Linked Data for a privacy-aware Smart Grid*

Andreas Wagner, Sebastian Speiser, Oliver Raabe, Andreas Harth
Karlsruhe Institute of Technology
{first.lastname}@kit.edu

Abstract: Recent developments around the Smart Grid promise more efficient power generation and distribution. In contrast to the current design of the electricity grid, the Smart Grid is heavily based on IT for managing communication flows. Given the large number of market participants, data exchanged should be marked up in a way that facilitates flexible modelling, extension and integration. In addition, since much of the data exchanged is highly sensitive – comprising e.g. consumption data – current organisational means for privacy enforcement are not sufficient. In this paper, we propose a novel framework, which addresses these privacy issues by means of a machine-interpretable representation of the user intent. More precisely, such machine-interpretable specifications enable an automated access control and enforcement of privacy rights on a technical level.

1 Introduction

The Smart Grid is expected to strongly increase energy efficiency, foster the usage of renewable energy sources and reduce greenhouse gas emissions [ETP06, Ger10, NIS10]. In order to do so, Smart Grids comprise a flow of information, enabling novel functionalities (e.g. prediction of future consumption). However, as more data is circulating within the Smart Grid, new challenges arise, most notably privacy related issues. The published data is in its nature highly sensitive, as it is personal information and thus must be handled in compliance with existing regulations, e.g. privacy laws [CPW10, MM09, Raa09, Raa10]. Traditional means, i.e. organisational control, for ensuring privacy are not sufficient, due to the high frequency and volume of the data transactions [Raa09, Raa10]. The Smart Grid requires a semantic data model, in particular w.r.t. a technical enforcement of privacy [Raa10]. However, currently proposed standards are either based on restricted technologies (such as EDIFACT[1]) or complex service-oriented architectures based on XML[2], both with no formal semantics associated. In contrast, we suggest a model based on *Semantic Web* technologies. More precisely, we employ *Linked Data* principles[3] using the *Resource Description Framework* (RDF) [Kly04], in combination with HTTP as a transfer protocol. In particular, the self-describing RDF facilitates the specification of user policies, and enables a framework for a privacy-aware grid. Thus, our main contribution is two-fold: i.) we show how Semantic Web technologies enable a privacy-aware framework for the Smart Grid and ii.) we evaluate the proposed model by means of a comparison with the

*This work was in part supported by the German Federal Ministry of Economics and Technology (MeRegioMobil, Grant 01ME09005).

[1] http://www.unece.org/trade/untdid/

[2] http://www.w3.org/TR/soap/

[3] http://www.w3.org/DesignIssues/LinkedData

basic privacy principles. The rest of the paper is structured as follows: we first outline an example scenario, illustrating our data model in Section 2. Section 3 introduces our generic policy model and details the data access. We evaluate our approach according to legal considerations in Section 4, cover related work in Section 5 and conclude with Section 6.

2 Linked Data in the Smart Grid

We outline an Smart Grid scenario showing how Linked Data may provide a suitable communication model for the grid. We distinguish between data associated with *legal consequences* (e.g. billing), which is handled by a trusted party, e.g. a smart meter or a metering provider, (*obligatory data*), and all other (*non-obligatory*) data. In the latter case, data may be managed by a smart meter or a device itself. The distinction is important, as for obligatory data legal regulations enforce publishing, plus guarantee data availability.

Furthermore, we use namespaces as follows: `ex` refers to a server providing a description for each actor in the customer domain, `sm` covers smart meter data, `washer` describes washing machine data, `gov` contains definitions from a trusted source (e.g. the government), `sg` covers general schema vocabulary for the Smart Grid (i.e. properties and classes), `xs` refers to the XML Schema vocabulary and `ical` covers a temporal vocabulary. Now, consider a user *Mary* (`ex:mary`), who lives at a premise (`ex:apt`) with a smart meter (`ex:sm`). She owns a *CoolWash* washing machine (`ex:coolWash`) and an *UltraAmp 760e* electric vehicle (`ex:uamp760e`). The scenario is illustrated in Fig. 1. The manufacturer of the *CoolWash* machine requests data from Mary's washing machine for after-sales services, a metering system provider requests power consumption data for billing, and an energy optimisation consultancy requests all energy-related data, in order to help Mary optimise her consumption.

