Conditional Access Smart Meter Privacy Based on Multi-Resolution Wavelet Analysis

Dominik Engel
Salzburg University of Applied Sciences
Urstein Sued 1
Salzburg, Austria
dominik.engel@fh-salzburg.ac.at

ABSTRACT

Smart Metering is an important component of Smart Grids. Detailed load profiles are available through smart metering at a high resolution. Load profiles allow inferring detailed information on the end user by non-intrusive load monitoring. Therefore, these load profiles need to be regarded as sensitive data, and treated with security and privacy in mind. We propose a method that allows conditional access to different resolution levels of the load data, allowing access on a "need-to-know" basis only. For this purpose, a multi-resolution representation of the load data is created using the simple Haar wavelet transform. Securing the portions of the wavelet representation pertaining to each resolution with a unique key allows to implement conditional access for smart meter data.

Categories and Subject Descriptors

E.3 [**Data**]: Data Encryption

1. INTRODUCTION

Smart Grids have recently come to the focus of attention of a large number of research programmes. The powerful combination of communication technology with electrical grids leads the energy infrastructure into a new paradigm.

However, this transition is not without challenges. There are security concerns: An overview of current research in smart grid security can be found in [3], a comprehensive proposal for securing smart grid infrastructure is given by [16]. Privacy is another critical area, related to security. Smart meters form a core component of smart grids. Each of these devices contains a processor, as well as storage and communication facilities and is capable of transmitting detailed usage statistics to the energy provider. While the exact granularity of the transmitted data is not finally specified, and may differ by country, it seems likely that the interval between single measurements will lie between 1 and 30 minutes. The availability of per-customer load profiles on such a fine granularity raises privacy concerns [17, 15, 8, 1]. A comprehensive discussion of such privacy concerns can be found in [14, 13].

There are a number of approaches for matching appliance signatures to load profiles to determine which appliances were used at what time and for how long, e.g. [6, 9, 12]. This type of method is termed "non-intrusive load monitoring" (NILM) or "non-intrusive appliance load monitoring" (NALM). Detection based on NILM is remarkably accurate: [14] reports over 90% accuracy in detecting presence and sleep cycle intervals. The results reported in [13] show that "even with relatively unsophisticated hardware and data-extraction algorithms, some information about occupant behavior can be estimated with a high degree of accuracy". [10] uses genetic algorithms for identification and report flawless identification for up to 10 types of appliances.

In [14] results of a collaboration between researchers from law and engineering are reported. The authors argue that there "exist strong motivations for entities involved in law enforcement, advertising, and criminal enterprises to collect and repurpose power consumption data." For example, burglars could use the data to determine occupancy patterns of houses to time break-ins, or NILM may be used to identify specific brands of appliances, which can then be used for targeted advertising. In summary, while there are many useful applications of smart meter data, such as energy saving and tailor-made energy rates, the privacy of this kind of data needs to be secured.

In this paper, we discuss the utility of wavelet multi-resolution analysis (MRA) to afford privacy to smart meter load profiles. We evaluate MRA-based privacy based on multi-layer conditional access. Access to resolutions can be defined individually, ranging from a low frequency dataset over multiple refinement datasets to the highest resolution representation. Conditional access to the different resolutions has the advantage that a information can be made accessible on a "need-to-know" principle.

Thereby it is possible to use data at lower resolutions, i.e. a (daily) average, for accounting purposes with the energy provider, while the high resolution data remain secured from access. Access to these higher resolutions can selectively be granted to third parties, e.g. to serve as input to energy-saving tools, which match load signatures to determine appliances with high power consumption. The contribution of the proposed approach is to provide both security and privacy to the level specified as needed. For evaluation we use data from a test project conducted by Salzburg AG, an Austrian energy provider.

ISABEL '11, October 26-29, Barcelona, Spain

The rest of this paper is organized as follows: Section 2 gives an overview of related approaches for smart meter privacy and security. Section 3 introduces the application of multi-resolution wavelet analysis for the representation of smart meter data. Section 4 details the proposed conditional access encryption scheme. Section 5 evaluates security and privacy provided by the proposed scheme, in comparison to other schemes. In Section 6 the complexity and computational demands of the proposed are discussed. Section 7 concludes and gives an outlook on future research.

