Keynote

Privacy-Preserving Smart Metering: Methods and Applicability

Dominik Engel, Josef Ressel Center for User-Centric Smart Grid Privacy, Security and Control, Salzburg University of Applied Sciences, dominik.engel@en-trust.at

Abstract Privacy-sensitive information can be extracted from the load data which is available in high resolution in smart metering. A number of approaches to privacy-enhancing technologies (PET) have been suggested to provide privacy in smart metering while maintaining (a certain level of) functionality. To date, most of these approaches have not been subjected to real-world use beyond test pilots in model regions. We review types of approaches for privacy-aware smart metering and discuss issues that may prove challenging for widespread adoption.

1. Introduction

With the introduction of smart grids, a more efficient utilization of electrical grid infrastructure is envisioned. For implementing smart grids, the method of choice is the introduction of Information and Communication Technology to monitor grid status and to balance generation and consumption. The data available in smart grids is vastly more accurate and timely than in traditional electrical grids. In the current conception of smart grids, smart meters are used to obtain fine-grained data in the distribution system. Moreover, smart metering provides essential functionality for a variety of use cases, such as demand response management. The collection of this fine-grained data has led to privacy concerns [1, 2]. Lisovich and Wicker [2] report results of a collaboration between researchers from law and engineering. They argue that there “exist strong motivations for entities involved in law enforcement, advertising, and criminal enterprises to collect and repurpose power consumption data” [2, p. 1]. For example, burglars could use the data to determine occupancy patterns of houses to time break-ins. Marketing agencies could identify specific brands of used appliances, which could then be used for targeted advertising. In summary, while there are many useful applications of smart meter data, such as energy saving, network monitoring and tailor-made energy rates, the privacy of this kind of data needs to be ensured.

It has been argued, that approaches relying on policy alone, may prove inadequate to provide a sufficient level privacy and that technological methods that enforce privacy by virtue of “strength of mechanism” need to be employed [3]. Indeed, a number of such technological approaches have been suggested to remedy the (perceived) loss in privacy and still enable smart metering functionality on a broad basis. However, on closer examination many of the approaches suggested in literature show some characteristics, such as data expansion, high computational demands, or excessive demands for the required number of participants, that may make their real-world application infeasible. In this paper, we will give an overview of types of approaches. We will then address the issue of real-world applicability and raise the question if it is indeed possible to provide both, privacy and functionality, in a sufficient amount. At the example of selected approaches, this conundrum will be explored.

2. Extractable Information and Privacy

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

It has been argued that both security and privacy need to be addressed from the earliest stages in the development and standardization process for smart grid technology. The terms "security by design" and "privacy by design", spearheaded by Cavoukian et al. [4], are used to describe principles to allow security and privacy to be built into the system, rather than be treated as add-ons.

There are two kinds of privacy approaches: regulatory-based and technology-based [5]. Important sources for regulatory scenarios and recommendations include the reports of the *M/490 SGCG-SGIS Smart Grid Information Security Working Group* and of the *European Commission Smart Grid Expert Groups Two* for regulatory recommendations for data safety, data handling and data protection, e.g. [6]. Other sources include *Common Criteria for Information Technology Security Evaluation* (ISO/EIC 15408) and country-specific recommendations, such as the *Federal Office for Information Security (BSI)* in Germany, e.g. [7].

The basis for both, regulatory-based and technology-based approaches, is detailed knowledge of what information can be extracted with which tools from the available user data. To date, there is little systematic research on this subject in the context of smart grids. Rajagopalan et al. present [8] an information-theoretic approach to abstract privacy and utility requirements. The authors aim at providing a measure for the amount of information leaked, and also for the utility that is retained in the data at different levels of abstraction.

There are a number of approaches for matching appliance signatures to load profiles to determine which appliances were used at what time and for how long, e.g., [9-11]. This type of method is termed "non-intrusive load monitoring" (NILM. Detection based on NILM is remarkably accurate: Lisovich and Wicker [2] report over 90% accuracy in detecting presence and sleep cycle intervals. The results show that "personal information can be estimated with a high degree of accuracy, even with relatively unsophisticated hardware and algorithms" [2, p. 2]. Leung et al. [12] use genetic algorithms for identification and report flawless identification for up to 10 types of appliances. Liang et al. [13] report on successful identification of appliances in relatively low resolution load profiles, e.g. 30 minute intervals, with the use of data-mining techniques.

