PRIVACY BY DESIGN FOR LOCAL ENERGY COMMUNITIES

Pol Van Aubel
Radboud University – Netherlands
pol.vanaubel@cs.ru.nl

Michael Colesky
Radboud University – Netherlands
mrc@cs.ru.nl

Jaap-Henk Hoepman
Radboud University – Netherlands
jhh@cs.ru.nl

Erik Poll
Radboud University – Netherlands
erikpoll@cs.ru.nl

Carlos Montes Portela
Enexis – Netherlands
carlos.montes.portela@enexis.nl

ABSTRACT

Electricity generation is becoming more and more decentralized. This trend requires management of the electricity grid at a more granular level by Distribution System Operators than before. To this end, local energy communities (LECs) are being piloted throughout the world. However, this granular control requires the use of more privacy-sensitive information than we are used to. In particular, this entails more frequent meter readings, over short intervals of 15 minutes or less. Therefore, privacy should be taken into account in the design of these LECs. In addition, they should comply with data protection laws. This paper reports on our experience with applying the Privacy by Design approach in GridFlex, an LEC pilot project. We describe and reflect on the process followed and summarize the results. We hope that this proves useful for other projects facing these issues.

INTRODUCTION

With renewable energy generation (solar power in particular) the classical view of the electricity grid, where energy flows from a few large production facilities out to the consumers, has become obsolete. Generation is now also decentralized. With solar panels in use, consumers might switch to being producers and back to being consumers multiple times per day, or even per hour. Distribution System Operators (DSOs) are responsible for ensuring that there is sufficient connection capacity to service each consumer – both for energy demand, but also for supplying energy back to the grid. Demand shaping, i.e. influencing demand to match the supply and capacity, is one technique to reduce the amount of copper needed – and hence the costs – to meet peaks in demand.

Different ways of demand shaping are being explored by DSOs. One option is to provide incentives to consume electricity at exactly those moments when there is locally generated power in abundance. Another option is in-home batteries. These provide a reliable form of power storage that can be charged during the day with the solar power being generated, then used during peak time or the night. Local energy communities, which we abbreviate to LECs, provide opportunities to test these techniques.

Both options require an accurate, fine-grained model of the grid. The information used for this can be privacy-sensitive [1,2]. Additionally, LEC projects often include some form of feedback to the consumer, or may try to stimulate collaboration between consumers. This also involves privacy-sensitive information.

To ensure that consumer privacy is adequately taken into account, it should be part of the design process. Privacy by Design is one approach to this [3]. Privacy by Design and Privacy by Default are mandatory since the European General Data Protection Regulation (GDPR) has come into force [4]. The GDPR uses the term data protection instead of privacy; though for brevity and clarity, we use privacy.

In this paper we explain what applying Privacy by Design means in general and for LECs in particular. We describe the approach we took, in a collaboration between Radboud University and GridFlex Heeten, to apply Privacy by Design in a Dutch LEC project. We highlight the issues encountered that are likely to be present in other LEC projects. We also show the ways in which we mitigate privacy concerns and explain the rationale behind these choices.

GridFlex Heeten

GridFlex Heeten (http://gridflex.nl/) is a pilot project in the Dutch village of Heeten to explore market and control models for an LEC. Dutch DSO Enexis is one of the project partners. The goal is to experiment with price incentives to optimally match local energy consumption, storage, and production. The project is centred on a single neighbourhood. Each household is equipped with solar panels, in-home batteries, or both. In addition, a large solar station is present nearby.

The project is both a research- and a production-project. The first four years are mainly for research into price incentives and grid control. After these four years, the research-project will end. However, it is likely that the project will continue as a controlled LEC afterwards. Fine-grained metering data will be collected and stored by the project partners running the project infra-structure. Additionally, a research database based on this data will be shared with the University of Twente.

Measurements from smart electricity meters will be used to monitor and manage the LEC and to build the models for studying incentivization. Smart electricity meters in the Netherlands can send this data directly to the DSO in intervals no shorter than 15 minutes [5]. However, for effective monitoring and analysis, more granular, one-minute-interval readings are necessary.

