How to have the cake and eat it, too: Protecting Privacy and Energy Efficiency in the Smart Grid.

Klaus Kursawe

Institute for Computing and Information Science,
Radboud University Nijmegen
P.O. Box 9010, NL-6500 GL Nijmegen, The Netherlands

kursawe@cs.ru.nl

Abstract

The ongoing introduction of the Smart Grid is making huge promises in terms of energy efficiency and cost savings, but at the risk of consumer privacy. This risk has shown to have a significant impact, and in some countries, privacy concerns have already caused a pushback that has significantly delayed smart grid deployment as well as increased deployment costs. In this paper we describe how much of the privacy risk can be mitigated, without hampering the business case or causing significant extra costs. By using modern Privacy Enhancing Technologies, it is possible to create a win-win situation, where both the business use case and the privacy protection profit at the same time.

1 Introduction

By adding more intelligence and measurements to the electricity distribution network, the Smart Grid promises vast energy savings due to a better alignment of demand and supply. Furthermore, the increasing use of weather depended green energy sources such as photovoltaic and wind energy, and the potential addition of electronic vehicles to the distribution network, a more sophisticated system is unavoidable. One important component in the future Smart Grid will be measurements and control in the consumers homes due to Smart Meters. Those meters will continuously monitor and report energy consumption, and potentially takes some degree of control on the attached devices. To enable this technology, the European Commission wants 80% of all European households to be equipped with a Smart Meter by 2020. Simultaneously, the continuous measurement is a substantial privacy risk, as it allows deriving a large amount of personal data of the consumer. Due to this risk, the Dutch government in 2010 halted a law for compulsory smart meter installation on the ground of privacy issues, causing a huge delay in the installation and significant economical damage.

In this report, we show that privacy protection does not need to contradict energy savings and business goals. Instead, we show that by using modern privacy enhancing technologies, privacy protection can turn from a zero-sum game to a positive sum game, leading to both better privacy and higher quality measurements and thus better energy efficiency. This is possible by performing all computations on privacy critical data under encryption, where the electricity provider only obtains the result of the computation, but does not learn anything about the in-

put data. This way, it is possible to use much more fine-grained measurements, e.g., measure in smaller intervals or report the energy consumption of different device types independently.

2 A short primer on the Smart Grid

This has become especially important with the increasing switch to renewable energy sources. While classical energy generation is largely predictable and under tight control of the operators, renewable sources such as wind or solar power can be highly variable, and generate both a sudden surplus of energy and a sudden shortfall. The promise of the Smart Grid is to match the unpredictability in the supply by more control over the demand. This is done by more precise measurements of the demand to allow for a better forecast, but also by actively controlling the demand side. To this end, energy prices will be highly dynamic and flexible, and some energy intensive appliances – such as electric heaters and climate controls – may be directly controlled by the grid provider. Especially important here is 'peak shaving', i.e., reducing the maximum spikes in power consumption.

While the Smart Grid encompasses many components, the part that is most interesting from a privacy point of view s the smart metering architecture. To this end, electricity meters in consumers' homes will get some computation power, as well as the ability to communicate with the backend, with other meters in the same household (e.g., gas- and water meters), and potentially with household appliances like the HVAC system. The goal is to provide the grid operator with detailed information on the power consumption, control flexible appliances either directly or through dynamic pricing, and to help a consumer to optimize their consumption pattern.

The projections on energy and cost savings of the smart grid are immense. In 2008 study, the Electric Power Research Institute estimated savings of up to 3.7 billion kWh due to peak demand reduction alone, and a total of $35 billion in electricity costs for the United States [EPRI2008]. In addition to the economic value, the smart grid is also seen as an important contributor to CO_2 savings, which has prompted major regulators to speed up its deployment. In the EU, the goal is to have 80% of all households equipped with smart meters by 2020, while the US is substantially subsidizing the deployment due to the Energy Independence and Security Act of 2007.