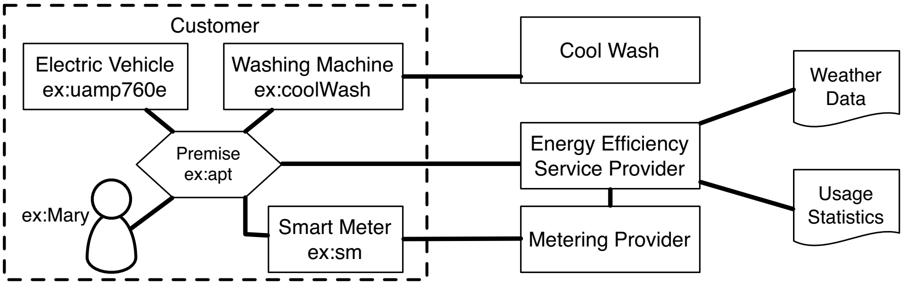

Figure 1: Data access scenario

Performing a HTTP request on the URI of a resource (e.g. a person, a smart meter, an appliance or a metering system) returns data describing the resource. E.g. a request on `ex:coolWash` would return:

```
ex:coolWash
  rdf:type sg:Appliance;
  sg:manufacturer <http://coolWash.com/company>;
  sg:owner ex:mary;
  sg:washingData washer:program40;
```

```
    sg:consumption sm:data20100310.
```

Performing a HTTP lookup on `sm:data20100310` (`washer:program40`) results in a data snippet, indicating a consumption of 1.04 kWh during a late-night wash (a program description and its associated count):

```
// lookup on sm:data20100310; data resides at smart meter
sm:data20100310
  rdf:type sg:Consumption;
  rdf:value "1.04"^^sg:kWh;
  ical:dtstart "2010-03-10T00:00:00";
  ical:dtend   "2010-03-10T01:00:00".

// lookup on washer:program40; data resides at washing machine
washer:program40
  rdf:type sg:WashingData;
  foaf:name "Program 40 C";
  sg:count "23"^^xs:int.
```

In contrast to on-premise appliances, the *UltraAmp 760e* is mobile. We assume a TCP/IP connection to the car (e.g. via 3G), so requests to `ex:uamp760e` may be performed as well. A lookup on `ex:uamp760e` would provide its model description and current location. Note that a request targets a device directly, rather than requiring a central data host. Also, we assume that access to the metering system is done via an encrypted channel (e.g. `https`), and recording of obligatory data adheres to legal requirements.

3 Policy Model

Policy Definition Our policy model gives users control over their data by enabling them to specify their intents in a machine-interpretable manner and thus allowing them to restrict or permit data access. More precisely, a `Policy` models a timespan during which it is valid via `ical:dtstart` and `ical:dtend`. A `Policy` further `allows` a number of `Usages`. An allowed usage is restricted to a specific `purpose` and to a `recipient`. Since allowing access to some data does not imply access to all data, we propose that policies contain data *perspectives*. A perspective is specified using SPARQL[4] queries. Below, find a policy, allowing CoolWash (`http://coolWash.com/company`) access to Mary's washing machine for after-sales service purposes, while omitting all consumption data.

```
washer:coolWashPol rdf:type sg:Policy;
   ical:dtstart "2010-01-01T00:00:00"^^xs:dateTime;
   ical:dtend   "2010-12-31T23:59:59"^^xs:dateTime;
   sg:allows #coolWashUse.

#coolWashUse rdf:type sg:Usage;
   sg:purpose gov:Purpose#service;
   sg:recipient <http://coolWash.com/company>;
   sg:perspective #coolWashPerspective.
```

[4] `http://www.w3.org/TR/rdf-sparql-query/`

```
#coolWashPerspective rdf:type sg:Perspective;
   sg:definition "PREFIX ... CONSTRUCT { ?s ?p ?o }
     WHERE { ?s rdf:type sg:Appliance .
        ?s sg:manufacturer <http://coolWash.com/company>.
        ?s ?p ?o .
        FILTER (?p != sg:consumption) }".
```

