones based on commercial equipment and operating systems [53].

In the distribution section of the electrical grid, the classical electrical energy meters are replaced with new ones, called *smart meters* [20]. A smart meter (SM) provides two-way communication with the central system allowing efficient monitoring and control of the electricity delivery process. Smart meters facilitate the data collection process which provides the system operator with an important source of information. These devices, together with the communication network connecting them, form the *Advanced Metering Infrastructure* (AMI) [13]. The AMI concept is relatively new and it brings together elements from electrical engineering and information technology. While in some countries the upgrade of the distribution network is almost complete, in others the AMI deployments are in different stages, depending on the local rules, regulations and technical implementations [94]. As a component of the electrical network, the Advanced Metering Infrastructure becomes a critical asset of out society. Studying its properties and particularities, finding solutions for its inherent problems and developing tools for the AMI becomes extremely important from both a research and a practical perspective.

Large datasets - multiple possibilities: Similar to the Internet of Things, the prognosis is that data produced by smart grid equipment will be considerable and the size of the smart grid may become larger than the size of the current Internet.[4] These data will play a key role in the development of the smart grid, and analyzing and building applications on these data will contribute towards improving electrical grid stability. Data collected in the Advanced Metering Infrastructure will provide, among others, better management of the electrical energy consumption and the integration of renewable energy sources [90]. Some of these improvements are closely related to other developing IoT areas. For example, these data will facilitate a close interaction with the vehicular domain, as electrical vehicles will have an active role in the electrical grid as energy sources during peak periods [24, 91].

[3] `http://ses.jrc.ec.europa.eu/sites/ses.jrc.ec.europa.eu/files/u24/2014/project_maps_28_april_2014.html`

[4] `http://news.cnet.com/8301-11128_3-10241102-54.html`

As previously mentioned, the Advanced Metering Infrastructure is composed of smart meters, i.e. devices that have two-way communication capabilities with a central system. Smart meters are capable of providing fine-grained data regarding the electrical energy consumed at the household level and the quality of electrical power delivered [13]. The electrical energy consumed by the household is measured using the kWh unit of energy. The billing is usually done over long periods of time (i.e. 1-3 months), but the fine-grained energy consumption data can also be used for differential tariffs or even for customer re-imbursements when renewable energy is produced locally [90]. Information regarding instantaneous values of voltage, current, active and reactive power are used for grid operation purposes and they can be very useful for a *low-voltage SCADA* system [83], especially when managing local renewable energy sources. Grid operational data need to be collected very often (i.e. less than a minute) in order to give an accurate overview of the electrical distribution network. Efthymiou and Kalogridis [30] use the term *high-frequency (HF) data* for data which are used mainly for grid operational purposes and *low-frequency (LF) data* for data used mainly for billing purposes. We will keep the same definitions for these two types of data throughout this thesis.

The work in this thesis is focused on data produced and collected in the Advanced Metering Infrastructure. Information from AMI data can be harnessed with the help of cloud processing [66, 88], and after a thorough data validation process [46], will allow for the development of a number of AMI applications such as peak energy consumption shaving [40], short-term energy consumption forecast [92], prevention of energy-related fraud [16, 70], securing critical infrastructures [6, 44, 45] and also educating the consumers towards efficient energy usage [71].

1.2.2 Large datasets - privacy issues, solutions and applications

As mentioned in Section 1.1, large quantities of data collected in the Internet of Things can raise privacy concerns, especially when a person is behind the data production process. This also applies to the Advanced Metering Infrastructure environment where energy consumption data can be used to infer information about the lifestyle of people living at the premises. Peoples' privacy can be preserved with the help of *privacy enhancing technologies (PET:s)* tailored to the AMI environment.

Data recorded and reported by the smart meters can contain sensitive information such as electrical equipment usage patterns [78], presence or absence from the premises [69] or even the channel displayed on the TV set [43]. With the help of a technique called Nonintrusive Load Monitoring [48] the type of the electrical appliances installed at premises may be inferred and this process is simplified by AMI data collected with high granularity [31]. The energy load profiles become distinctive biometric behavioral traits [8] and can be used to identify individuals or group of persons based on their energy consumption patterns.

Privacy-enhancing technologies for the Advanced Metering Infrastructure

The examples above show some of the privacy concerns raised by data collected in the Advanced Metering Infrastructure. Thus, preserving data producers' privacy is an important goal and employing PET:s in the AMI environment can help to achieve it. Before presenting the main PET:s for AMI we briefly mention the legal framework covering the smart metering data. At the time of writing, there is no enforced European Directive that covers smart metering data in particular and this type of data falls under the general incidence of the EU Data Protection Directive 95/46/EC [25]. The EU Data Protection Directive 95/46/EC is being replaced by the EU General Data Protection Regulation (GDPR) [35], whose enforcement for all EU entities will take place on 25^{th} of May 2018. As a consequence, until the enforcement of GDPR, deployment of the Advanced Metering Infrastructure can be significantly slowed down in countries where privacy preservation of AMI data is not guaranteed by law [23]. Article 35 of the GDPR which refers to Privacy by Design will become the legal requirement for implementing Privacy Enhancing Technologies in all domains, including the process of developing and improving the smart grid. The EU Commission also provides a Data Protection Impact Assessment Template for Smart Grid and Smart Metering Systems [34] whose scope is "[...] to help ensure the fundamental rights to protection of personal data and to privacy in the deployment of smart grid applications and systems and smart metering roll-out [...]" [34]. In addition, the EU Smart Grid Task Force[5] offers documentation containing guidance on data protection and privacy for smart grid investors and data controllers[6].

There are a number of surveys that cover the privacy enhancing technologies proposed for the Advanced Metering Infrastructure [10, 56, 87]. Generally, there are two main types of PET:s for AMI: techniques that operate on the personally identifiable information (PII) attached to the energy data (i.e. data anonymization, data pseudonymization) and techniques that operate on the energy consumption data (i.e. data aggregation, data obfuscation, verifiable computation). Combinations of these technologies are also possible for additional enhancing of data' privacy.

Data anonymization and usage of pseudonyms: We return to the two types of AMI data presented earlier: high-frequency (HF) data used for grid operational purposes and low-frequency (LF) data used for billing. Efthymiou and Kalogridis [30] consider billing data to be privacy neutral, as they are seldom collected (LF data), thus showing overall information about the energy consumption process over the time period considered. Also, for a correct billing, these data need to be attributable to a specific customer. On the other hand, grid operational data need to be collected often (HF data), thus they might show detailed information about the cus-

[5] http://ec.europa.eu/energy/en/topics/markets-and-consumers/smart-grids-and-meters

[6] https://ec.europa.eu/energy/sites/ener/files/documents/2014_dpia_smart_grids_forces.pdf

tomer's lifestyle. For privacy reasons, the connection to the real identity of these data needs to be severed. This can be accomplished with the help of third party entities which are responsible of collecting, anonymizing and then delivering the sensitive data to its beneficiary [9, 30, 98]. Borges et al. [11] propose a similar solution based on anonymity networks where a customer uses different identities for transmitting the billing and grid-operational data. Rottondi et al. [84] propose a pseudonymization protocol which relies on a secret sharing scheme, enabling a set of nodes in the network to perform pseudonymization without having access to the measurements themselves. The protocol also provides a Identity Recovery phase which can be performed in case of alarms or faults, which can be used to connect a pseudonym with the real identity of the data producer.

Data aggregation and obfuscation: These methods can employ simple aggregation, when the values are aggregated together, or data can be obfuscated with the help of noise addition. Homomorphic cryptography methods can be employed to provide an extra layer of privacy for the aggregated values [63, 69, 99] and they can also be combined with noisy aggregation [63]. Bohli et al. [9] propose a solution where each meter adds a random value extracted from a known distribution to each of its reported consumption values. If enough smart meters participate, then the energy provider which knows the parameters of the distribution can compute the (approximated) aggregated consumption. Methods based on differential privacy [29] where noise extracted from a Laplacian distribution is added to the aggregated consumption were also proposed for the Advanced Metering Infrastructure environment. Barthe et al. [3] propose a protocol which is able to aggregate smart meter readings into statistics and bills in a privacy-friendly fashion. Ács and Castelluccia [1] present a similar solution, relying on a different distributed noise generation method. Shi et al. [86] propose a solution where homomorphic encryption is integrated with differential privacy to prevent untrusted aggregators from gaining knowledge from the participant's data. Rottondi et al. [85] propose a technique inspired from differential privacy and multiparty computation which is based on white noise addition. There are also methods that provide obfuscation via technical means [59, 60], but they are outside the scope of this thesis.