3 Privacy Risks

While the major part of the smart grid does not deal with consumer data at all, one important component – the smart meter – will be in the consumers home, and measure and report data that allows deep insights into the consumer behavior. This ranges from obvious ones such as vacation time and sleeping patterns to information about health (e.g., sleep disorders), social behavior, or even religion. Furthermore, there is no good way for a consumer to temporarily opt out – while a mobile phone can be left at home when a consumer does not want to be traced, the meter will always be active (there are some proposals to include batteries in consumers' homes to allow them to mask some sensitive data, though the practicality of this approach can be questioned). In theory, the measurement at the meter allows for even greater detail – prototypes have already shown that it is possible to determine relatively precisely

which appliances are turned on at any given time, and to some extend even the activity on those appliances.

This privacy risk is not only a problem for end consumers, but has already proven to be a substantial business risk. In the Netherlands, the parliament has passed a law to speed up the smart meter rollout by making the deployment of a smart meter compulsory. While the grid operators where preparing for roll-out and making substantial investments, the Dutch Senate blocked the law on the ground that it violated the European Convention of Human Rights because of privacy reasons. This unforeseen issue stopped the roll-out completely, and cost the companies involved millions of Euros. Privacy is since playing an important role in regulatory guidelines, both in North America [NIST10] and Europe [EC11].

3.1 The Limits of Anonymisation

As a reaction to the increased privacy worries, a number of proposals have been made on how to add better privacy to the Smart Grid. The traditional way towards more privacy is to separate data that can directly identify users from the actual measurement data, and only associate the measurement with a pseudo-anonymous device identifier. However, with this approach, the data is still there, and the only protection is that the link between device identifier and the identifying data is not directly available; protecting this link is a tedious, expensive and risky task, as various data breaches in the past have show. An additional risk is that it is not entirely clear anymore what kind of data is actually personally identifiable, which has been identified as a generic limit to the anonymisazion approach [Ohm09]. As smart metering data may reveal a lot about the circumstances and the habits of a person, it cannot be excluded that a re-identification is possible based purely on the measurement data, when seemingly harmless data does contribute towards identifiable data once collected in sufficient amounts and cross linked to other, public data sources. That this kind of re-identification is possible has been shown in past studies, e.g., on Netflix move preference data [NaSh08]. In all those cases, data that was anonymised (such as movie preferences, or anonymised health data) could be de-anonymised with a surprising efficiency. It is therefore no longer possible to cleanly separate between personal identifiable data and harmless data, as each additional data item makes identification a little bit easier. Due to the wealth of data that can be derived in smart grid readings, there is a clear indication that the approach of simply separating identifiable and anonymous data is a good start, but will quickly reach its limits.

As a more concrete example, grid data may reveal that a person always stays up late when a particular TV show is on, which in return may give some demographic data. It also can be linked with some semi-public data (e.g., people who ‘like’ this show on social networks) to assist in the de-anonymisation. Additional data mining may give information about my occupation, holiday schedule, religious preference, etc, which all narrow down the anonymity.

In addition to the possibility of re-identification, another risk- and cost factor is the secure storage of critical data. As we have seen in numerous data leaks in the recent past, preventing data leakage is hard, even for well funded and reasonably competent organizations. This shows that collecting data and then attempting to protect it from unauthorized access is risky, expensive, and in many cases inefficient. We therefore argue that the best protection from such events is if the data is not there in the first place.

4 Privacy by Design

The *Privacy by Design* principle has been developed the privacy commissioner of Ontario, and is rapidly becoming a standard in the deployment of privacy sensitive systems; for the

smartgrid, the concept has been adopted by the Ontario smart grid deployment, and is recommended to be applied for European smart girds by the European Commission [EC11]. The basic concept builds on seven relatively abstract principles [Cavo11], namely, respect for users, lifecycle protection, proactive/preventive rather than reactive privacy, visibility and transparency, embedded privacy in the system design, privacy by default, and positive sum privacy. A concrete analysis and guidance on the how these principles can be used in a practical setting can be found in [GuTD11].