Our approach focuses on the expression of a user intent (and matching the intent with an incoming request), but could be extended to enforce access control by verifying the requestor authentication. The following two works illustrate how to realise such checks within the Semantic Web: First, the work by Weitzner et al., who require that a requestor presents a logical proof for his authorisation, based on axioms from trusted sources [WHBLC05]. Second, the approach by Agarwal and Sprick employs decentralised issuing of certificates, stating the possession of credentials [AS05].
Please note, purpose (`gov:purpose#service`) and recipient (`http://coolWash.com/company`) are externally defined resources. Using concepts defined by a (trusted) third-party organisation, reliable and trusted definitions may easily be integrated and reused all-over the grid. Such a solution is similar to the Creative Commons approach, where e.g. a non-commercial clause is defined that can be referenced by different licenses.
Lastly, not only the user intent, but also laws and other regulations may be modelled by means of our policy approach. These *law-based* policies may be similarly integrated within the system, thereby providing additional means for privacy enforcement.

Policy-Aware Data Access According to our model, a policy-aware data access involves the following steps: i.) the requestor performs a HTTP lookup on a URI (e.g. `ex:uamp760e`), ii.) the web server returns an *authorisation required* response, iii.) the requestor sends a request, i.e. a specification of identity and purpose, and iv.) the device matches the request with an applicable policy (either a law-based or a user policy). If both match, the requested data and (signed) policy is sent.[5]
The matching procedure is implemented as a rule, checking whether i.) the requestor is subsumed by the recipient description and ii.) the requested purpose is subsumed by the allowed purpose (both w.r.t. the applicable policy). We assume that the purpose and the recipient/requestor description link to the same trusted (hierarchical) definition. Thus, a *subclass-of* or *same-as* check is sufficient for realising the subsume-operation.

4 Evaluation

Below, we briefly evaluate how the basic privacy principles may be enforced in our approach. Due to space reasons, we will focus on the *data economy*, *purpose limitation*, *transparency* and *information security* principle.
Let us first consider *data economy*, i.e. the principle of using as little personal information as possible. Thus, an ideal system w.r.t. *data economy* would employ an anonymisation directly at the data source (§ 3 III 6 BDSG). Such a solution could be applied for a global-statistical analysis. However, early anonymisation is not possible in general, as consumption data is also required for billing purposes and hence personal information is necessary. Thus, anonymisation of consumption data by means of pseudonyms, would satisfy the *data economy* principle, while allowing a regular billing process. The *prin-*

[5] Note, the signing is optional, however, it may be necessary for proofing privacy infringements later on.

ciple of purpose limitation specifies that information has to be used in accordance with the purpose it was originally published for. Our approach supports purpose limitation, as requested data is always released together with a policy describing the intended purpose. Assuming the integrity of the policy, there is no mechanism to modify the original purpose later on. Additionally, one can implement automatic checks for purpose modifications and legitimate usage at each host, thereby enabling a technical enforcement. Next, consider the *transparency* principle, i.e. data may only be used, if the affected person is informed about the usage details. Transparency is fully integrated within our solution, as we assume that for each task the data is requested at its source. With each lookup, the user is notified about the request, its purpose and the recipient. Furthermore, a technical enforcement can be implemented, as data and policies are represented in a machine-interpretable manner. Lastly, according to the *information security* principle, data privacy has to be enforced by technical and organisational means. Note, currently organisational procedures (in compliance with § 9 BDSG) are regarded as sufficient. Our model, however, enables mechanisms for technical access control and thereby avoids the weaknesses of organisational privacy enforcement, such as missing legal expertise or deliberate exploitation of the system. In addition, we allow a direct representation of the user intent and its technical enforcement. We conclude that by means of pseudonymised data and technical procedures for access control (which implement privacy rules), the outlined privacy principles could be enforced best. The proposed policy approach provides techniques for articulation of machine-interpretable user intents and thereby enables an integration of privacy aspects at a technical level. Our solution lacks, however, means for anonymisation via pseudonyms, which should be addressed in future work.