Limitations of privacy-enhancing technologies for the Advanced Metering Infrastructure

Recent studies showed that through adversarial means, the effect of privacy-enhancing technologies for the Advanced Metering Infrastructure can be diminished. Jawurek et al. [55] present the problem of breaking smart meter privacy by employing a technique called de-pseudonymization. They rely on support vector machines and present procedures that can be employed to link consumption traces by correlating anomalies that happen at the same time (such as consumption spikes or blackouts) or by finding similar customer behaviors in different consumption traces.

Buchmann et al. [15] show that with the help of simple statistical tools such as mean and standard deviation, individual houses can be identified based on their energy consumption records. Eibl and Engel [31] describe how a characteristic of AMI data, called *data granularity* can influence the efficiency of detection methods employed in nonintrusive load monitoring algorithms [48], thus diminishing the effect of privacy-enhancing methods that hide appliance consumption patterns. Faisal et al. [36] show how data granularity and the quantity of collected data affects the re-identification efficiency. Their results show that even very low sampling consumption traces (two samples per year) can still be viable to re-identify customers with an accuracy of 20%.

Studying the characteristics of the AMI data and especially how they influence the efficiency of the privacy-preserving methods will help the improvement of these methods, the efficient collection of AMI data and also the development of privacy-preserving smart grid services. We further describe this problem and our research contribution in Research Question 1 (RQ1).

Enhancing the utility of privacy-enhanced data in practical applications

Aggregation and obfuscation solutions are successful in preserving the privacy of the customers involved, but depending on how they are performed, some important information might be lost, thus narrowing the applications where resulting data can be used. In Section 1.2.2 we presented some of the applications where data produced in the Advanced Metering Infrastructure can be employed. Some of these applications require data which is unaltered and identifiable with the real identity of the customer who produced it, while others can employ data that undergoes privacy-enhancing processing. Due to privacy issues many applications are enhanced and become privacy-preserving. Applications which employ obfuscated Advanced Metering Infrastructure data may suffer a loss in accuracy caused by the extra noise introduced by the privacy-enhancing methods employed. Limiting the accuracy loss while maintaining the customers' privacy is an interesting investigation venue.

Erkin and Veugen [32] indicate that AMI data collected for management purposes can also be used by third parties and propose solutions to provide new personalized services for smart homes while protecting the privacy-sensitive data. Gong et al. [42] propose a privacy-preserving scheme for demand-response programs which also enables rewarding customers for reducing their load during demand peaks. In the following we focus on obfuscation methods based on differential privacy [29]. Ács and Castelluccia [1] propose an aggregation method based on differential privacy which can hinder the adversary's ability to infer customers' activity during a specific period of time. Their results show that the application's utility increases together with the number of participating customers. Barthe et al. [3] propose and evaluate a distributed solution that can be used to aggregate smart meter readings into statistics and bills, but no experiments based on real data are provided. Jelasity and Birman [57] assume the existence of a

bound in order to limit the global sensitivity of the aggregation and discuss a method to prevent differentially private queries run continuously over time from enabling the adversary to learn the readings' underlying distribution. Bounding the global sensitivity may reduce the quantity of noise added, which will improve the utility of aggregation. We further describe the utility enhancing challenge and our research contribution in Research Question 2 (RQ2).

1.2.3 Challenges in monitoring critical infrastructures

In addition to their benefits, Internet of Things devices also raise cyber security challenges. By exploiting their vulnerabilities, IoT devices can become part of a botnet and used in large-scale network attacks.[7] This becomes even more dangerous when these devices are part of critical infrastructures which control physical processes. Equipment malfunction and malicious activities need to be detected early by monitoring IoT devices and their communication network.

In recent years, the integration of commercial off-the-shelf solutions in critical infrastructures enabled cyber-attacks which can be carried out in the same way as in classical IT systems [72]. One of the attack techniques that were recently employed was that of malware tailored for these types of systems. Stuxnet [79], Duqu [4] and Flame (sKyWIper) [64] are relevant examples of ICS related malware discovered recently and their complexity shows the amount of resources and effort that was put in their development. These pieces of malware, together with a recent proof-of-concept called Irongate [52] show that some adversaries are willing to allocate significantly more resources compared with many examples of malware for classical systems [33]. Two recent attacks, one against a power station[8] and one against a public heating system[9] give a glimpse of the effects of these attacks on critical infrastructures.

The research community is actively working towards identifying and mitigating vulnerabilities that affect critical infrastructures such as the Advanced Metering Infrastructure (AMI). Carpenter et al. [18] present a number of vulnerabilities that exist in devices in the AMI together with an attack methodology. Subsequently, Foreman and Gurugubelli [39] present the attack surface of the AMI with respect to hardware and network configurations, protocols, and software. McLaughlin et al. [73] and Grochocki et al. [44] describe possible attack scenarios for the Advanced Metering Infrastructure, starting from potential attacker goals covering denial of service, energy fraud, and even targeted disconnect of electrical services. Distributed denial of service (DDoS) against a power station [77] might affect the control equipment and the energy supply in the area served by the targeted power station. Smart meters can be tampered with to report a lower energy consumption in order to lower the electricity bill, thus committing fraud. A fraud

[7] https://www.wired.com/2016/12/botnet-broke-internet-isnt-going-away/

[8] https://www.wired.com/2016/03/inside-cunning-unprecedented-hack-ukraines-power-grid/

[9] https://thehackernews.com/2016/11/heating-system-hacked.html

case involving a large number of smart meters, which may have cost the utility several hundreds of millions of dollars was reported in Puerto Rico [62]. A similar case occurred in Malta [68] where the authorities discovered that at least $1,000$ smart meters had been tampered with, affecting 10% of the total local generation of electricity and causing a loss of approximately $41 million in 2012 alone. Abusing the smart meters' remote disconnect feature [22] may influence the power quality, leading to negative effects on the devices connected to the electrical network.

These reported events and also the potential attacks described by the research community reveal the need for monitoring solutions for the Advanced Metering Infrastructure. The monitoring is performed by a specialized system, called *Intrusion Detection System (IDS)* tailored to the environment under observation. IDS:s [89] have three logical components: the sensors (data collectors), the analyzers (data processors) and the user interface (presenting the information to the operator). Depending on their location and on the type of data analyzed, the IDS:s can be host-based (monitoring the events in a single host), network-based (monitoring the network traffic) or distributed/hybrid (mix between host-based and network based).

Due to the large scale of the AMI environment, a distributed IDS is apparently the most cost-effective in a long-lived deployment, according to Cárdenas et al. [17]. Zhang et al. [103] propose a distributed architecture with multiple IDS nodes deployed at different points in the AMI, while Grochocki et al. [44] recommend a distributed IDS model that can monitor traffic between peers and also scales with the size of the AMI communication network. Besides monitoring network traffic, there is also a need to monitor devices' internals, as some of the attacks' manifestations may not reflect into exterior traffic. Raciti and Nadjm-Tehrani [82] present a model for a host-based IDS which detects anomalies inside AMI devices. They build a module based on this model and test four possible attack types against the smart meters' internals: data manipulation, recalibration (changing registers' values), reset (deleting the records regarding consumed energy) and sleep mode (the meter is put into sleep mode and the energy consumed is not registered). If not detected in time, these attacks can cause serious economic loss.

Depending on the detection method employed, IDS:s can be signature based (detect attacks by comparing current activity with specific patterns) or anomaly based (compare current activity with a model of trustworthy behavior and alert for deviations) [89]. Due to the lack of known attack signatures, Mitchell and Chen [75] stress that a behavior (anomaly) based IDS is preferred in the Advanced Metering Infrastructure environment over a signature-based one. Behavior models can be built with the help of large data collected in the AMI environment. As mentioned in Section 1.2.2, AMI data undergoes privacy-enhancing processing in order to protect them against privacy invasive attacks. Therefore, network traffic between AMI devices is usually encrypted to protect the confidentiality of the communication process and customers' data privacy. One important source of information for behavior models comes from monitoring the traffic between devices [51] but this process might become cumbersome, especially when

the AMI traffic is encrypted [7]. Monitoring the AMI communication network while preserving confidentiality and privacy becomes a major challenge in the process of developing security solutions for the Advanced Metering Infrastructure environment. This shows that there is a need to develop monitoring solutions that are also privacy preserving and the starting point for these monitoring tools relies in the study of the large data produced in this environment. We further describe this problem and our contribution in Research Question 3 (RQ3).