An important aspect to make privacy happen in real systems is the aim for 'positive sum' technologies. This means, that a privacy technology is designed with the business case in mind, and optimized to not inhibit the business case – rather, if done right, it can even be supportive and allow for better quality data and cheaper deployment costs. This contradicts the common view on privacy as being opposed to business goals, and being in the way of allowing companies to make money.

The second of the principles we will focus on in the rest of this paper is the concept data minimization; once this is deployed properly, many of the others (e.g., embedded privacy) follow implicitly. The basic idea to this end is that data cannot only be minimized by not collecting it; it can also be minimized by making only processed data visible to the grid operator, while hiding the unprocessed data items.

The concept of data minimization is a useful tool where the concept of removing personally identifiable information hits its natural limits. Rather than trying to separate personally identifiable data from the rest, the idea here is to determine which data is really needed to perform the task given, and then deploy technologies that assure that only that data is used.

Concretely, for the smart grid, the main use for smart grid related data is demand-control, i.e., getting a detailed overview on the nature of the current electricity command, and regulate it by means of price changes or direct communication with heavy electricity consumers, such as HVACs and electric vehicles.

For this business-case, the grid operator does not need to know any individual consumption data – it is enough for them to have the aggregate of the consumption of all appliances in a particular area. The primary reason why individual information is collected is that grid operators do not know how to derive the aggregates directly, and thus they collect privacy critical data they actually do not need.

More precisely, the grid operator only sees data that is aggregated over a sufficiently large set of customers, and never needs to see any of the individual data items. This aggregated data is enough for all use cases we found with the two exceptions. The first one is billing, which can be done in a much more coarse resolution, and the second one consumer awareness - for which the data can be directly transferred to the consumer or a third party acting on behalf of the consumer, without any need of the grid operator to ever see that data. There also are privacy preserving protocols that allow computations on masked data of an individual meter [RiDa10,JaJK11], which build on a similar principle as the aggregation protocols but are out of scope here.

For the other published usecases [EPRI2008] especially for the most promising ones in terms of energy savings (performing demand/response and peak shaving functionality), aggregated data is fully sufficient. Thus, with a privacy preserving aggregation protocol, most of the energy- and cost saving benefits of the Smart Grid are preserved, while largely eliminating the privacy issues.

Once such protocols are deployed, they open additional opportunities for the grid operator. For example, it is of great interest to what extend the consumption would be reduced if the electricity price changes. This is difficult to see from the normal measurements, as the energy could be consumed by a device that can react quickly on price change (e.g., a fridge or an electric heater, which can easily be tuned down) or one that cannot (e.g., a computer). As the privacy issue is critical enough on overall consumption data, there is not even a debate on making such more detailed data available to the grid operator. By aggregating data over several households without revealing the individual input data, this is suddenly no longer an issue – the grid operator can see what all fridges in a neighborhood consume together, but does not see the consumption of individual fridges. Similarly, it is no longer a problem to have a higher measurement frequency (e.g., one measurement per minute), as no privacy relevant data is transmitted anymore.

This way, the grid operator gets more detailed data where it is relevant for them, because they do not see data where they do not need it.

In the next section, we will describe how the aggregation is done concretely. Our approach does not need any additional trusted hardware; all necessary operations can be done on the meter and the backend, using existing hardware for computation and communication.

4.1 Computing under Encryption

It is known since quite some time that it is possible to allow certain computations to be performed on encrypted data without the need of decryption – that is, an untrusted party can perform the computation on encrypted input, and return the encrypted output, while never learning the concrete values. A variant of such a scheme has for example been deployed to help with double auctions for Danish farming industry [BCDG+08]. So far, this technology has mostly been seen as too expensive to use in embedded systems, and thus few applications have been tested in the real world. An excellent demonstration is shown in, Balasch et al. , where the authors have build a prototype that can use this approach for privacy friendly road tolling [BRTG+10], and have implemented their protocol on a device of comparable computation resources as real on board units for road tolling.

Thus, the data never needs to leave the meter in unprotected form – each individual data item is protected, while the grid operator can derive the results of some computations derived from the data.

