3

Data Privacy and Confidentiality

Alberto Fernandez[1] and Karwe Markus Alexander[2]

[1]Sensing & Control, Barcelona, Spain
[2]University of Freiburg, Freiburg, Germany

Abstract

The transformation of the current electrical gird to a smart grid, enabling a real time analysis as well as response of electrical consumption, poses new security and privacy electricity grid challenges. It is of crucial interest for utilities to obtain precise consumption data, in order to manage the grid. From the security perspective confidentiality as well as integrity must be kept to ensure utilities receiving of correct data. From privacy perspective precise data poses a threat to customers. Precise energy data allows to gain a view into each participating household, which is beyond the original needs of performing grid management. The iUrban pilot builds a bridge between both contrary goals. Data needed for grid management is delivered in a precise form, while data for additional use cases, like analyzing energy consumption of a house, is delivered in a privacy preserving form.

Keywords: Security, Confidentiality, Smart Metering, Privacy, Privacy Enhancing Technologies.

Confidentiality (confidentiality, integrity, and availability)[1] is an information security requirement. We understand under confidentiality that only authorized entities shall be able to gain access to data. To ensure confidentiality, we consider the two concepts: access control and encryption.[2] Encryption means that

[1] http://security.blogoverflow.com/2012/08/confidentiality-integrity-availability-the-three-components-of-the-cia-triad/.

[2] Encryption can be seen also as a form of access control. Only entities which have the according key are able to decrypt properly encrypted data. Key management can be seen as access control.

cryptographic operations are performed on plaintext so that only authorized entities are able to decrypt it and read the retrieved plaintext. Access control means that only authorized entities are able to retrieve data from a system; in this case, the iURBAN platform. To ensure that only authorized entities are able to have access to iURBAN platform data, authentication is needed. To guarantee authentication, digital certificates as well as the distribution of log in credentials for platform users are performed. Three main concepts are used to protect the confidentiality of data in iURBAN:

- Security and encryption of the communication channels,
- Encryption of stored data,
- Access control to the iURBAN platform.

Figure 3.1 provides an overview of the iURBAN components connection. Between the components, between the data providers and the components, between the data providers and the users of the system as well as the smart city database (SCDB) component and the energy utility, the communication channel is protected in diverse ways. Data stored within the components is encrypted. Access to the system can only be obtained after an access control mechanism checks the validity of the request. The digital certificates as well as the access control rights need to be distributed under the platform users. The "iURBAN platform confidentiality manager" performs this work; this entity takes care for granting people and organization access rights to the platform, takes care for distributing the digital certificates, and is able to revoke both.

This is part of the administration area in iURBAN project. It is formed by a number of system administrators that have to be in charge of the system

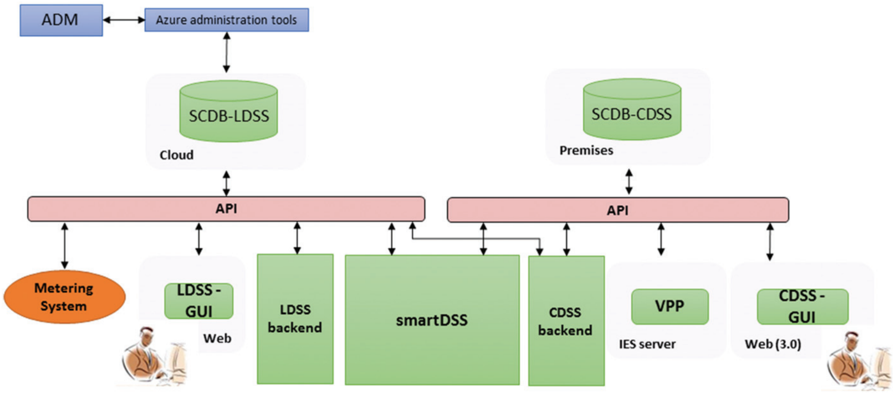

Figure 3.1 iURBAN components connection overview.

operational and security care. As in any other critical IT system, these technical personnel are in a trusted state by the company members of the consortium, but at the same time, they are not exempt of the security control to ensure all the operations they perform are conducted following the highest security levels. Lately, all the actions performed by these personnel are logged and securely stored for further review if necessary.

3.1 Confidentiality

The iURBAN system framework, as a whole, is seen as a service-oriented architecture (SOA) paradigm application and therefore constituted by several tiers of processing: data acquisition, data presentations, etc.

Data transmission networks transmit messages and commonly interconnect the several application tiers, which can include a mix of air and cable implementations. The end-to-end communication, across tiers, must be secured in order to ensure the basic security requirements of confidentiality, integrity, accessibility, availability, authenticity, and nonrepudiation and therefore protect the communication channels and messages that run across them (and even the ICT infrastructures).

The iURBAN security framework (SEC) is considering only Internet as the data transmission networks, because we are considering the utility networks as secure (Figure 3.2).