1.3 Research questions

The research presented in this thesis is based on the analysis of large quantities of data produced in large scale environments such as large scale cyber-physical systems or Internet of Things and we pose research questions which stem from the challenges presented in Section 1.2. Our study has three primary objectives, filling some of the identified gaps and advancing the current state of research. The first one is to identify relevant characteristics of the collected datasets and study their influence on previously proposed anonymization techniques [15, 31, 36, 55]. The second objective is to investigate how to enhance the utility of applications which employ data that undergo privacy-enhancing technologies based on noise addition [1, 3, 57]. The third is orthogonal to the first two and focuses on the balance between privacy requirements and the requirement to monitor the communication network [7, 17, 75]. Next, we define and motivate our research questions, while in the following sections we present the methodology employed and the contributions of this thesis.

RQ1: Which dataset characteristics influence the efficiency of privacy-enhancing technologies and what is their effect?

RQ2: How to enhance the utility of data that undergo privacy-enhancing technologies based on addition of noise?

RQ3: How to balance the need for confidentiality and customers' data privacy with the need to monitor the communication network?

RQ1: Data gathered in different environments can be used to infer sensitive information regarding the individuals that are behind the data producing process, raising privacy concerns. In order to alleviate these concerns, privacy-enhancing technologies can be applied during the data collection process [30]. However the efficiency of these technologies may be influenced by properties of the data itself. Recall from Section 1.2.2 that in the Advanced Metering Infrastructure, information about customers' lifestyle can be inferred from fine-grained energy data collected by smart meters [69]. We identify and investigate AMI datasets' characteristics [36] and how they influence privacy-enhancing techniques which were previously proposed for this environment.

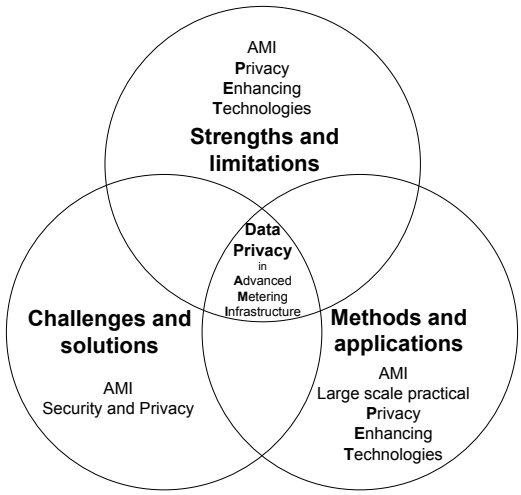

Figure 1.1: The research directions applied to the Advanced Metering Infrastructure environment

RQ2: Large data produced in cyber-physical systems offer many advantages but they also raise significant challenges regarding private data processing and enhancing privacy oblivious applications. One technique that can be used to increase the resilience of these large data sets against privacy violations such as de-anonymization and de-pseudonymization is to aggregate individual data values with the help of privacy-enhancing technologies based on noise injection [1]. Recall from Section 1.2.2 that the quantity of injected noise can affect the utility of these data, and might decrease the efficiency of the cyber-physical systems' applications employing them [57]. Fine-tuning the noise addition process may improve data utility, but this may come at a privacy cost. We investigate ways of enhancing the utility of data that undergo privacy-enhancing technologies based on addition of noise.

RQ3: Monitoring solutions are required in order to ensure the correct behavior of CPS devices. These monitoring solutions might often interfere with the privacy requirements by inspecting sensitive data, hence they need to be adjusted accordingly. Please recall from Section 1.2.3 that in order to ensure confidentiality, devices in the Advanced Metering Infrastructure rely on encryption in their network communication [5]. This makes the communication network more resilient against malicious adversaries but at the same time reduces the monitoring capabilities of the network operator, making it more difficult to detect any misbehaving user or equipment. We investigate how to harmonize the privacy and confidentiality requirements with the devices' monitoring requirements in the AMI environment [7]. This investigation is important in the context of developing intrusion detection solutions for the Advanced Metering Infrastructure [6].

Figure 1.1 depicts the research directions applied to the Advanced Metering Infrastructure environment. In the following we present each research question in the context of the Advanced Metering Infrastructure, where we describe the methodology employed and our contributions

1.4 Research methodology

In this section we present an overview of the methodology used in this thesis with regard to the research questions formulated in Section 1.3. A detailed description of these methods and their specific implementation is provided in Parts II, III and IV. Some of the methods employed are shared (with small adaptations) between the different research questions, while others are employed in close connection with each of the questions under study.

In our work we focus on the large quantities of data collected in the Advanced Metering Infrastructure. In order to explore the possible answers to our research questions, we employ methods that extract, analyze and process the useful information from the aforementioned data.

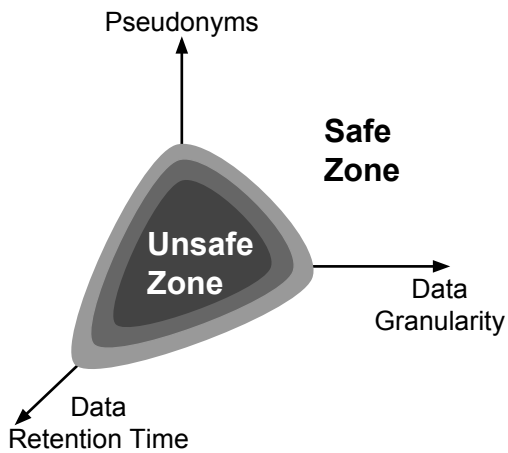

Figure 1.2: Advanced Metering Infrastructure datasets' characteristics

RQ1: Our first research question addresses the influence of dataset's characteristics on the efficiency of privacy-enhancing technologies [36]. Figure 1.2 (proposed by us in Paper I) depicts three main characteristics we focus on: the usage of *pseudonyms* in the process of reporting/storing data for the same customer, the *retention time* of data stored by the utility provider under the same pseudonym for the same customer and the *granularity* [36] of reported/stored data. We briefly describe them and motivate their influence on the data privacy:

Pseudonyms: Using a pseudonym instead of the real identity is the first step in the data anonymization process. This offers a privacy-preserving layer but the connection between the real identity of the customer and the pseudonym needs to be kept secret. This protection can be enhanced by increasing the number of pseudonyms used for each customer and changing them frequently, thus shortening the period where each pseudonym is used, offering a better separation between the real identity and the pseudonyms used.

Data retention time: A long period of stored data will help create an accurate energy consumption profile, which can be used to extract sensitive information about the data producer's lifestyle. Shortening the data retention time or employing multiple pseudonyms over a long retention period may help in enhancing the privacy of the participating customers, increasing the difficulty of tracking them through the consumption pattern.

Data granularity: AMI data can be reported and stored under different granularities depending on their application utilization. Data used for billing needs to be reported using the exact consumption values for accuracy and to prevent fraud, but it can be collected with a low frequency, recording only the total quantity of energy consumed in a time period. High-frequency data needs to be collected more often as some grid operation applications require a short response time and fine-grained data. Decreasing the granularity of the data may help in improving customers' privacy, by hiding consumption artifacts.

We study the influence of these three characteristics on two previously proposed privacy violation methods: de-anonymization [30] and de-pseudonymization [55]. We propose algorithms describing a two-staged adversarial model which targets Advanced Metering Infrastructure datasets with the purpose of extracting sensitive information in relation to the data producer. In each of the stages, the adversarial model employs one privacy violation method: de-anonymization in the first stage and de-pseudonymization in the second one. For the de-anonymization stage, we propose a framework based on probabilistic analysis (bins and balls) [76] which assumes a specific distribution (Poisson) of the customers' energy consumption values. Based on the uniqueness of these values and using the probabilistic tools available for the bins and balls problem [76] we estimate the expected number of customers that can be de-anonymized by employing the adversarial strategy. For the de-pseudonymization stage, we estimate the expected number of customers that can be identified by randomly matching the pseudonyms use for storing customers' data. In addition to the probabilistic frameworks we perform a practical experimental evaluation of the adversarial capabilities with the help of a dataset consisting of smart meter readings from a large number of consumers in a medium-sized city. The data has undergone a sanitization process where a number of data collection artifacts has been removed, such as gaps in reporting, double conflicting records or decreasing consumption indexes.