The meters can provide this data only on a local interface, requiring equipment connected directly to the smart meter to send this data to the project partners. Collecting measurements at this high level of detail implies several privacy issues, which we discuss below.

In order to build accurate models, household composition is used to characterize the consumption profile of a connection. This information is also privacy-sensitive.

PRIVACY BY DESIGN

Privacy by Design (PbD) [3,6] is a design and engineering approach intended to ensure privacy protection from the earliest stages of a project, not just in hindsight. The idea is that privacy concerns are considered throughout the entire project lifecycle, from the earliest concept formulation, to design process, implementation, deployment, and, if applicable, decommissioning. By considering privacy from the beginning, costs and complexity of redesign when privacy issues are discovered can be largely avoided.

The GDPR makes application of PbD mandatory [4]. PbD has consequences such as forbidding data processing that is disproportionally invasive, and requiring allocation of resources towards ensuring consumer privacy. The GDPR also requires Privacy by Default, meaning that the strictest privacy settings should be the default.

PbD is a somewhat vague concept. To make its underlying goals more concrete, more specific privacy design strategies have been proposed [7]:

1. **minimize**: only collect that data which is strictly necessary, and remove that which no longer is.
2. **hide**: encrypt, pseudonymize, and take other measures that protect and obscure links between elements of data and their source.
3. **abstract**: reduce the granularity of data collected; combine or aggregate data from multiple sources so that the sources are no longer uniquely identifiable.
4. **separate**: store and access data only where it is used; process data at the source instead of centrally.
5. **inform**: explain to data subjects how their personal data is processed, and how profiles and automated decision-making based on their personal data work. A subject can only provide valid consent to data processing if they understand how their data is being processed.
6. **control**: allow data subjects to provide and revoke consent to process, and to access, correct, and delete their provided and derived data.
7. **enforce**: build technical and organizational measures that ensure the design decisions taken with regard to privacy are actually implemented, and log the actions of the systems.
8. **demonstrate**: document, audit, and report on the operational and PbD processes.

The first four strategies are more focused on data. The last four are about policies and the surrounding processes. Given these strategies, the PbD process could then ideally be implemented as follows: look at each project requirement, figure out what potential privacy impacts it has, and apply strategies to mitigate those impacts. This should be an iterative process, which is repeated as the design becomes more detailed or changes in other ways. The first step in each iteration involves performing a Privacy Impact Assessment (PIA) [8], or rather, refining the assessment from the previous iteration. Unfortunately, standardization of the PbD process in general is still lacking and the subject of further research [9].

Still, in absence of such standardization, the PbD process can take the form of several meetings between project stakeholders where for each project requirement the privacy impact on the end-user is estimated, and all these strategies are considered. For this to be effective, people who possess sufficient experience and domain knowledge to deduce privacy issues from project (data) requirements must be present. The outcome is ideally twofold: a set of design documents stipulating in detail the measures that must be taken in implementing the project design, and a keen awareness of the privacy considerations among project architects and developers.

In the case of GridFlex, however, we became involved after the architecture had, for the most part, already been designed. Therefore, our approach was a retroactive one, applying strategies with the intent to redesign the architecture where possible [10]. We held three meetings, each half a day long. Business architects from all project partners were present at these meetings, not just from the partners developing the soft- and hardware. The reason for this is that design requirements are made for business reasons. In order to determine appropriate implementations of strategies, it is essential for the process to have a representative present who is able to explain and discuss those business reasons. Lead software developers were not present during these meetings. In retrospect, we believe that they should also have been included in this process. The project partners feel that this would improve engagement and awareness of the software developers, which privacy requirements in documentation may not be able to achieve.

The outcome of this process was a set of applied strategies in a document, with rationale included. The next section summarizes these. Additionally, a PbD manual, for use by the project partners, is being developed.

PRIVACY ISSUES IN GRIDFLEX

This section summarizes the main privacy issues in the GridFlex project, and the design choices that were made as part of the PbD process.

Fine-grained metering data

For grid control in an LEC such as GridFlex, real-time, fine-grained measurements are required. The measurements of household connections are used because in GridFlex, smart grid equipment is managed per household. The granularity here is a problem, because it provides a detailed insight into the personal lives of the members of a household.