Due to the high costs already associated with a smart meter rollout, it is vital that the privacy technology does nod add a substantial new cost factor. The most critical factor in there are the meters - given the large number of meters that will be deployed, every cent of extra costs will translate into millions of additional costs for the grid providers. Thus, it is vital to keep the extra code to be executed on the meter small enough to fit into existing memory, and fast enough to run on the current processors without causing a visible slowdown. Furthermore, the communication bandwidth of the meters is quite restricted, as a large number of smart meters may share a single, noisy powerline channel for communication with the backend. Finally, communication standards have already been developed (e.g., the DLMS/COSEM), and additional protocols need to fit into the framework of those standards

For our protocols, this means that meter functionality has to be absolute minimal in terms of computation (on the meter side), and the meters – once initialized - should only need to send one single (short) message, without requiring any further interaction with the backend or with

other meters. The latter is a challenge few comparable systems need to meet, and where most protocols in the literature fall short [GaJa10].

.

Furthermore, we have to consider that the bandwidth available to an individual meter may be very limited; there can be a large number of meters attached to a noisy powerline communication channel, which does exclude communication heavy protocols. In addition, in some legislations it is not allowed that meters of different households communicate with each other, as that is seen as a potential vector for meter based malware. Ideally, after the system setup, meters would only unicast messages to the backend, and not be involved in any further computation.

4.2 The Data Aggregation protocols DiPA and LoPA

In this section, we outline two concrete protocols for privacy protecting data aggregation. For the scope of this work, we only give an intuition on the protocol mechanism and a basic outline; the more detailed protocols as well as the protocols for efficient key initialization can be found in [KDK11].

The first protocol, DiPA (Diffie-Hellman based Private Aggregation) is a simple cryptographic protocol based on the Diffie-Hellman public key scheme. The main mechanism here is a *homomorphic commitment* scheme, which can be implemented very efficiently using Diffie-Hellman on elliptic curves.

A commitment scheme is a simpler tool than an encryption scheme. It allows a user to fix some secret (*commit* to it), and to later reveal the secret and prove that this was the value she committed to. As opposed to an encryption, a commitment scheme is easier to implement, and it does not need a secret decryption key, which means that there is less key management required.

As visualization, one can think of the original value as a Lego-car, the commitment scheme as a rubber hammer, and the commitment as a heap of Lego-bricks. Committing to a value (car) means to smash it with the hammer, and showing the heap of stones to the verifier. It is now computationally hard for the verifier to reconstruct the car from the heap, while it is easy to verify that a given car corresponds to the given heap.

The special property of *homomorphic commitments* is that it is possible to perform computations on the commitments, which then correspond to computations on the original plaintext, i.e.,

$$\text{Commit}(A+B) = \text{Commit}(A) * \text{Commit}(B)$$

In our visualization, this means that two Lego cars can be added up to a transformer (which is the addition on the original value side). Similarly, if one adds the heaps generated by the two individual cars, one gets the heap generated by the transformer, so the addition on the original values has an equivalent operation on the commitments.

These commitments form is the basis of the aggregation protocol. Because of the homomorphism, we can sum up commitments various parties to a commitment of the aggregate.

Public parameters:

G: Group of prime order p

H: {0,1}* → G : Hash function (Random Oracle)

Private Meter Keys

$r_i \in G$: random values

Aggregator Key(s)

$r_A = \Sigma r_i$

Meter side:

For measurement m_{ik} iwith index k, meter i :

computes g=H(j)

sends $g^{r_i+m_{ik}}$ to the aggregator

Aggregator:

Compute $\Pi g^{r_i+m_{ik}} = g^{\Sigma(r_i+m_{ik})} = g^{\Sigma r_i} g^{\Sigma m_k}$

Using knowledge of Σr_i, compute $g^{\Sigma m_{ik}}$

Compare if this fits with the expected measurement

Figure 1:DiPA (Diffie Hellman Private Aggregation) Protocol

While this protocol only allows us to compare values we already know, it is easy to transform into a protocol that computes actual aggregates. To this end, we simply perform the standard aggregation protocol on individual bytes, and then brute force those on the backend server (given the small domain of measurement values, this should not be more than a few hundred tests, which a modern PC can easily handle).