RQ2: The second research question focuses on privacy-enhancing technologies which add noise to sensitive values in order to protect them against privacy invasive adversaries. We fo-

cus on differential privacy [28, 29], a state of the art privacy-enhancing noise addition method which offers probabilistic guarantees on the privacy leaks of individual values when a statistical result is released. Following the original definition of differentially private [29] and building on previously proposed solutions for the AMI environment [1], the noise drawn from a Laplacian distribution is proportional to the sensitivity of the function (given by the highest possible contributing value) employed in the release mechanism. In the case of differentially private aggregation of real values, the quantity of noise can be very large if the sensitivity is not bounded, which reduces the utility of the aggregation result [57]. We analyze the bounding mechanism, we propose applied practical methods which can be used to compute bounds and we investigate their impact on the privacy of AMI customers. In addition to the bounded sensitivity we evaluate how other data collection parameters can be fine-tuned in order to maximize data utility, by minimizing the mean absolute percentage error [1] between the differentially private and the simple aggregation.

We build on the adversarial model employed in RQ1 [95] and we investigate the complementary protection offered by differentially private aggregation when combined with other privacy-enhancing technologies, such as anonymization. We run an evaluation on a fine-grained AMI energy consumption dataset in order to evaluate how differentially private bounded aggregation reduces the adversarial efficiency in performing de-anonymization attacks and what is the effect on the utility of the aggregated statistic. Furthermore, we study how data that undergo privacy-enhancing technologies based on noise injection (differential privacy) can be employed in practical Advanced Metering Infrastructure applications such as short-term load forecasting [54]. We propose a framework that can be used to enhance an AMI application with the help of differentially private aggregated data and to evaluate the effect of this enhancement on the application's efficiency [96].

RQ3: Our third research question addresses the capacity of monitoring solutions for the Advanced Metering Infrastructure communication network to handle encrypted communication without affecting the data privacy. Employing encryption makes the communication network more resilient against malicious adversaries but at the same time reduces the monitoring capabilities of the network operator, making it more difficult to detect any misbehaving user or equipment [7]. We start from a previously proposed command recognition methodology [51] and we refine and adapt it to the AMI environment. We analyze the properties of the AMI communication network and we identify features [65] that characterize the communication protocols used. In the feature selection process we analyze and motivate the relevance of each individual feature for the problem considered. The selected features reflect aspects related to commands' timing and duration, making them applicable also to other IoT protocols. We employ supervised learning and we propose an AMI command classifier based on k-nearest-neighbor algorithm, a fast and accurate solution for the problem at hand [100]. We evaluate the efficiency of our

classifier on two Advanced Metering Infrastructure protocols currently used in EU AMI deployments. One of the protocols, called DLMS/COSEM [26], uses encrypted communication while the other, called M-BUS [67], is difficult to parse due to its proprietary implementation. In our evaluation we assess the efficiency of both individual and combined features, while for the combination we also apply dimensionality reduction [58] and analyze its effect on the classification method. In our analysis we also cover the case of commands that can only be differentiated based on their payload, which circumvents the privacy requirement of our solution and we investigate solutions for this special case taking into account the commands' impact on security.

1.5 Thesis contributions

In this section we present an overview of the contributions of this thesis, with regard to the research questions formulated in Section 1.3. We start by enumerating the main contributions, followed by a short summary. A full description of these contributions is provided in Parts II, III and IV.

Main contributions and advancement on existing research:

RQ1: Advanced Metering Infrastructure data characteristics and anonymization efficiency

- We study the effects of AMI dataset characteristics (data granularity [36], retention time, usage of pseudonyms) on two privacy violation methods: de-anonymization [30] and de-pseudonymization [55].
- We define and refine an adversary model [15, 55] and present her methodology which covers both privacy violation methods in order to assess her limitations.
- We demonstrate both through probabilistic estimation and evaluation on a real dataset how small changes in the data collection process [36] can scale down the efficiency of the adversary.

RQ2: Efficient applications of differential privacy in AMI data processing

- We provide a method that can maximize the utility [57] of a differentially private statistic by controlling its aggregation parameters, allowing for differential privacy to be practically [1] deployed in similar cyber-physical systems such as AMI
- We provide a thorough evaluation, based on a real prototype [3] and conducted with events collected from a real-world Advanced Metering Infrastructure, showing the accuracy [57] of differentially private aggregation based on our proposed method.

- We present a methodology that can be employed to enhance an existing application with data that is processed with privacy-enhancing technologies based on noise addition [1] and qualitatively and quantitatively evaluate the effect of this enhancement on the application's utility.
- We apply this methodology to an AMI short-term load forecasting application [54] and we demonstrate that the error introduced by the noisy aggregated data does not have a major effect on the accuracy of the forecast algorithms in question.

RQ3: Monitoring Advanced Metering Infrastructure devices by analyzing encrypted traffic

- We perform an analysis of the AMI communication network and its properties in order to identify features important from a security perspective [51].
- Based on these features, we propose a methodology to identify the type of commands [51] exchanged between AMI devices, which can handle proprietary and/or encrypted AMI protocols [7].
- We provide a validation of our approach using collected traffic [7, 51] from two testbeds using different AMI protocols currently used in EU AMI deployments.

Summary of contributions:

RQ1: We focus on two types of data that can be collected in the AMI. The first type is called Low-Frequency (LF) data [30] and is seldom (sometimes yearly) collected and has billing as a primary application utility. The second type is called High-Frequency (HF) data [30], which is collected very often, and is mainly used in grid operation applications. LF data is somewhat privacy neutral and, because of the legal implications of the billing process, needs to be identifiable with the data producer's real identity. On the other hand, fine-grained HF data raises privacy concerns, thus needs to be collected and stored under a pseudonym in order to preserve the privacy of the customer that produced the data.

We identify, couple together and investigate the three data characteristics presented in Section 1.4: the usage of pseudonyms, the retention time and the granularity. In adition, we build upon two previously proposed privacy-enhancing methods that operate on the personal identifiable information (PII) stored in AMI datasets: anonymization [30] and usage of pseudonyms [84]. We advance the current state of research by defining an *adversarial model* comprising of two stages, each covering a different privacy violation: de-anonymization and de-pseudonymization.

In addition to existing literature covering these privacy violations [15, 55], we propose a probabilistic framework to better define the adversarial capabilities and methods that allow evaluation on large AMI datasets. In the de-pseudonymization stage HF datasets originating from the same customer but stored under different pseudonyms are linked together using a method

based on distances in a multi-dimensional feature space which employs features extracted from the energy consumption data, broadening the work presented in [15]. In the de-anonymization stage customers' HF and LF datasets are linked using a probabilistic method [76] based on the uniqueness of LF energy consumption values. We employ this framework and we evaluate how changes in the data collection procedure [36] which modify the datasets' characteristics can help mitigate the outcome of privacy violations such as de-anonymization and de-pseudonymization. Our findings show how tuning the aforementioned characteristics of the data collection process can reduce the efficiency of the adversary. Our proposed methodology and results can be used by data custodians to better understand the properties of their Advanced Metering Infrastructure datasets and provide a foundation for developing privacy-preserving release methods.

RQ2: We focus on differential privacy [28, 29], a method which offers probabilistic guarantees on the privacy leaks of individual values when a statistical result is released and we extend the existing research on employing differential privacy in practical AMI applications [1, 3]. We present practical solutions to one problem previously raised in the literature, that of enhancing the utility of the data by bounding the sensitivity of the release mechanism [57]. We show the strong complementary protection offered by differentially private aggregation when combined with other privacy-enhancing technologies, such as anonymization [30]. Our results show that differentially private bounded aggregation reduces the adversarial efficiency in performing de-anonymization attacks while enhancing the utility of the aggregated statistic compared with unbounded differentially private aggregation [95]. Our solution, based on a state-of-the-art stream processing engine, can be efficiently deployed in environments similar to existing AMI environments.