2. RELATED WORK

There are a number of proposals for secure transmission of smart meter data, e.g. [19] proposes a secure multi-cast protocol that automatically derives group memberships and verifies configuration performance; [2] proposes a security protocol for smart meter aggregation that provides hop-by-hop security, while still providing end-to-end security. In principle, the approach proposed here is compatible with many of the basic methods used in secure transmission, such as aggregation along a spanning tree. The advantage of the approach proposed here, is the possibility to determine the available granularity of the data along the spanning tree.

There are suggestions for security methods that also preserve privacy. [11] proposes to employ homomorphic encryption for smart meter data. Specifically a Paillier cryptosystem is used, which supports the additive homomorphic property, to enable aggregation of smart meter data in the encrypted domain. The approach is evaluated in an honest-but-curios adversary model.

There are some approaches that propose to install rechargeable batteries at the end user home to mask the load profile [7, 18]. While in theory this is an effective approach, the practical applicability remains questionable due to the high costs of installing batteries and the energy loss introduced by using a battery buffer.

The authors of [14] use their NILM algorithm on power consumption data from a two-week experiment to infer individual information and usage patterns. This endeavor is highly successful on high resolution data (15 second intervals). The authors then investigate the performance of their algorithm in the face of downsampled data, i.e. decreased resolution. They report that the algorithm performance "degrades quite gracefully". "Meaningful estimates" are possible even for 20 minute intervals. The observed graceful degradation for load signature detection supports the representation of the smart meter data in different resolutions, as each resolution will exhibit different detection properties.

[5] proposes an anonymization scheme that is based on two different resolutions. This scheme employs a trusted third party escrow service. Two smart meter data sets are generated: One of low frequency that can be used for billing purposes, and one of high frequency that allows further investigation. Only for the low frequency data set a mapping to the corresponding user is provided, foremost to allow the energy provider to invoice the user for the consumed energy. The high frequency data set is not attributable to a user, at least not by the energy provider.

In this scenario, two IDs have to be used by the smart meter hardware: HFID, or High-Frequency ID, which remains anonymous, and LFID, or Low-Frequency ID, which is attributable. Each message that is communicated from the smart meter, needs to include one of these IDs. In order to keep the mapping HFID and smart meter hidden from the energy provider, the authors propose an agnostic data aggregator (operated by the third party escrow service) that collects high frequency profiles from a number of smart meters. The validity of the used IDs is verified by the third party escrow service. Low-frequency data is forwarded including the LFID. HFID data is aggregated across multiple smart meters and forwarded without the corresponding HFIDs.

As [3] points out, there are a number of security issues with the approach proposed by [5]. As with all systems that rely on a trusted third party, a compromise of this party is devastating. Furthermore, an attacker could try to match high-frequency data to low-frequency data (and thereby to a unique user) by collecting high-frequency information over time, summing up this information and matching it to observed low-frequency data. A third issue is the fact that both LFID and HFID need to be stored in hardware which is in the sphere of control of the users. This may lead to the possibility to tamper the hardware and manipulate the IDs.

3. WAVELET MULTI-RESOLUTION-ANALYSIS OF SMART METER DATA

In [5] representations of the smart meter data in two resolutions are created, a low frequency and a high frequency resolution. In principle, this is a valid approach in terms of granting each party access only to the information that is needed for the processing demands of this party.

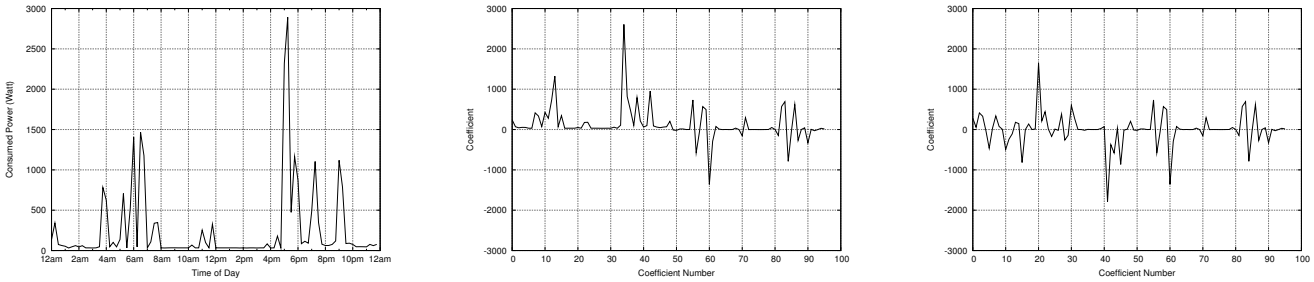
As a first stage of our proposed approach, we generalize this idea. Instead of creating only two variants that are separate, we generate a hierarchy of resolutions in an integrated representation. The best suited tool for this endeavor is the discrete wavelet transform (DWT).