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

In general, there is a close relation between the resolution in which the load data is available and the extractable information. As not all extractable information is necessarily privacy-sensitive, a comprehensive and formal account on how extractable information, such as type or brand of appliance, relates to personal information, and how such data items could be combined by a potential attacker. To date there is no formal investigation on what information can be extracted by which method at what resolution, and what kind of threat this may represent to an individual's privacy.

3. Methods and Applicability

A number of technological privacy-enhancing technologies (PET) have been proposed for smart metering. Recent surveys have been conducted by Jawurek et al. [3] and Erkin et al. [14]. In the following, we give an overview of the types of approaches, without aiming at listing or detailing all existing approaches, and point out properties that may prevent real-world use or at least prove a challenge should these approaches be deployed in the real world.

3.1 Anonymization/Pseudonymization

The classic approach, and the only approach that is widely used in the real world at this point in time, is anonymization or pseudonymization of smart metering data. The consumption data and the personal data are split and stored separately.
Methods for ***de-anonymization*** are a major threat for these types of approaches. It has been shown that even after anonymization or pseudonymization, data items can still be attributed to the individual that originated them. For example, in the area of social networks, it has been shown by Backstrom et al. [15] that anonymization is somewhat difficult, because individual users can be traced based on structural cues evident in the network even after anonymization. Jawurek et al. [16] show that de-anonymization can also be done in the smart grid user domain. This structural traceability is a problem for schemes that rely on anonymization or pseudonymization only without the use of additional encryption.

3.2 Simple Aggregation

Simple aggregation tries to hide data related to individuals by aggregating over a number of households, e.g., all households in a neighborhood are network (NAN). For example, Bohli et al. [17] propose a privacy scheme in which high resolution smart meter readings are aggregated at NAN level and only the aggregate is sent to the utility. They introduce two solutions both with and without involvement of trusted third parties.
A possible issue with this kind of approaches is the number of households required. If a NAN only has a small number of households, traces of individual data can still be identified in the aggregate. Furthermore, these approaches often assume complete trust between the households in a NAN, as the data is aggregated in a hop-by-hop manner. If one participant should start an attack, the schemes can be easily compromised. Introducing a dedicated aggregator in each NAN only moves the issue to a different part of the system, as in this case, the aggregator needs to be afforded complete trust by all parties. In general, the ***adversary models*** which are used to analyze PET in smart

grids often exclude malicious attackers. Most authors evaluate their approaches in honest-but-curious adversary models.

3.3 Multiple Resolutions

Due to the inherent link between load data resolution and privacy, splitting the load data into a variety of different resolutions, each associated with different authorization levels, has been proposed by a number of contributions.

For example, the anonymization scheme proposed by Efthymiou and Kalogridis [18] is based on two different resolutions: a low resolution that can be used for billing purposes, and a high resolution that allows further investigation. This scheme employs a trusted third party escrow service. Engel [19, 20] proposes the use of the wavelet transform to generate a whole cascade of different resolutions. The approach is combined with a conditional access scheme: each wavelet resolution is encrypted with a different key, allowing differentiated access management. By using a suitable wavelet filter, it is ensured that the sum of the original data is preserved over all resolutions.

For application in the real world, the ***requirements of use cases*** with respect to data resolution need to be clarified. It could turn out that most of the more interesting use cases (except for billing), such as distribution system monitoring, may require high resolution data, rendering a cascade of lower and medium resolutions useless. Furthermore, many of these use cases may require the data in (near) real-time. Using the wavelet transform to create a number of resolutions is at odds with this requirement, as a sufficient amount of data needs to be available for transformation.

3.4 Masking

Masking relates to approaches which add numerical artifacts, e.g., random sequences to the original load data to obfuscate individual contribution. The added artifacts are constructed in such a way that they cancel each other out upon aggregation. The aggregator can therefore combine the data of all participant to create an accurate aggregation, but cannot gain access to individual contribution. For example, Kursawe et al. [26] propose such an aggregation protocol, which compared to other approaches has the advantage of relatively low computational complexity.