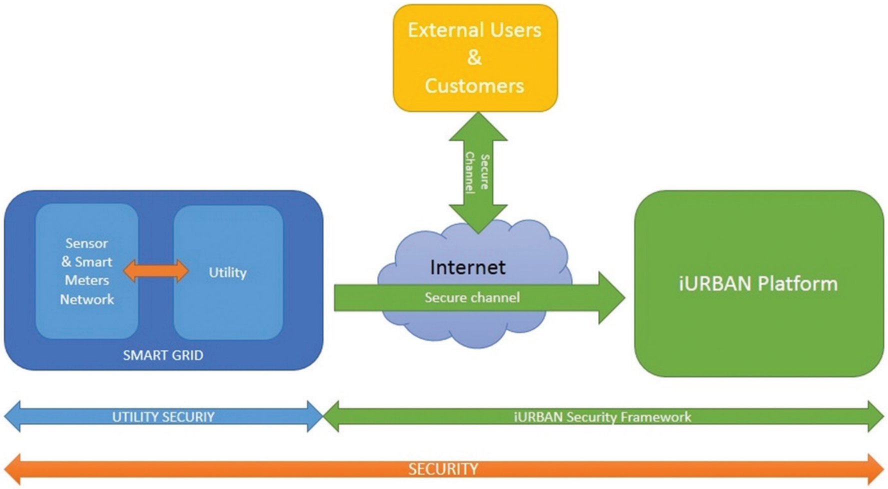

Figure 3.2 iURBAN security framework scope.

Confidentiality is an aspect of iURBAN nonfunctional requirements aimed at limiting information access and disclosure to authorized users and preventing access by or disclosure to unauthorized ones.

3.2 Confidentiality and General Security Requirements

To summarize, the security solutions implemented within the iURBAN framework in regard to confidentiality and other security-related aspects (EU directive 95/46/EC; EU recommendations 2012/148/EU; CEN-CENELEC-ETSI):

I. Confidentiality
   a. TLS/SSL protocol
   b. X.509 v3 certificates Class 1 with one-way authorization
   c. Azure traffic manager

II. Integrity
   a. TLS/SSL protocol
   b. X.509 v3 certificates Class 1 with one-way authorization

III. Authenticity
   a. MAC addresses control
   b. Authentication and authorization (the later existing on the administrators premises)
   c. TLS/SSL protocol
   d. X.509 v3 certificates Class 1 with one-way authorization

IV. Nonrepudiation
   a. TLS/SSL protocol
   b. X.509 v3 certificates Class 1 with one-way authorization
   c. MAC addresses control
   d. Authentication and authorization (the later existing on the administrators premises)

V. Accessibility and availability
   a. Offline data retention procedures
   b. High availability on components (load balancing/redundancy provided by Microsoft Azure Cloud)
   c. Backups/restore plans (provided by Microsoft Azure Cloud)

3.3 The iURBAN Privacy Challenge

The data minimization principle for smart metering data is in general implemented at the smart meter level. Such an implementation requires knowing in advance which question will be asked. The iURBAN platform is about to be able to ask questions about the data, which are not known in advance, creating the challenge to use a different approach.

Privacy preserving data publishing is a concept enabling to ask those questions, while preserving data privacy.

These privacy-preserving data publishing (PPDP) mechanisms are based on the interaction model where a data publisher uses collected data to issue them in a privacy-preserving manner (Figure 3.3).

PPDP mechanisms can be roughly classified into two categories based upon different attack scenarios and according to protection requirements. The concepts of the first one try to prevent that an attacker links records, attributes, or tables to a single person. The concepts in the second are based upon the uninformative principle: "*The published table should provide the adversary with little additional information beyond the background knowledge. In other words, there should not be a large difference between the prior and posterior beliefs*" [1]. Both categories allow to infer information about howl groups of participants but information gained upon an individual is limited which is achieved by reduced data accuracy [2].

While the business potential for fine granular consumption values, provided by smart meters is promising [3], those values pose the threat of privacy sphere invasion. In the Netherlands, this threat was sufficient to put a smart

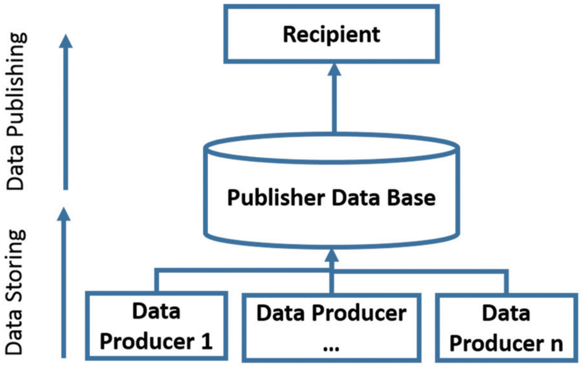

Figure 3.3 Privacy-preserving data publishing model.

meter roll out on hold [4]. By using customer data and energy values, a demand response aggregator (DRA) is able to find out customers' incentive and representing a considerable advantage on determining the compensation amount of an offer [5]. Computer science tries to avoid intermediaries and provides a variety of privacy-enhancing technologies (PETs) such as zero knowledge proofs (ZKP) where only one bit of information is divulged instead of a full consumption trace of fine granular energy data, creating the situation where the question be asked, must be known in advance [6]. The computing penalties are high and the approach is not promising in respect of managing smart grid.