A suitable wavelet transform is applied to the original smart meter data. This leads to a low frequency and a high frequency band. As an example, we use the Haar wavelet transform, which in principle only consists of calculating averages and differences.

To obtain a multi-resolution representation of the original signal, the wavelet analysis step is recursively applied to the low pass subband, up to a maximum level m. The low-frequency portion at each step presents the data at a resolution with half the number of samples of the next higher resolution. The resolution level corresponding to the highest decomposition depth m is referred to as R_0, and has a size of 2^{-m} samples.

The synthesis step of the inverse wavelet transform starts with R_0. Each next higher resolution can be obtained by applying the inverse wavelet transform to the low-pass subband (i.e., the lower resolution) and the corresponding high-pass subband. In this way, m further resolutions can be obtained. Typically, the resulting subbands are represented in a single bitstream.

Figure 1 shows an example of the wavelet decomposition of a smart meter signal. Figure 1(a) shows the original signal. Figure 1(b) shows the first level of decomposition into a low-pass and a high-pass subband. Figure 1(c) shows a wavelet decomposition with a maximum decomposition depth $m = 5$. In this example, the interval between smart metering values is 15 minutes. Therefore, 96 values are produced within 24 hours. The lowest resolution of a level 5 wavelet decomposition in this case contains 3 values, which roughly correspond to the average load during



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

Figure 1: Example for Wavelet Decomposition of Smart Meter Data

the first, second and third eight-hour interval, respectively, of the 24-hour interval. By combining subbands L_0 and H_1 the resolution R_1 can be obtained, which contains double the number of samples as R_0. Depending on the needed granularity, higher resolutions can be reconstructed by applying the inverse wavelet transform, up to the highest resolution R_5, which contains the original 96 values.

To implement multi-resolution analysis in a manner that is suitable for smart metering devices, wavelet lifting [4] is the best approach. This view on the wavelet transform factors wavelet filters into lifting steps, which for many filters encompass only basic operations.

For the Haar wavelet, the lifting steps can easily be implemented by hardware with lowest computing power, such as the smart meter environment. Furthermore the transformation is lossless, and the aggregation is equivalent to subsampling.

4. CONDITIONAL ACCESS FOR SMART METER DATA

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

We propose to use the conditional access paradigm for smart-metering data in multi-resolution representation. Each high-pass subband is encrypted with a different key. The lowest resolution is left unencrypted. The whole datastream can be transmitted over the Smart Grid communications network. Access to the different resolutions is thereby only granted to parties that hold the needed keys, as illustrated by Figure 2. The lowest resolution remains accessible to the network provider at all times to enable invoicing. In the example above this is done by leaving R_0 unencrypted (which in principle mirrors the situation in current energy networks). Alternatively, R_0 could also be encrypted with an appropriate key that allows access to the energy provider.

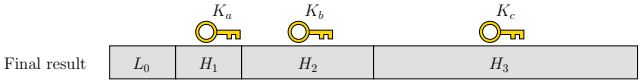


Figure 2: Final Bitstream Produced by Smart Meter

This scheme allows flexible control by the end-user how access is granted to smart meter data. For example, the energy network operator may be granted access to the lowest resolution for billing purposes, but the end-user may not be willing to provide any detailed usage statistics to the energy network operator. Through the conditional access scheme, end-users can also decide to use the services of third parties, by encrypting the relevant portions of their metering data so that the third party offering a service can access the needed data. Such services could include load signature matching to identify appliances with high energy consumption. Aggregation services through a trusted third party are another possible type of service.

The proposed scheme also enables the relaying of data. For example, the data needed by a third party analysis lab can be forwarded in encrypted form by an aggregator, or even the energy network operator. While the parties forwarding the data on route to the destination can access the low-frequency data, there is no access to the higher resolutions.

Of course, the proposed scheme requires the smart meter hardware to provide functionality for wavelet lifting and encryption, and to support the manual or automatic setting of encryption keys for the higher resolutions.

5. SECURITY

Following [11], we assume an "honest-but-curious" adversary model, i.e. all parties are assumed to follow the protocol ("honest"), but within this limitation try to infer as much information about the other participating parties as possible ("curious"). In this setting, the proposed scheme is very successful, as detailed data of the high resolutions remains protected from unauthorized access, including eavesdroppers. There are no methods to infer the higher resolutions by using the information from lower resolutions. Therefore, the higher resolutions are secure, provided that state-of-the art cryptographic ciphers are used.