For real-world use, the issue of creating the random secret shares among each group of participants needs to be addressed. In [26] this is achieved by either selecting a leader among the participants, or by relying on a trusted third party to create the final shares (which exhibit the property of cancelling each other out) from the individually generated random shares. Again, this relates to the assumed underlying ***adversary and trust models***. Another issue, as Jawurek et al. [3] point out, is ***fault tolerance***: if a single participant fails (e.g., due to a hardware error), the whole aggregate is affected.

3.5 Differential Privacy

As Dwork [27] puts it, differential privacy, roughly speaking, “ensures that (almost, and quantifiably) no risk is incurred by joining a statistical database”. Adding or removing an item from the database will not (or only to a very limited degree) affect the result of statistical computations. This is commonly achieved by the distributed generation of noise which is added to the individual data contribution.

Shi et al. [28] propose a scheme for adding random noise to time series data using a symmetric geometric distribution. An advantage of this scheme is that the participants need not trust each other, nor rely on a trusted aggregator. As another example, Acs and Castelluccia [29] obscure individual data sets by adding Laplacian noise, which is jointly generated by the participants.

As Shi et al. [28] point out themselves, the issue of ***data pollution***, i.e., a malicious participant or a group of malicious participants injecting false data. Furthermore, although keeping the contribution of each participant private, the protocols exhibit little to no ***fault tolerance*** of participants [3]. Finally, in order to achieve a high level of (differential) privacy, the ***number of participants*** needs to be large.

3.6 Secure Signal Processing

Secure Signal Processing (SSP) refers to the possibility to perform certain computations, such as aggregation in the encrypted domain. A commonly employed mechanism in SSP is *homomorphic encryption*, which allows some specific manipulations of the ciphertext to be reflected in the plaintext domain.

For example, Li et al. [21] propose an overlay network in a tree-like topology and the use of a Paillier cryptosystem [22]. Garcia and Jacobs [23] combine *secret sharing* with a Paillier cryptosystem to add flexibility in the aggregation (at the expense of additional computational complexity). Erkin and Tsudik [24] extend the idea of homomorphic encryption of smart meter readings by splitting the module into random shares, which, in combination with a modified Pailler cryptosystem, allows flexible spatial and temporal aggregation for different use cases, such as billing or network monitoring. The complexity of this approach is lower than that presented in [23]. Engel and Eibl [25] show that SSP can be combined with multi-resolution signal processing, increasing the degrees of freedom.

For real-world applicability, a number of factors need to be taken into account. For most schemes, homomorphic additivity comes at the cost of ***data expansion***. For example, when a Paillier cryptosystem is used, a plaintext of size n is encrypted to a ciphertext modulo n^2, thus doubling the number of bits needed for data representation in the encrypted domain. The ensuing data expansion, which grows with the number of participating nodes, may prove a challenge, especially if communication is done over low-bandwidth power line carrier. ***Computational complexity*** is another issue to be considered. Compared to other ciphers, homomorphic encryption systems are often more demanding. Furthermore, unlike standardized cryptographic ciphers, such as AES and RSA, homomorphic encryption schemes are not commonly supported by standard crypto hardware (this of course may change if a standard for homomorphic encryption is brought forward). For a smart meter roll-out to be successful, the required computational complexity may prove to be too high to allow manufacturing devices that satisfy ***economic feasibility***. Furthermore, high computational demands may lead to energy demands that are significantly higher than traditional meters, and low ***energy efficiency*** for smart meters may negatively impact consumer acceptance.

Another issue, as with previously discussed approaches, lies with the ***number of required participants*** and the underlying ***trust model***, i.e., what level of mutual trust needs to be afforded among the participants. For real-world use both need to be carefully investigated. In many homomorphic encryption scheme, participants are required to use the same key, which implies that they need to trust each other with their meter readings.

3.7 Rechargeable Batteries

There are a number approaches that propose to install rechargeable batteries at the end-user home to mask the real profile. In the approach presented by Kalogridis et al. [30], a flat load curve is produced by constant charging of a battery as far as possible, matching the household consumption over time. Varodayan and Khisti [31] argue that with this best-effort approach, privacy may still leak through lower frequencies. They propose the use of a "stochastic battery" which instead of constant charging employs a randomized model to decrease information leakage.

While in theory this is an effective approach, the practical applicability remains questionable due to the ***high costs*** of installing batteries. Furthermore, the energy loss introduced by using a battery buffer leads to low ***energy efficiency*** of this approach, which, as mentioned above, is not desirable in general, but specifically detrimental in the context of smart grids.