We contribute to the process of developing practical privacy-preserving applications for the AMI environment [32] and we study how data that undergo privacy-enhancing technologies based on noise injection can be employed in practical Advanced Metering Infrastructure applications. We propose a methodology that can be used to enhance an AMI application with the help of differentially private aggregated data and to evaluate the effect of this enhancement on the application's efficiency. We use this methodology on an application that relies on fine-grained AMI energy consumption data in order to compute accurate predictions of short term energy consumption, building on and expanding the work presented in [54]. We identify different information sources that can be employed by this application and make an analysis of their characteristics with respect to the trade-off between application's accuracy and their privacy impact. We identify privacy neutral sources of information which enhance the accuracy and we use them in the prediction process. We provide a quantitative evaluation in which we compare the accuracy of a differentially private enhanced prediction application with its non-enhanced counterpart. Our results show that there is a minor trade-off between the differentially private aggregation privacy benefits and the loss in accuracy.

RQ3: So far our contributions covered privacy issues and applications of privacy-enhanced data for cyber-physical systems. Orthogonal to this, we focus on and advance the current state of research concerning monitoring solutions for the CPS environment which need to handle encrypted traffic [7]. Building on the work on encrypted industrial control traffic [51], we provide solutions in employing information extracted from encrypted AMI traffic, without affecting the privacy of customers.

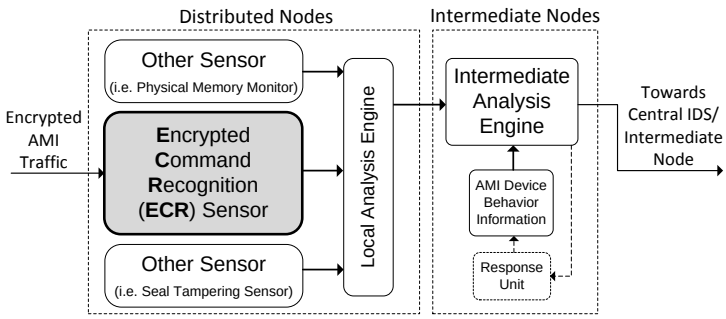

Figure 1.3: An Encrypted Command Recognition Sensor as an AMI IDS module

We develop one essential component for an AMI Intrusion Detection System (IDS) [82], which we call the *Encrypted Command Recognition (ECR)* sensor (depicted in Figure 1.3 and proposed by us in Paper IV). The ECR sensor can accurately determine individual AMI commands exchanged between AMI devices, commands relying either on an encrypted protocol or one that is difficult to parse. The command identification is performed in a privacy-preserving fashion, without decrypting the traffic and without accessing the sensitive customer data reported via the AMI communication network. The command identification is performed with the help of a classifier which employs features based on side-channel information [51] and can be applied to both encrypted or hard to parse protocols. We also cover the special case of commands that can only be differentiated based on their payload contents and we offer solutions for this special case based on the commands' impact on security.

Our main contribution to already existing research is the ECR module, which can become an important component of a distributed Intrusion Detection System for the Advanced Metering Infrastructure environment [103]. It will help in early detection of misbehaviors and attacks [7], giving the network operator a better overview on the AMI communication network's status, while preserving the privacy of customers' sensitive data [69]. We expect that the features identified for the two protocols we have studied will facilitate the study of other proprietary or encrypted protocols employed in the IoT, while preserving the privacy of the exchanged data.

1.6 Summary of appended research articles

1.6.1 Paper I: The Influence of Dataset Characteristics on Privacy Preserving Methods in Advanced Metering Infrastructure

In the first paper we investigate **RQ1** in detail. We focus on previously proposed privacy-enhancing technologies and we study how their efficiency is affected by the characteristics of the datasets they are applied to. We present the first steps towards an analytical framework which models a privacy-invasive adversary, a framework that can be used to estimate the effect of the datasets' characteristics and the conditions under which such datasets can be released to third parties.

Our main focus is on datasets collected from the AMI environments and on two privacy violations they are susceptible to: de-anonymization and de-pseudonymization. We start from formalizing the process of creating low-frequency and high-frequency datasets which are employed in different Advanced Metering Infrastructure applications. Next we define an adversary model and present her methodology for each of the two privacy violation techniques to better understand her limitations. We propose a formalization for the two privacy violations based on a probabilistic framework and we compare the theoretical model estimation results with evaluation results obtained from a large dataset of energy consumption data captured in a live Advanced Metering Infrastructure environment. Our results show how Advanced Metering Infrastructure datasets' characteristics (data retention time, data collection granularity and the frequent changes of pseudonyms) can be tuned in order to mitigate the efficiency of an adversary that intends to perform de-anonymization and de-pseudonymization.

1.6.2 Paper II: BES: Differentially Private Event Aggregation for large-scale IoT-based Systems

In the second paper we investigate the first part of **RQ2** and we focus on enhancing the utility of previously proposed privacy-enhancing technologies for the Advanced Metering Infrastructure. We study how differential privacy, a privacy-enhancing technology offering probabilistic guarantees, can be used to complement other technologies to allow for controlled disclosure of statistics computed on sensitive data.

We propose methods to limit the noise introduced by differential privacy in real-world applications, by bounding the parameters of differential privacy based on information extracted in a differentially private fashion from the system under study or from other similar systems, thus keeping the differential privacy's guarantees. We provide an evaluation based on a fully implemented prototype using real-world data from the Advanced Metering Infrastructure. We

show how a large number of events can be aggregated in a private fashion with low processing latency by a single-board device, similar in performance to the devices deployed in the Advanced Metering Infrastructure. With the help of a previously published de-anonymization scenario (presented in Paper I), we also study the complementary protection offered by differentially private aggregation when compared to other privacy-enhancing technologies. We show that differentially private aggregation can reduce the efficiency of an adversary whose goal is to perform de-anonymization of Advanced Metering Infrastructure datasets.

1.6.3 Paper III: Employing Private Data in AMI Applications: Short Term Load Forecasting Using Differentially Private Aggregated Data

In the third paper we continue the investigation of **RQ1** and **RQ2** and we focus on the need to balance the privacy requirements with utility benefits in practical applications that rely on sensitive information. More specifically we study the possibility of employing in practical applications data that undergo privacy-enhancing technologies.

We propose a methodology which can be used to analyze and enhance an Advanced Metering Infrastructure application with data that was preprocessed with privacy-enhancing technologies. This methodology facilitates an analysis of privacy concerns raised by the different sources of information that could benefit the application. We apply this methodology to Short Term Load Forecasting, an Advanced Metering Infrastructure application that relies on sensitive AMI data to perform accurate prediction of electrical energy consumption. We conduct an exploratory study focused on the effects of differentially-private aggregation on linear Short Term Load Forecasting methods that can be employed in the Advanced Metering Infrastructure. We show that the noise introduced, in the case of a bounded sensitivity, has a minor effect on the forecast accuracy.

1.6.4 Paper IV: Harnessing the Unknown in Advanced Metering Infrastructure Traffic

In the fourth paper we investigate **RQ3** and we study how to balance the need for privacy and confidentiality with the need to monitor the communication between devices in the Advanced Metering Infrastructure environment. We develop a component for an Intrusion Detection System which can recognize the type of individual commands exchanged between Advanced Metering Infrastructure devices which might employ encrypted or hard to parse communication protocols.

We identify a number of features which summarize Advanced Metering Infrastructure traffic characteristics and we propose a traffic recognition methodology that employs them. We show, for two different Advanced Metering Infrastructure protocols, that our methodology and identified traffic characteristics can be employed in order to identify a set of commands exchanged between devices. This component can be used in an Intrusion Detection System and it is one important source of information for building and maintaining a behavior model for each and every device that the Advanced Metering Infrastructure is comprised of. The operator will get a better view of the network's status and early insights of possible attacks and misbehaviors.

1.7 Conclusion and future research directions

The proliferation of Internet of Things (IoT) with many small devices and their communication capabilities will produce large quantities of data which can be processed and transformed into valuable information, opening the path for new applications and improvement of the IoT environment. The work presented in this thesis tackles three data related challenges previously raised by the research community and focuses on one instantiation of an IoT environment, the Advanced Metering Infrastructure (AMI), a critical infrastructure. The first challenge is to study the characteristics of AMI datasets and their influence on the efficiency of privacy enhancing technologies (RQ1). The second one is to enhance the utility of applications which employ AMI data that undergo privacy-enhancing technologies based on noise addition and to better understand the effect of the noise added (RQ2). Orthogonal to the first two, the third challenge (RQ3) is to investigate the balance between data confidentiality and the requirement to monitor the AMI communication network, with practical applications. This thesis addresses these three challenges and proposes new methods for solving and analyzing the problems, as well as presenting extensive experimental evaluations in real usage scenarios.