To put this into perspective: the initial smart meter roll-out in the Netherlands was postponed because the original law proposed that the DSOs take 15-minute-interval measurements by default. This was found to be in violation of article 8 of the European Convention on Human Rights [11,12].

In GridFlex, the granularity of measurements has been chosen to be even smaller. It is currently one minute, and may be reduced further during the project. The goal is to determine whether these measurements provide advantages in running an LEC, and whether those advantages outweigh the additional cost of privacy-protecting measures. During the research phase of the project, the metering data will also be used as basis for a research database by the University of Twente.

For this privacy problem, the following strategies were deemed most effective on the data level: **Abstract** is applied by taking the measurements on a minute-granularity, rather than the 10 seconds that the interface allows. **Minimize** takes the form of only collecting the energy usage, not other information available on the interface. **Hide** is then applied by pseudonymizing the data before transmission to the central system and by encrypting the data in transit. **Separate** is applied by limiting access to the data to those parties that strictly require it: the project partner doing actual grid control, and the University of Twente for research purposes.

On the process level, we **inform** the data subject via informational meetings before they sign up, and in the customer's project contract when signing up for the project. **Control** is provided through a customer portal.

Once the research phase ends, the data collection for the university also ends, but the infrastructure will remain and will probably become part of the normal grid. This infrastructure will still have additional control capabilities, and if deemed successful, the project will likely enter its production phase. If the fine-grained metering data is still required at that point, the architecture should be reviewed and changed to not store metering data any longer than is necessary for the actual grid control decisions.

Customer identity and location information

Another class of personal data processed includes customer names and addresses, which are needed when the project partners need to contact the consumer. Storage of this data by all project partners, and in relation to the energy usage information, however, is not needed and (hence) not acceptable.

Instead, **separate** is applied by having a single project partner store a database linking customer name and address to a pseudonymized identifier. This project partner does not need access to the measurement data, so concerns are kept separate. For partners that store information that may need to be relinkable to the customer name and address, we follow the **hide** strategy and use the pseudonymized identifier instead. For example, the energy measurements might indicate a need to perform on-site maintenance by an engineer. The party determining this could simply contact the project partner which holds the linking database, and tell them that an engineer needs to visit the household linked to that pseudonym. There is then no need for that partner to learn the customer's identity.

Household composition

In GridFlex, the goal is not just testing an LEC, but also creating standard profiles for prediction and planning. To this end, users are asked to provide information on the composition of their household, which is needed to understand different usage patterns to base the profiles on. However, there are several ways to go about this, and similar to the previous point, it is not needed and (hence) not acceptable to store this information with all project partners.

The impact is mitigated by several strategies. First, **hide** is applied by storing household data with the measurement data under a pseudonymous identifier, rather than with the customer identity and location information. Second, **abstract** is applied by only characterizing the household, rather than using the precise composition or even identities of people in the household. Third, **separate** is applied by only storing the household composition in the research database, and not at each project partner. Finally, because this data is only required for research purposes, it suffices to sample part of the project participants. Therefore, **minimize** can be applied by making household composition optional: customer opt-in is requested, and participation in the project is still possible if the customer prefers not to provide this data. Of course it may impede the accuracy of the established profiles if many households choose to withhold this data, but this trade-off was deemed acceptable. Finally, this data will be destroyed after the research phase is finished.

Customer opt-in is only valid if the customer is accurately and understandably informed. Therefore, we **inform** the customers through informational meetings where the project is explained and where they can ask questions about the data processing. In addition, the project's contract with the customer will have explanatory text accompanying the opt-in form.

CONCLUSIONS

Even though we were not involved from the start of the design phase, the Privacy Design sessions that were held in GridFlex proved useful, considering that several measures were taken to significantly reduce potential privacy impacts. Examples include storing data only with the parties that actually need it, separating the data from direct identifiers such as name and address through the use of pseudonymization, and minimizing the data collected, as explained in the previous section. Ideally, however, Privacy by Design should have been applied from the beginning of the project.