4. Conclusion

In summary, there are a number of issues that need to be considered when PET for smart metering are to be deployed in the real world:

- De-anonymization
- Data expansion
- Usability of (aggregated, reduced) data for smart grid use-cases
- Computational complexity
- Scalability
- Number of required participants
- Fault tolerance
- Realistic adversary model
- Energy efficiency
- Economic feasibility

Addressing these issues will prove a challenge. To date there is not a single approach that will readily solve all of these issues completely, but there are a number of promising suggestions. The answer to the question, whether the requirements regarding privacy and functionality can be balanced, can be formulated as a cautious "Yes, but…". Yes, there are good and promising suggestions, but we need to make sure to address all issues that may occur in the real world carefully. A more definite answer will only be provided over time, when researchers and experts from both domains, privacy and energy, put their heads together and work out the details in terms of a standardized method for privacy-aware smart metering, possibly under European mandate M/490. An even more definite answer will be given, if and when such standardized methods are actually applied and rolled out all over Europe.

Acknowledgements

The financial support by the Austrian Federal Ministry of Economy, Family and Youth and the Austrian National Foundation for Research, Technology and Development is gratefully acknowledged. Thanks are due to the experts of our company partners Salzburg AG and Salzburg Wohnbau for providing an insight on real-world issues of Smart Grids.

References

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

[2] M. A. Lisovich and S. B. Wicker, "Privacy concerns in upcoming residential and commercial demand-response systems," *IEEE Proceedings on Power Systems*, vol. 1, no. 1, 2008.

[3] M. Jawurek, F. Kerschbaum, and G. Danezis, "Privacy technologies for smart grids - a survey of options," Microsoft Research, Tech. Rep., 2012.

[4] A. Cavoukian, J. Polonetsky, and C. Wolf, "SmartPrivacy for the smart grid: embedding privacy into the design of electricity conservation," *Identity in the Information Society*, vol. 3, pp. 275–294, 2010, 10.1007/s12394-010-0046-y. [Online]. Available: http://dx.doi.org/10.1007/s12394-010-0046-y

[5] Z. Fan, P. Kulkarni, S. Gormus, C. Efthymiou, G. Kalogridis, M. Sooriyabandara, Z. Zhu, S. Lambotharan, and W. Chin, "Smart grid communications: Overview of research challenges, solutions, and standardization activities," *IEEE Communications Surveys & Tutorials*, vol. PP Issue:99, no. 99, pp. 1–18, 2012.

[6] European Commission Task Force Smart Grids, Expert Group 2: Regulatory Recommendations for Data Safety, Sata Handling and Data Protection, "Report," http://ec.europa.eu/energy/gas_electricity/smartgrids/doc/expert_group2.pdf, Feb. 2011, online.

[7] Bundesamt für Sicherheit in der Informationstechnik (German Federal Office for Information Security), "Protection profile for the gateway of a smart metering system," https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/SmartMeter/PP-SmartMeter.pdf, March 2013, Version 1.2. [Online]. Available: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/SmartMeter/PP-SmartMeter.pdf

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

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

[10] H. Y. Lam, G. S. K. Fung, and W. K. Lee, "A novel method to construct taxonomy of appliances based on load signatures," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 2, pp. 653–660, 2007.

[11] J. Liang, S. Ng, G. Kendall, and J. Cheng, "Load Signature Study Part I: Basic concept, structure, and methodology," *IEEE Transactions on Power Delivery*, vol. 25, no. 2, pp. 551–560, 2010.

[12] S. K. J. Leung, S. H. K. Ng, and W. M. J. Cheng, "Identifying appliances using load signatures and genetic algorithms," in *Proceedings International Conference on Electrical Engineering (ICEE)*, Hong Kong, Jul. 2007.

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

[14] Z. Erkin, J. Troncoso-Pastoriza, R. Lagendijk, and F. Perez-Gonzalez, "Privacy-preserving data aggregation in smart metering systems: An overview," *Signal Processing Magazine, IEEE*, vol. 30, no. 2, pp. 75–86, March.