We begin our study by focusing on RQ1 where we propose a framework to analyze the effect of data characteristics on the efficiency of two currently proposed privacy-preserving methods for the AMI: anonymization and usage of pseudonyms. Here, we identify and investigate three main characteristics of AMI data: the granularity of the data reported, its timespan and the number of pseudonyms used for reporting these data. Based on our results we provide practical means to tune these characteristics in order to enhance the efficiency and strengthen the resilience under adversarial hindrance of the two AMI data privacy-enhancing technologies under study. Our methodology can be employed by AMI data custodians to better understand the utility and the properties of their datasets and provides a stepping stone for developing and testing privacy-preserving release methods.

Without losing RQ1 from sight, we focus on RQ2 by investigating how to improve the utility of AMI data that undergoes privacy-enhancing technologies based on noise addition, such as differential privacy. Here we propose methods that can be used to tune the quantity of noise added in real-world AMI applications which employ differentially private aggregation. Based on our investigation, we demonstrate the complementary protection offered by differentially private aggregation when combined with other privacy-enhancing technologies, complementing our study on RQ1. In a particular example, we validate with the help of an adversarial scenario that the effect of de-anonymization can be mitigated if AMI data is aggregated in a differentially private fashion.

Moving onwards with RQ2, we propose a methodology which can be employed to analyze the different sources of information present in datasets which can benefit Advanced Metering Infrastructure applications. We focus on sources of information which can be privacy enhanced with the help of differentially private aggregation and we evaluate the effect of this enhancement on the applications' efficiency. We apply this methodology on an energy consumption prediction technique and our results show that, with proper tuning, the noise introduced by the differentially private aggregation has a negligible effect on the accuracy of the prediction. This opens the path for further possible applications' extension which enable extended use of data collected in AMI in a privacy-preserving fashion.

Orthogonal to RQ1 and RQ2, we continue with RQ3 and we investigate how to balance the need for confidentiality with the need to monitor the Advanced Metering Infrastructure communication network. We propose a methodology for an encrypted command recognition component, based on side-channel information, which can be used in an Intrusion Detection System for the AMI. Our results show that statistical information extracted from network traffic can be used to correctly identify AMI commands, even when they are sent over an encrypted channel or embedded in a protocol that is hard to parse. This is especially useful in environments where the network operator employs encrypted traffic, both for security reasons and for customers' privacy. Our proposed solution can successfully complement already existing monitoring techniques or it can be employed as a command recognition component in future developing ones.

Throughout this thesis, by harnessing the information contained in the large AMI data collected, we show how the security and the privacy of the entities can be enhanced, we pose and answer new questions, we validate previous findings, and we provide means to aid further extensions of this work. One important outcome of the exploratory work presented in this thesis is that it can complement the Article 35 regarding Privacy by Design, of the EU General Data Protection Regulation, for implementing Privacy Enhancing Technologies in the smart grid domain. Furthermore, our findings can be extended to other similar large scale Internet of Things deployments such as sensor and vehicular networks which share many common characteristics through the large data produced by similar IoT components.

Future research directions

The work presented in this thesis advances the current state of the research regarding privacy and security issues in the Advanced Metering Infrastructure. Based on the knowledge and technical resources available we have provided thorough answers to the research questions presented at the beginning of this chapter. Furthermore, our work stands as a starting point to further extend some of these answers while also identifying a number of new challenges. We briefly present these open challenges with respect to the research questions considered.

RQ1: In our first research question we have addressed the influence of Advanced Metering Infrastructure datasets' characteristics on the efficiency of privacy-enhancing technologies. One direction for future work is extending and improving the adversarial model in order to explore her full range of capabilities in performing de-anonymization and de-pseudonymization. Another research direction is to analytically bound the success rate of the adversary with respect to the characteristics of the targeted dataset. These will give a better understanding of the adversary while offering privacy-preserving options for releasing and processing AMI datasets.

RQ2: In our second research question we have focused on enhancing the utility of data that undergo privacy-enhancing technologies based on noise addition. Our study is focused on differential privacy and it can be extended by considering other distributions for the noise addition process and studying their effect on the utility of the data. Our study on short-term load forecasting applications employing noisy aggregated data can be extended to cover also other Advanced Metering Infrastructure applications in order to examine the benefits and possible limitations of privacy-enhancing based on noise addition. This will help in improving current and further developing applications based on AMI data.

RQ3: Finally, to answer the last research question we have investigated the possibility to balance the need for confidentiality and customers' data privacy with the need to monitor the communication network. In our study we have considered two protocols currently used in EU Advanced Metering Infrastructure deployments. Two limitations of our study are the classification method used and the number of commands considered. Our classifier is based on the k-nearest-neighbor algorithm, with a complexity linear in the size of the learning set. Other classification algorithms also need to be considered in order to obtain a good performance on the limited capabilities hardware installed in the AMI, especially in the case of protocols comprising a large set of commands. We have taken into consideration commands available in the current implementation of the studied protocols, and in the future our solution can be easily extended by considering larger datasets comprising of multiple commands.

Bibliography

[1] G. Ács and C. Castelluccia. I have a DREAM!: Differentially private smart metering. In *Proceedings of the 13th International Conference on Information Hiding*, IH'11, pages 118–132, New York, NY, USA, 2011. Springer-Verlag.

[2] L. Atzori, A. Iera, and G. Morabito. The internet of things: A survey. *Computer networks*, 54(15):2787–2805, 2010.

[3] G. Barthe, G. Danezis, B. Gregoire, C. Kunz, and S. Zanella-Beguelin. Verified computational differential privacy with applications to smart metering. In *Computer Security Foundations Symposium (CSF), 2013 IEEE 26th*, June 2013.

[4] B. Bencsáth, G. Pék, L. Buttyán, and M. Félegyházi. Duqu: Analysis, detection, and lessons learned. In *ACM European Workshop on System Security (EuroSec)*, volume 2012, 2012.

[5] R. Berthier, J. G. Jetcheva, D. Mashima, J. H. Huh, D. Grochocki, R. B. Bobba, A. A. Cárdenas, and W. H. Sanders. Reconciling security protection and monitoring requirements in advanced metering infrastructures. In *Smart Grid Communications (SmartGridComm), 2013 IEEE International Conference on*, pages 450–455. IEEE, 2013.

[6] R. Berthier and W. H. Sanders. Specification-based intrusion detection for advanced metering infrastructures. In *2011 IEEE 17th Pacific Rim International Symposium on Dependable Computing (PRDC)*, pages 184–193, December 2011.

[7] R. Berthier, D. I. Urbina, A. A. Cárdenas, M. Guerrero, U. Herberg, J. G. Jetcheva, D. Mashima, J. H. Huh, and R. B. Bobba. On the practicality of detecting anomalies with encrypted traffic in AMI. In *Proceedings of the IEEE Conference on Smart Grid Communications (SmartGridComm)*, 2014.

[8] M. Bicego, F. Recchia, A. Farinelli, S. D. Ramchurn, and E. Grosso. Behavioural biometrics using electricity load profiles. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 1764–1769. IEEE, 2014.

[9] J.-M. Bohli, C. Sorge, and O. Ugus. A privacy model for smart metering. In *Communications Workshops (ICC), 2010 IEEE International Conference on*, pages 1–5, May 2010.

[10] F. Borges. *On privacy-preserving protocols for smart metering systems.* PhD thesis, Springer, 2015.

[11] F. Borges, L. A. Martucci, and M. Mühlhaüser. Analysis of privacy-enhancing protocols based on anonymity networks. In *Smart Grid Communications (SmartGridComm), 2012 IEEE Third International Conference on*, pages 378–383. IEEE, 2012.

[12] D. Boyd and K. Crawford. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15(5):662–679, 2012.

[13] R.E. Brown. Impact of smart grid on distribution system design. In *Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*, pages 1–4, July 2008.

[14] S. Buchegger. Coping with misbehavior in mobile ad-hoc networks. 2004.

[15] E. Buchmann, K. Böhm, T. Burghardt, and S. Kessler. Re-identification of smart meter data. *Personal and ubiquitous computing*, 17(4):653–662, 2013.

[16] J. E. Cabral, J. O. P. Pinto, and A. M. A. C. Pinto. Fraud detection system for high and low voltage electricity consumers based on data mining. In *Power & Energy Society General Meeting, 2009. PES'09. IEEE*, 2009.