Although Privacy by Design as a concept is becoming well-known, it turns out that there is not much standardization in how to actually apply it [9]. One document of interest to local energy communities is the standardized Data Protection Impact Assessment template for Smart Grids [13,14].

In retrospect, for the effectiveness of Privacy by Design, we believe both project architects *and* the lead software developers should be included in Privacy by Design meetings. This ensures engagement and awareness amongst the people implementing the decisions, which may not be achieved by only communicating privacy requirements through documentation.

One thing we noticed at the meetings is that pseudo-nymization and anonymization are still often confused. It is important to realize that an identifier is still an identifier, whether that is a full name or a random number. So pseudonymization, where we replace full names by some random number, does not necessarily provide anonymization. Even though these numbers look more anonymous than the names, it may well be possible to reconstruct the associated name [15,16].

Local energy communities may become the new standard for managing the electricity grid. If that happens, opting out may be difficult due to social or political pressure. This makes it even more important to adequately protect consumer privacy. Privacy by Design provides a structured approach towards achieving this. We have shown how this can be applied to a local energy community pilot, which issues such a pilot is likely to encounter, and how they can be mitigated.

A major design decision in the electricity grid of the future is the trade-off between usage of smart grid technology and concepts, and usage of additional grid infrastructure. This is beyond the scope of individual pilot projects such as GridFlex, but a broader discussion seems warranted. Putting more copper in the ground would accommodate higher peak demand. More advanced measures – including more IT – may manage supply and demand in an effort to reduce peak demand. Lower peak demand reduces the amount of copper needed. Copper is expensive, but so is rolling out and securing a complex IT infrastructure. The optimal trade-off may not be easy to determine. This trade-off will also have an impact on privacy, so copper could effectively act as a privacy-enhancing technology.

REFERENCES

[1] P. McDaniel and S. McLaughlin, 2009, "Security and privacy challenges in the smart grid", *IEEE S&P* vol. 7, no. 3, 75–77.
[2] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin, 2010, "Private memoirs of a smart meter", *Procs. SenSys'10*, ACM, 61–66.
[3] A. Cavoukian et al., 2009, *Privacy by design: The 7 foundational principles*, IPC of Ontario, Canada.
[4] European Commission (EC), 2016, *General Data Protection Regulation*.
[5] Netbeheer Nederland, 2014, *Dutch Smart Meter Requirements version 4.0.7 – Main Document*.
[6] A. Cavoukian, 2012, *Operationalizing privacy by design: A guide to implementing strong privacy practices*, IPC of Ontario, Canada.
[7] M. Colesky, J.-H. Hoepman, and C. Hillen, 2016, "A critical analysis of privacy design strategies", *Procs. IWPE'16*, IEEE, 33–40.
[8] D. Wright, 2012, "The state of the art in privacy impact assessment", *CLSR*, vol. 28, no. 1, 54–61.
[9] J. van Puijenbroek and J.-H. Hoepman, 2017, "Privacy impact assessments in practice: Outcome of a descriptive field research in the Netherlands", *Procs. IWPE'17*, IEEE, to appear.
[10] A. Cavoukian and M. Prosch, 2011, *Privacy by ReDesign: Building a Better Legacy*, IPC of Ontario, Canada.
[11] C. Cuijpers and B.-J. Koops, 2008, *Het wetsvoorstel 'slimme meters': een privacytoets op basis van art. 8 EVRM*, Tilburg University, the Netherlands.
[12] C. Cuijpers and B.-J. Koops, 2013, "Smart metering and privacy in Europe: Lessons from the Dutch case", *in European data protection: coming of age*, Springer, 269–293.
[13] Smart Grid Task Force – Expert Group 2, 2014, *Data protection impact assessment template for smart grid and smart metering system*, EC.
[14] A29 WP, 2013, *Opinion 04/2013 on the data protection impact assessment template for smart grid and smart metering system*, EC.
[15] M. Barbaro and T. Zeller Jr., 2006, *A face is exposed for AOL searcher no. 4417749*, New York Times.
[16] A. Narayanan and V. Shmatikov, 2006, "How to break anonymity of the Netflix prize dataset", *CoRR*.