[15] L. Backstrom, C. Dwork, and J. Kleinberg, "Wherefore art thou R3579X? : anonymized social networks, hidden patterns, and structural steganography," in *Proceedings of the 16th international conference on World Wide Web*, ser. WWW '07. New York, NY, USA: ACM, 2007, pp. 181–190.

[16] M. Jawurek, M. Johns, and K. Rieck, "Smart metering de-pseudonymization," in *Proceedings of the 27th Annual Computer Security Applications Conference*, ser. ACSAC, New York, NY, USA: ACM, 2011, pp. 227–236. [Online]. Available: http://doi.acm.org/10.1145/2076732.2076764

[17] J.-M. Bohli, C. Sorge, and O. Ugus, “A privacy model for smart metering,” in *Proc. IEEE Int Communications Workshops (ICC) Conf*, 2010, pp. 1–5.

[18] C. Efthymiou and G. Kalogridis, “Smart grid privacy via anonymization of smart metering data,” in *Proceedings of First IEEE International Conference on Smart Grid Communications*, Gaithersburg, Maryland, USA, Oct. 2010, pp. 238–243.

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

[20] ------, “Wavelet-based load profile representation for smart meter privacy,” in *Proc. IEEE PES Innovative Smart Grid Technologies (ISGT’13)*, Washington, D.C., USA, Feb. 2013, pp. 1–6. [Online]. Available: http://dx.doi.org/10.1109/ISGT.2013.6497835

[21] F. Li, B. Luo, and P. Liu, “Secure information aggregation for smart grids using homomorphic encryption,” in *Proceedings of First IEEE International Conference on Smart Grid Communications*, Gaithersburg, Maryland, USA, Oct. 2010, pp. 327–332.

[22] P. Paillier, “Public-key cryptosystems based on composite degree residuosity classes,” in *Proceedings of Eurocrypt ’99, Advances in Cryptology*, ser. Lecture Notes in Computer Science, J. Stern, Ed., vol. 1592. Prague, Czech Republic: Springer, May 1999, pp. 223–238.

[23] F. Garcia and B. Jacobs, “Privacy-friendly energy-metering via homomorphic encryption,” in *Security and Trust Management*, ser. Lecture Notes in Computer Science, J. Cuellar, J. Lopez, G. Barthe, and A. Pretschner, Eds. Springer Berlin / Heidelberg, 2011, vol. 6710, pp. 226–238. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-22444-7_15

[24] Z. Erkin and G. Tsudik, “Private computation of spatial and temporal power consumption with smart meters,” in *Proceedings of the 10th international conference on Applied Cryptography and Network Security*, ser. ACNS’12. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 561–577. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-31284-7_33

[25] D. Engel and G. Eibl, “Multi-resolution load profile representation with privacy-preserving aggregation,” in *Proceedings of IEEE Innovative Smart Grid Technologies (ISGT) 2013*. Copenhagen, Denmark: IEEE, Oct. 2013, to appear.

[26] K. Kursawe, G. Danezis, and M. Kohlweiss, “Privacy-friendly aggregation for the smart grid,” in *Privacy Enhanced Technology Symposium*, 2011, pp. 175–191.

[27] C. Dwork, “Differential privacy: A survey of results,” in *Theory and Applications of Models of Computation*, ser. Lecture Notes in Computer Science, M. Agrawal, D. Du, Z. Duan, and A. Li, Eds. Springer Berlin Heidelberg, 2008, vol. 4978, pp. 1–19. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-79228-4_1

[28] E. Shi, R. Chow, T. h. Hubert Chan, D. Song, and E. Rieffel, “Privacy-preserving aggregation of time-series data,” in *Proc. NDSS Symposium*, February 2011.

[29] G. Acs and C. Castelluccia, “I have a dream! (differentially private smart metering),” in *Proc. Information Giding Conference*, 2011, pp. 118–132.

[30] G. Kalogridis, C. Efthymiou, S. Z. Denic, T. A. Lewis, and C. R., “Privacy for smart meters: Towards undetectable applicance load signatures,” in *Proceedings of First IEEE International Conference on Smart Grid Communications*, Gaithersburg, Maryland, USA, Oct. 2010, pp. 232–237.

[31] D. Varodayan and A. Khisti, “Smart meter privacy using a rechargeable battery: minimizing the rate of information leakage,” in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2011)*, Prague, Czech Republic, May 2011.