A main advantage of this protocol is that it allows for different sets of meters to be aggregated on, without requiring any change to the meter configuration. Instead, the aggregator needs to know the sum of the corresponding masking values. This would allow, for example, to separately aggregate over all meters in one particular district, as well as over all meters of consumers that also generate energy.

The LoPA (Low overhead Private Aggregation protocol) is even simpler, but does sacrifice some flexibility for this simplicity. In this protocol, the group of meters whose measurements are aggregated is fixed, and all meters in one group know of each other. Each two meters in one aggregation group share a common secret x (i.e., in a group of ten meters, each meter needs to keep nine such secrets). When a measurement is to be protected, one meter adds its corresponding x for all its peers to its output value, while the other one subtracts it. Thus, the overall effect of the secrets cancels out completely, and an aggregator summing up all values gets the exact sum of all measurements. However, if only one measurement is missing, not all the secrets cancel out, and the reading is unreadable. To protect privacy over several readings, the secrets need to be changed after each reading; this can easily be done without any interaction and little computational overhead by applying a hash function such as SHA-256.

Public parameters:

H: {0,1}* → {0,1}* : Hash function

Private Meter Keys

r_{ij} : Pairwise shared key between meters i and j

Aggregator Key(s)

--

Meter side:

For measurement with index k and value m_{ki} in meter i :

computes $R_{ki} = \Sigma\ s(i,j) * H(k,r_{ij})$, where s(i,j) = -1 if i<j, 1 otherwise

send $R_{ki} + m_{ki}$ to the aggregator

Aggregator:

Compute $\Sigma\ R_{ki} + m_{ki} = \Sigma\ m_{ki}$

Figure 2: LoPA (Low overhead Private Aggregation) protocol

One major advantage of this approach is that there is no public key cryptography involved once the system is initialized, and all operations are simple additions as well as a hash function. This not only reduces the computation overhead to the absolute minimum, but also allows the message size to stay exactly the same – a masked 32-bit value still is a 32-bit value, as opposed to the homomorphic commitment based protocol, where it needs to be long enough to be cryptographically secure. Thus, this protocol integrates very neatly into the existing DLMS/COSEM standard, and no changes have to be made to the message format.

The price is a somewhat a smaller flexibility – in this protocol the aggregation group is fixed by the keys the meters have, and it is not possible to aggregate over different sets of meters simultaneously. Also, a meter does need enough memory to store all the shared keys with its peers. This does not have a large impact in practice, however, as the sets of meters should be kept small anyhow for stability reasons.

5 Implementation & Practical Issues

To validate the concept, both protocols presented above have been prototyped by Elster SG on their production meters, together with the billing protocol presented in [RiDa10], and the application scenarios have been discussed intensively with a Dutch grid operator. The low overhead protocol has been fully integrated into the existing protocol stack. In the end, the meters could successfully report masked measurements to a fourth meter, which was used to compute the aggregate. The cryptographic protocol has been implemented to verify the performance impact, but not yet integrated into the communication stack due to the change in the message formats this would have required. For both protocols, the computation overhead for the masking was essentially instant (i.e., far below one second), and thus easily within the scope of 15 minute measurement intervals.

One additional issue that comes in the nature of the aggregation protocols is that it is impossible to compute a sensible aggregate if a single measurement is missing. This is unavoidable – if it were possible to compute such an aggregate, the aggregator could easily compute the difference of an aggregate with all meters and with all meters bar one, and then derive the measurement of that meter by comparing the aggregates. In practice, however, that means that a single failed meter brings down its entire aggregation group.

For some use cases, this is entirely tolerable. For example, if the protocols are used for fraud detection, the failure of a meter is exactly the event we want to detect, and the protocols will detect it. For load balancing and demand response, however, this is a different issue.