5 Related Work

There have been various proposals for a communication infrastructure for the Smart Grid [Ger10, NIS10]. In these works, however, privacy issues were solely addressed as a pointer for future work. On the other hand, there is also work addressing these dangers in more detail [CPW10, MM09, Raa10]. Some of this work [MM09, Raa10] outlines means for privacy enforcement, i.e. organisational means and legislative measures, in particular an adjustment of current regulations. However, we aim at a technical solution, i.e. an automated enforcement of user rights by means of policies. Furthermore, related to our policy model is work on access control for RDF [ACH+07]. Current work, however, solely targets restrictions on initial data access and not its ongoing usage. On the other hand, Hanson et al. developed a data-purpose algebra, which allows a modelling of purpose-restricted data [HBLK+07]. However, their work focuses on the verification of processes, whereas our work targets policy expression and enforcement. Lastly, Euzenat et al. introduce the notion of *personal infospheres*, which can be used to specify, what data may be shared with whom [He10]. Our policy model may be regarded as an instantiation of the infosphere model, in which users specify policies that regulate data sharing with external parties.

6 Conclusion

In this paper, we have realised a privacy-aware Smart Grid via royalty-free and open standards. In particular, we have presented a scenario illustrating how Linked Data principles may be applied. Our approach leads to a distributed system, where users retain full control over their personal information using policies. Furthermore, we showed how a machine-interpretable intent can be employed for technical enforcement of privacy rights. Finally, we evaluated the proposed approach via a study of privacy principles and their realisation within our model. We plan to extend the current work in several directions, most notably: i.) adding *cryptographic* procedures for fostering technical privacy enforcement and ii.) means for *anonymisation*, in order to fully realise the basic privacy principles.

References

[ACH⁺07] Fabian Abel, Juri Luca De Coi, Nicola Henze, Arne Wolf Koesling, Daniel Krause, and Daniel Olmedilla. Enabling Advanced and Context-Dependent Access Control in RDF Stores. In *The Semantic Web*, pages 1–14, 2007.

[AS05] Sudhir Agarwal and Barbara Sprick. Specification of Access Control and Certification Policies for Semantic Web Services. In *E-Commerce and Web Technologies*, pages 348–357, 2005.

[CPW10] Ann Cavoukian, Jules Polonetsky, and Christopher Wolf. SmartPrivacy for the Smart Grid: embedding privacy into the design of electricity conservation. *Identity in the Information Society*, April 2010.

[ETP06] European Technology Platform - SmartGrids Vision and Strategy for Europeâs Electricity Networks of the Future. Technical report, European Comission, 2006.

[Ger10] The German Roadmap - E-Energy / Smart Grid. Technical report, German Commission for Electrical, Electronic & Information Technologies of DIN and VDE, 2010.

[HBLK⁺07] Chris Hanson, Tim Berners-Lee, Lalana Kagal, Gerald Jay Sussman, and Daniel Weitzner. Data-Purpose Algebra: Modeling Data Usage Policies. In *Eighth IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY'07)*, pages 173–177, Juni 2007.

[He10] Andreas Harth and Rudi Studer (eds.). Dagstuhl Perspectives Workshop Report: Semantic Web Reflections and Future Directions. Technical report, 2010.

[Kly04] Brian McBride Graham Klyne. Resource Description Framework (RDF): Concepts and Abstract Syntax. Technical report, World Wide Web Consortium, February 2004.

[MM09] Patrick McDaniel and Stephen McLaughlin. Security and Privacy Challenges in the Smart Grid. *IEEE Security and Privacy*, 7:75–77, 2009.

[NIS10] NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 1.0. Technical report, National Institute of Standards and Technology, 2010.

[Raa09] Oliver Raabe. Datenschutz im Internet der Energie. In Stefan Fischer, Erik Maehle, and Rüdiger Reischuk, editors, *GI Jahrestagung*, volume 154 of *LNI*, page 191. GI, 2009.

[Raa10] Oliver Raabe. Datenschutz im SmartGrid. *Datenschutz und Datensicherheit*, 2010.

[WHBLC05] Daniel J Weitzner, Jim Hendler, Tim Berners-Lee, and Dan Connolly. *Creating a Policy-Aware Web : Discretionary , Rule-based Access for the World Wide Web*, chapter I, pages 1–31. IRM Press, 2005.