If the adversary model is changed to malicious, an adversary will also tamper the data to change aggregation or billing data. Under these settings, the homomorphic encryption model proposed by [11] suffers from the fact that all

homomorphic encryption schemes are malleable and allows in-transit tampering. The scheme proposed here is widely agnostic to the used encryption scheme, and supports traditional, non-homomorphic encryption, which do not suffer from this malleability. Integrity checks can be added to prevent in-transit tampering. However, as in the scheme proposed by [11], the proposed scheme does not prevent tampering of smart meter data at the point of origin, i.e. if a tampered smart meter produces fake data, this is not recognized. To prevent this kind of tampering, the proposed scheme needs to be combined with trusted computing (e.g. [16]).

Other than the scheme proposed by [11], the method proposed here does not allow a non-trusted aggregator. While the proposed method allows to control the level of detail an aggregator is allowed to process, there is no restriction on readability on this level of data for the aggregator. The recipients of the data, i.e. the key holders for a certain resolution, need to be trusted with the data at the given resolution.

Regarding successful privacy protection of the higher resolutions, the proposed scheme has an advantage over the scheme proposed by [5]. As stated by [3], the privacy afforded by the scheme proposed by [5] may be compromised by data aggregation through an eavesdropper. The eavesdropper could link LFID and HFID by summing up high frequency data that he observes. Such an attack is not possible in the scheme proposed here, as all high frequency data is transmitted in encrypted form.

6. COMPUTATIONAL COMPLEXITY

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

The computational demands for encryption depends on the used encryption scheme. For standard encryption schemes, efficient hardware implementations exist that can be integrated into the smart meter hardware. Depending on the desired scenario, symmetric encryption alone can be used, or in combination with asymmetric encryption. The latter case is computationally more demanding but benefits from the support for public key infrastructures, such as proposed by [16].

Some overhead is introduced for key management, and potentially for the creation of session keys. Both are computationally inexpensive, and implementations should be easily transferable to smart meter hardware, given that the other tasks above should be implementable on the underlying smart meter hardware.

To explore the complexity of the proposed approach empirically, we use smart meter data from an Austrian energy provider, which was generated by real households over a period of 18 months. In our tests, we use 400 load profiles. The load profiles originate from Siemens smart meter hardware, model TD3510 (3 phase, 100 Amp.). The sampling interval is 15 minutes, i.e. 96 readings a day. Three scenarios were implemented: (i) Symmetric encryption: AES with 128 bit keys, (ii) Asymmetric encryption only: RSA with 2048 bit keys, (iii) Hybrid encryption: 128 bit AES session keys encrypted with 2048 bit RSA keys. In each scenario the following steps are executed: (i) Level 5 wavelet transform of the load profile, (ii)) Generation of different keys to encrypt resolutions R_1 through R_5 (R_0 is left unencrypted), (iii) Encryption of R_1 through R_5, each with a different key.

	WAV	AES	RSA	HYB
Avg. Exec. Time (ms)	35.9	246.4	1366.3	1425.5
Std. Deviation	3.4	17.7	18.9	76.8

Table 1: Execution Times for Test Implementation: Total for 400 load profiles, averaged over 1000 executions

The implementation was done in Java (OpenJDK 64-Bit Server VM). The Java standard implementation of the cryptographic routines were used. The Haar wavelet transform was implemented as lifting steps. No special optimization was performed. The tests were run on a Sun Fire V20z with two AMD Opteron Processors 244 and 8GB of RAM, running 64-bit Ubuntu Linux 10.04.1 with kernel version 2.6.32. We note that the execution environment for smart meter hardware will be extremely restricted and generally not be comparable to this test setup. However, the test setup is suited for obtaining empirical data on the comparative performance of the possible scenarios, and to estimate the computational complexity of the wavelet transform in relation to encryption.

Table 6 shows timing results in milliseconds comparing wavelet transform only (WAV) with AES (128 bit), pure RSA (2048 bit) and hybrid encryption (HYB) using an AES 128-bit session key encrypted with RSA (2048 bit). Due to the small amount of time needed for the transformation and encryption of a single load profile, in order to get valid results, timing was done for a batch of 400 load profiles. Each load profile was subjected to 1000 encryptions in each of the aforementioned categories. In Table 6, each entry corresponds to the total transformation/encryption time of 400 different load profiles, averaged over 1000 encryptions.