The most practical solution for this problem is to group meters in small groups of ca 20 meters, and then computes the larger aggregate by summing up the aggregates of those groups. A single failed meter will still bring down 19 of its peers – however, on the larger scale of a demand response system, the loss of 20 meters is still tolerable (one should also note that meters are not overly likely to fail) as long as it is detected and the output is not used in further computations. This approach also makes it easier to accommodate the more efficient low overhead protocol – once there are only small groups of meters, the limited scalability of this protocol is no longer an issue.

6 Conclusions

In this paper, we have shown a privacy technology for data aggregation in a smart metering setting, which allows a grid provider (or any other authorized party) to compute aggregates of measurement data without needing access to the individual – privacy critical - measurements themselves.

The main message is twofold. Firstly, modern privacy enhancing technologies have reached a level of practicability that does allow them to work in a real system – and while larger scale tests still have to be done before a real deployment is possible, the implementation already demonstrates that an integration into existing architectures and hardware is feasible. Secondly, we show a practical example of 'positive-sum' privacy, i.e., a privacy technology that has been developed together with the businesses, and that does fit into the overall business model and its requirements. In doing so, the technology even can generate positive value for the business – not only by helping to comply to regulation and saving costs on otherwise needed technology, but by allowing to have more privacy *and* actually use more data.

7 References

[BRTG+10] Josep Balasch, Alfredo Rial, Carmela Troncoso, Christophe Geuens, Bart Preneel and Ingrid Verbauwhede. PrETP: Privacy-Preserving Electronic Toll pricing. In 19th USENIX Security Symposium, pp 63-78, 2010.

[Cavo11]: Cavouican, Ann: Privacy by Design: The 7 foundational principles, 2011

[BDDG+08]: Peter Bogetoft, Dan Lund Christensen{, Ivan Damgard,Martin Geislerz, Thomas Jakobsen,Mikkel Kroigaard, Janus Dam Nielsen, Jesper Buus Nielsen,, Kurt Nielsen,, Jakob Pagter, Michael Schwartzbach, and Tomas Toft: Secure Multiparty Computation Goes Live, 2008

[EC11] Smart Grids: From innovation to deployment. Communication from the Commission, 2011

[EPRI08]: The Green Grid: Energy Savings and Carbon Emissions Reduction Enabled by a Smart Grid. EPRI Report no 1016905, 2008.

[GaJa1] Flavio Garcia, and Bart Jacobs: Privacy Friendly Energy Metering via Homomorphic Encryption. Workshop on Security and Trust Management (STM 2010) Lecture Notes in Computer Science, Vol. 6710, pp 226-238, 2011

[GoJu04] Phillipe Golle, Ari Jules: Dining Cryptographers Revisited. Eurocrypt 2004 ,pp 456-473, 2004.

[GuTD11] Seda Gürses, Carmella Troncoso, and Claudia Diaz: *Engineering Privacy by Design*. In Conference on Computers, Privacy & Data Protection (CPDP), 2011.

[JaJK11] Marek Jawurek, Martin Johns, Florian Kerschbaum: Plug-in privacy for Smart Metering Billing. Privacy Enhancing Technologies Symposium (PETs),pp 192-210, 2011.

[KDK11] Klaus Kursawe, George Danezis and Markulf Kohlweis: *Privacy Friendly Aggregation for the Smart Grid*. Privacy Enhancing Technologies Symposium (PETs 2011), pp 175-191, 2011

[NaSh08] Arvind Narayana, Vitaly Shmatikov: Robust De-anonymisation of Large Sparse Datasets. IEEE Symposium on Security and Privacy,Oakland, 2008

[NIST10] NISTir 7628: Guidelines for Smart Grid Cyber Security, 2010

[Ohm09] Paul Ohm: Broken Promises of Privacy: Responding to the Surprising Failure of Anonymizaion. 57 UCLA Law Review, pp. 1701 – 1778, 2010

[RiDa10] Alfredo Rial & George Danezis: Privacy-friendly smart metering.. Microsoft Research Technical Report MSR-TR-2010-150, 2010.

Index

Smart Grid, Privacy, Privacy by Design, Aggregation, Homomorphic Encryption, Implementation