[17] A. A. Cárdenas, R. Berthier, R. B. Bobba, J. H. Huh, J. G. Jetcheva, D. Grochocki, and W. H. Sanders. A framework for evaluating intrusion detection architectures in advanced metering infrastructures. *IEEE Transactions on Smart Grid*, 5(2):906–915, 2014.

[18] M. Carpenter, T. Goodspeed, B. Singletary, E. Skoudis, and J. Wright. Advanced Metering Infrastructure Attack Methodology. `http://inguardians.com/pubs/AMI_Attack_Methodology.pdf`, 2009.

[19] M. Chan, D. Estève, C. Escriba, and E. Campo. A review of smart homes - present state and future challenges. *Computer methods and programs in biomedicine*, 91(1):55–81, 2008.

[20] Federal Energy Regulatory Commission. Assessment of demand response and advanced metering. 2008.

[21] D. J. Cook and S. K. Das. How smart are our environments? an updated look at the state of the art. *Pervasive and mobile computing*, 3(2):53–73, 2007.

[22] M. Costache, V. Tudor, M. Almgren, M. Papatriantafilou, and C. Saunders. Remote control of smart meters: Friend or foe? In *Computer Network Defense (EC2ND), 2011 Seventh European Conference on*, pages 49–56, 2011.

[23] C. Cuijpers and B.-J. Koops. Smart metering and privacy in Europe: Lessons from the Dutch case. In *European data protection: coming of age*. Springer, 2013.

[24] S. Deilami, A.S. Masoum, P.S. Moses, and M.A.S. Masoum. Real-time coordination of plug-in electric vehicle charging in smart grids to minimize power losses and improve voltage profile. *Smart Grid, IEEE Transactions on*, 2(3):456–467, Sept. 2011.

[25] EU Directive. 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal of the EC*, 23:6, 1995.

[26] DLMS User Association. DLMS/COSEM protocol `http://www.dlms.com/index2.php`.

[27] C. Doukas and I. Maglogiannis. Bringing iot and cloud computing towards pervasive healthcare. In *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2012 Sixth International Conference on*, pages 922–926. IEEE, 2012.

[28] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In Shai Halevi and Tal Rabin, editors, *Theory of Cryptography*, number 3876 in Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, January 2006.

[29] C. Dwork, M. Naor, T. Pitassi, and G. N. Rothblum. Differential privacy under continual observation. In *Proceedings of the ACM Symposium on Theory of Computing*, STOC '10, pages 715–724, New York, NY, USA, 2010. ACM.

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

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

[32] Z. Erkin and T. Veugen. Privacy enhanced personal services for smart grids. In *Proceedings of the 2nd Workshop on Smart Energy Grid Security*, SEGS '14, pages 7–12, New York, NY, USA, 2014. ACM.

[33] S. Etalle, C. Gregory, D. Bolzoni, E. Zambon, and D. Trivellato. Monitoring industrial control systems to improve operations and security. 2013.

[34] European Commission. 2014/724/EU: Commission recommendation of 10 october 2014 on the data protection impact assessment template for smart grid and smart metering systems, 2014.

[35] European Commission. General data protection regulation, 2016.

[36] M. Faisal, A. A. Cárdenas, and D. Mashima. How the quantity and quality of training data impacts re-identification of smart meter users? In *2015 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 31–36, Nov 2015.

[37] N. Falliere, L. O. Murchu, and E. Chien. W32.Stuxnet Dossier. *White paper, Symantec Corp., Security Response*, 5, 2011.

[38] S. Fischer-Hübner. *IT-security and Privacy: Design and Use of Privacy-enhancing Security Mechanisms*. Springer-Verlag, Berlin, Heidelberg, 2001.

[39] J. C. Foreman and D. Gurugubelli. Identifying the cyber attack surface of the Advanced Metering Infrastructure. *The Electricity Journal*, 28(1):94–103, 2015.

[40] G. Georgiadis and M. Papatriantafilou. Dealing with storage without forecasts in smart grids: Problem transformation and online scheduling algorithm. In *Proceedings of the 29th Annual ACM Symposium on Applied Computing*, SAC '14. ACM, 2014.

[41] M. Gerla and L. Kleinrock. Vehicular networks and the future of the mobile internet. *Computer Networks*, 55(2):457–469, 2011.

[42] Y. Gong, Y. Cai, Y. Guo, and Y. Fang. A privacy-preserving scheme for incentive-based demand response in the smart grid. *IEEE Transactions on Smart Grid*, PP, 2015.

[43] U. Greveler, P. Glösekötterz, B. Justusy, and D. Loehr. Multimedia content identification through smart meter power usage profiles. In *Proceedings of the International Conference on Information and Knowledge Engineering (IKE)*, page 1, Athens, 2012. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), ProQuest.

[44] D. Grochocki, J. H. Huh, R. Berthier, R. Bobba, W. H. Sanders, A. A. Cárdenas, and J. G. Jetcheva. AMI threats, intrusion detection requirements and deployment recommendations. In *Smart Grid Communications (SmartGridComm), IEEE Third International Conference on*, pages 395–400, 2012.

[45] V. Gulisano, M. Almgren, and M. Papatriantafilou. METIS: a two-tier intrusion detection system for advanced metering infrastructures. In *International Conference on Security and Privacy in Communication Systems*, pages 51–68, Beijing, China, 2014. Springer International Publishing.

[46] V. Gulisano, M. Almgren, and M. Papatriantafilou. Online and scalable data validation in advanced metering infrastructures. In *The 5th IEEE PES Innovative Smart Grid Technologies (ISGT) European Conference*, pages 1 — 6, Istanbul, 2014. IEEE.

[47] S. Gunelius. The Data Explosion in 2014 Minute by Minute - Infographic. `http://goo.gl/drrrxG`, July 2014. [last visited March 2017].

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

[49] Harvard Business Review Staff. With Big Data Comes Big Responsibility. `https://hbr.org/2014/11/with-big-data-comes-big-responsibility`, November 2014. [last visited March 2017].

[50] M. Hermann, T. Pentek, and B. Otto. Design principles for industrie 4.0 scenarios. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pages 3928–3937, Jan 2016.

[51] M. Hoeve. Detecting intrusions in encrypted control traffic. In *Proceedings of the First ACM Workshop on Smart Energy Grid Security*, SEGS '13, pages 23–28. ACM, 2013.

[52] J. Homan, S. McBride, and R. Caldwell. IRONGATE ICS Malware: Nothing to See Here... Masking Malicious Activity on SCADA Systems. `https://www.fireeye.com/blog/threat-research/2016/06/irongate_ics_malware.html`, June 2016. [last visited October 2016].

[53] V. M. Igure, S. A. Laughter, and R. D. Williams. Security issues in SCADA networks. *Computers & Security*, 25(7):498–506, 2006.

[54] Y. Iwafune, Y. Yagita, T. Ikegami, and K. Ogimoto. Short-term forecasting of residential building load for distributed energy management. In *Energy Conference (ENERGYCON), 2014 IEEE International*, pages 1197–1204, May 2014.

[55] M. Jawurek, M. Johns, and K. Rieck. Smart metering de-pseudonymization. In *Proceedings of the 27th Annual Computer Security Applications Conference*, pages 227–236. ACM, 2011.

[56] M. Jawurek, F. Kerschbaum, and G. Danezis. Sok: Privacy technologies for smart grids - a survey of options. *Microsoft Res., Cambridge, UK*, 1:1 – 16, 2012.

[57] M. Jelasity and K. P. Birman. Distributional differential privacy for large-scale smart metering. In *Proceedings of the 2nd ACM workshop on Information hiding and multimedia security*, pages 141–146, New York, NY, 2014. ACM.

[58] I. Jolliffe. *Principal component analysis*. Wiley Online Library, 2005.

[59] G. Kalogridis, R. Cepeda, S.Z. Denic, T. Lewis, and C. Efthymiou. Elecprivacy: Evaluating the privacy protection of electricity management algorithms. *Smart Grid, IEEE Transactions on*, 2(4):750–758, Dec. 2011.

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

[61] S. K. Khaitan and J. D. McCalley. Design techniques and applications of cyberphysical systems: A survey. *IEEE Systems Journal*, 9(2):350–365, 2015.

[62] KrebsonSecurity. FBI: Smart Meter Hacks Likely to Spread. `http://krebsonsecurity.com/2012/04/fbi-smart-meter-hacks-likely-to-spread/`, April 2012. [last visited March 2015].