It can be seen that indeed the computational demands for the wavelet transform are negligible. On average, the transformation of 400 load profile takes 36 ms. Regarding demand for encryption, as expected, symmetric encryption with AES is the fastest method by far. In the used test setup, this approach outperforms the other two encryption approaches by a factor of 5. In many application scenarios, where the superior key management of public key cryptography is not needed, this advantage will make symmetric encryption a prime candidate. Due to the limited size of the subbands, public key cryptography can be used directly on the load data. For our test setup, all subbands can be encrypted using 2048 bit RSA keys. We compare this approach to a hybrid approach, in which a symmetric session key is encrypted with a public key. For larger plaintext data sets the hybrid approach allows to combine the speed of symmetric encryption with asymmetric key management. It can be observed that the hybrid approach in our scenario is slower than the pure asymmetric approach. This is due to the fact that the load profile subbands are limited in size. Of course, for larger data sets using pure asymmetric encryption is not feasible and the hybrid approach would have to be used. However, it can be rated an advantage that the multi-resolution representation of the load profiles allows the direct application of public key cryptography.

7. CONCLUSION AND OUTLOOK

We have shown that a multi-resolution representation of smart-meter data is a way to balance the need for privacy with the additional functionality introduced by the smart meter load profiles. By using multiple keys to encrypt each resolution separately, the scheme affords end-user control of access to different granularities of the data. Apart from privacy, due to encrypting the higher resolutions, the proposed scheme also implements secure transmission of the load profiles and prevents unauthorized access by eavesdroppers.

The scheme proposed here fits neatly into the larger frameworks proposed to date, such as [16], as it is compatible with other approaches for securing smart grid communication, including authentication, integrity checking, and the integration into smart grid public key infrastructure.

Regarding computational complexity, some overhead is introduced. However, we have shown that the simple Haar wavelet transform, implemented as lifting steps, has very low demands. The computational demands for encryption of the higher resolution subbands are higher, especially if an asymmetric or hybrid approach is chosen. However, communication in the smart grid will have to be secured by cryptographic means, and smart meters are no exception. Therefore, it is likely that smart metering hardware will be required to support encryption. The additional computational overhead for multi-resolution representation and multiple key-handling is acceptable, especially when seen with the background of this requirement for secure communication and authentication.

Future work will focus on the details of integrating the proposed approach into larger smart grid communication frameworks. To support privacy preserving data aggregation by non-trusted data aggregators, we will investigate if, based on the simple operations used in Haar wavelet lifting, the proposed scheme is compatible with the homomorphic encryption approach proposed by [11]. Finally we will explore possible advantages of using more sophisticated wavelets for representing the load profiles.

8. ACKNOWLEDGMENTS

The author would like to thank Salzburg AG for providing test data.

9. REFERENCES

[1] R. Anderson and S. Fuloria. On the security economics of electricity metering. In *Proceddings of the Ninth Workshop on the Economics of Information Security (WEIS 2010)*, June 2010.

[2] A. Bartoli, J. Hernández-Serrano, M. Dohler, A. Kountouris, and D. Barthel. Secure lossless aggregation for smart grid m2m networks. In *Proceedings of First IEEE International Conference on Smart Grid Communications*, pages 333–338, Gaithersburg, Maryland, USA, Oct. 2010.

[3] T. Baumeister. Literature review on smart grid cyber security. Technical report, University of Hawaii at Manoa, Dec. 2010.

[4] I. Daubechies and W. Sweldens. Factoring wavelet transforms into lifting steps. *J. Fourier Anal. Appl.*, 4(3):247–269, 1998.

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

[6] G. Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12), 1992.

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

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

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

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

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

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

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

[14] M. A. Lisovich and S. B. Wicker. Privacy concerns in upcoming residential and commercial demand-response systems. *IEEE Proceedings on Power Systems*, 1(1), 2008.

[15] P. McDaniel and S. McLaughlin. Security and privacy challenges in the smart grid. *IEEE Security Privacy Magazine*, 7(3):75–77, 2009.

[16] A. R. Metke and R. L. Ekl. Security technology for smart grid networks. *IEEE Transactions on Smart Grid*, 1(1):99–107, June 2010.

[17] E. L. Quinn. Privacy and the new energy infrastructure. *Social Science Research Network (SSRN)*, Feb. 2009.

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

[19] J. Zhang and C. A. Gunter. Application-aware secure multicast for power-grid communications. In *Proceedings of First IEEE International Conference on Smart Grid Communications*, pages 339–344, Gaithersburg, Maryland, USA, Oct. 2010.