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

[64] Laboratory of Cryptography and System Security (CrySyS Lab), Budapest University of Technology and Economics - Department of Telecommunications. sKyWIper (a.k.a. Flame a.k.a. Flamer): A complex malware for targeted attacks. `http://www.crysys.hu/skywiper/skywiper.pdf`, May 2012. [last downloaded August 2016].

[65] W. Lee and S. J. Stolfo. A framework for constructing features and models for intrusion detection systems. *ACM Trans. Inf. Syst. Secur.*, 3(4):227âĂŞ261, November 2000. 00757.

[66] B. Lohrmann and O. Kao. Processing smart meter data streams in the cloud. In *Innovative Smart Grid Technologies (ISGT Europe), 2nd IEEE PES International Conference and Exhibition on*, pages 1–8, Manchester, UK, 2011. IEEE.

[67] M-Bus Usergroup. M-Bus protocol `http://www.m-bus.com/`.

[68] Malta Independent. Sparks fly over smart meter theft scandal. `http://www.independent.com.mt/articles/2014-02-16/news/sparks-fly-over-smart-meter-theft-scandal-3968892934/`, February 2014. [last visited March 2015].

[69] F.G. Mármol, C. Sorge, O. Ugus, and G.M. Pérez. Do not snoop my habits: preserving privacy in the smart grid. *Communications Magazine, IEEE*, 50(5): 166–172, May 2012.

[70] D. Mashima and A. A. Cárdenas. Evaluating electricity theft detectors in smart grid networks. In *Research in Attacks, Intrusions, and Defenses*, pages 210–229. Springer, 2012.

[71] K. Matsui, H. Ochiai, and Y. Yamagata. Feedback on electricity usage for home energy management: A social experiment in a local village of cold region. *Applied Energy*, 120(0), 2014.

[72] S. McLaughlin, C. Konstantinou, X. Wang, L. Davi, A. R. Sadeghi, M. Maniatakos, and R. Karri. The cybersecurity landscape in industrial control systems. *Proceedings of the IEEE*, 104(5):1039–1057, 2016.

[73] S. McLaughlin, D. Podkuiko, S. Miadzvezhanka, A. Delozier, and P. McDaniel. Multi-vendor penetration testing in the Advanced Metering Infrastructure. In *Proceedings of the 26th Annual Computer Security Applications Conference*, pages 107–116. ACM, 2010.

[74] M. Meingast, T. Roosta, and S. Sastry. Security and privacy issues with health care information technology. In *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, pages 5453–5458. IEEE, 2006.

[75] 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, September 2013.

[76] M. Mitzenmacher and E. Upfal. *Probability and computing: Randomized algorithms and probabilistic analysis*. Cambridge University Press, 2005.

[77] Y. Mo, T. H.-J. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli. Cyber–physical security of a smart grid infrastructure. *Proceedings of the IEEE*, 100(1):195–209, 2012.

[78] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin. Private memoirs of a smart meter. In *Proceedings of the second ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, BuildSys '10, pages 61–66, New York, NY, USA, 2010. ACM.

[79] N. Falliere and L. O. Murchu and E. Chien, Symantec. W32.Stuxnet Dossier version 1.4. `http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/w32_stuxnet_dossier.pdf`, February 2011. [last downloaded August 2016].

[80] A. Pfitzmann and M. Hansen. A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management. *URL: http://dud. inf. tu-dresden. de/literatur/Anon_Terminology_v0*, 34, 2010.

[81] C. W. Potter, A. Archambault, and K. Westrick. Building a smarter smart grid through better renewable energy information. In *Power Systems Conference and Exposition, 2009. PSCE'09. IEEE/PES*, pages 1–5. IEEE, 2009.

[82] M. Raciti and S. Nadjm-Tehrani. Embedded cyber-physical anomaly detection in smart meters. In *Critical Information Infrastructures Security*, volume 7722 of *Lecture Notes in Computer Science*, pages 34–45. 2013.

[83] S. Repo, D. Della Giustina, G. Ravera, L. Cremaschini, S. Zanini, J.M. Selga, and P. Jarventausta. Use case analysis of real-time low voltage network management. In *Innovative Smart Grid Technologies (ISGT Europe), 2011 2nd IEEE PES International Conference and Exhibition on*, pages 1–8, Dec. 2011.

[84] C. Rottondi, G. Mauri, and G. Verticale. A data pseudonymization protocol for smart grids. In *Online Conference on Green Communications (GreenCom), 2012 IEEE*, pages 68–73, Sept. 2012.

[85] C. Rottondi, M. Savi, D. Polenghi, G. Verticale, and C. Krauß. A decisional attack to privacy-friendly data aggregation in smart grids. In *Global Communications Conference (GLOBECOM), 2013 IEEE*, pages 2616–2621. IEEE, 2013.

[86] E. Shi, T.-H. H. Chan, E. G. Rieffel, R. Chow, and D. Song. Privacy-preserving aggregation of time-series data. In *NDSS*, volume 2, pages 1–17, San Diego, CA, 2011. Internet Society.

[87] F. Siddiqui, S. Zeadally, C. Alcaraz, and S. Galvao. Smart grid privacy: Issues and solutions. In *Computer Communications and Networks (ICCCN), 2012 21st International Conference on*, pages 1–5, 2012.

[88] Y. Simmhan, B. Cao, M. Giakkoupis, and V. K. Prasanna. Adaptive rate stream processing for smart grid applications on clouds. In *Proceedings of the 2nd international workshop on Scientific cloud computing*, pages 33–38, New York, NY, 2011. ACM.

[89] W. Stallings and L. Brown. *Computer Security Principles and Practice, Third edition*. Pearson Education, 2015.

[90] G. Strbac. Demand side management: Benefits and challenges. *Energy policy*, 36(12):4419–4426, 2008.

[91] W. Su, H. Eichi, W. Zeng, and M-Y. Chow. A survey on the electrification of transportation in a smart grid environment. *Industrial Informatics, IEEE Transactions on*, 8(1):1–10, Feb. 2012.

[92] J. W. Taylor. An evaluation of methods for very short-term load forecasting using minute-by-minute british data. *International Journal of Forecasting*, 24(4), 2008.

[93] T. Toledo, O. Musicant, and T. Lotan. In-vehicle data recorders for monitoring and feedback on drivers' behavior. *Transportation Research Part C: Emerging Technologies*, 16(3):320 – 331, 2008. Emerging Commercial Technologies.

[94] J. Torriti, M. G. Hassan, and M. Leach. Demand response experience in Europe: Policies, programmes and implementation. *Energy*, 35(4):1575–1583, 2010.

[95] V. Tudor, M. Almgren, and M. Papatriantafilou. Analysis of the impact of data granularity on privacy for the smart grid. In *Proceedings of the 12th ACM Workshop on Workshop on Privacy in the Electronic Society*, WPES '13, pages 61–70, New York, NY, USA, 2013. ACM.

[96] V. Tudor, M. Almgren, and M. Papatriantafilou. Employing private data in AMI applications: Short term load forecasting using differentially private aggregated data. In *2016 Intl IEEE Conferences on Ubiquitous Intelligence Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)*, pages 404–413, July 2016.

[97] SmartGrids – European Technology Platform `http://www.smartgrids.eu/?q=node/163`, June 2011. [last visited March 2015].

[98] S. Wang, L. Cui, J. Que, D-H. Choi, X. Jiang, S. Cheng, and L. Xie. A randomized response model for privacy preserving smart metering. *Smart Grid, IEEE Transactions on*, 3(3):1317 –1324, Sept. 2012.

[99] Y. Yan, Y. Qian, and H. Sharif. A secure data aggregation and dispatch scheme for home area networks in smart grid. In *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, pages 1–6, Dec. 2011.

[100] R. B. Yanai, M. Langberg, D. Peleg, and L. Roditty. Realtime classification for encrypted traffic. In Paola Festa, editor, *Experimental Algorithms*, number 6049 in Lecture Notes in Computer Science, pages 373–385. Springer Berlin Heidelberg, January 2010. 00010.

[101] S. Zanero. Behavioral intrusion detection. In *International Symposium on Computer and Information Sciences*, pages 657–666. Springer Berlin Heidelberg, 2004.

[102] S. Zanero. Analyzing tcp traffic patterns using self organizing maps. *Image Analysis and Processing–ICIAP 2005*, pages 83–90, 2005.

[103] 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. *Smart Grid, IEEE Transactions on*, 2(4):796–808, 2011.