Figure 3.4(A) shows a feedback loop of tasks for indirect load control (ILC). ILC is one option to create load shifts, and one form of it was implemented in iURBAN. Another option is direct load control, where a DRA can directly control devices at customer side or tariff-based programs where different prices at different times create incentives to shift load. Without an intermediary in ILC, a DRA makes a proposal directly to residential customers, e.g., households which accepts or rejects it. Out of the accepted, DRA selects a sufficiently large group to perform the intended load shift and sends this group the participation acknowledgement. For program verification and forecasts, DRA obtains energy data from the according household smart meter. We assume that the only way to obtain plain energy data is via access to the smart meter.

In price-based DR as shown in Figure 3.5, a DRA sends price signals to the respective smart meter to influence customer's behavior consumption. In an

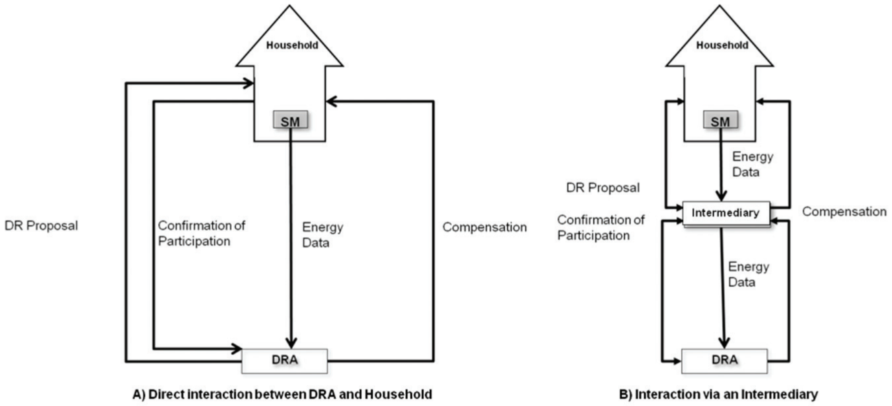
Household
SM
DR Proposal
Confirmation of Participation
Energy Data
Compensation
DRA
A) Direct interaction between DRA and Household
Household
SM
Energy Data
DR Proposal
Intermediary
Compensation
Confirmation of Participation
Energy Data
DRA
B) Interaction via an Intermediary

Figure 3.4 Demand response data and interaction.

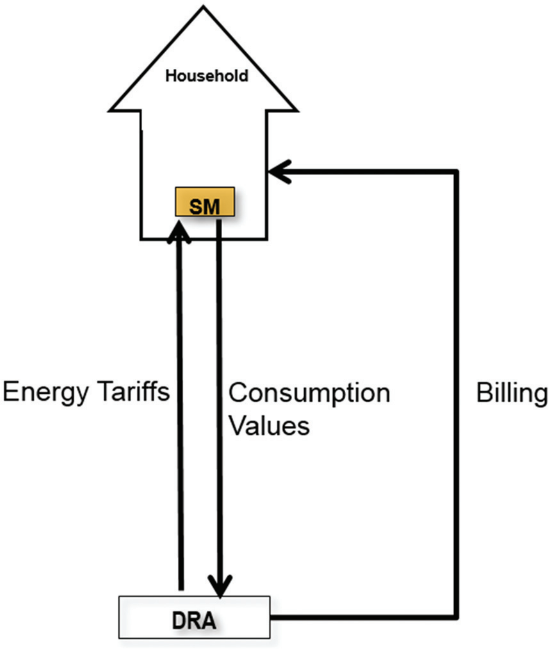

Figure 3.5 Price-based demand response and interaction.

intermediary setting, the signals are sent through the intermediary forwarding it to related recipients.

In both cases, the feedback loop allows DRA to exercise control of households. The loop needs data traces in order to be performed. Those traces leak privacy-sensitive information and thus are a potential privacy threat. Figure 3.4(B) shows how this feedback loop is interrupted by an intermediary. The iURBAN platform acts as the intermediary as shown in Figure 3.6.

In this feedback loop, all energy data are stored within the SCDB. The other components are accessing the data via the SCDB interfaces. This approach considers SCDB as a central cardinal point to protect PPDP. iURBAN follows the privacy by design approach which consists of seven principle steps:

- Proactive not reactive, preventative not remedial
- Privacy as the default setting
- Privacy embedded into the design
- Full functionality—positive sum, not zero sum
- End-to-end security—full life cycle protection

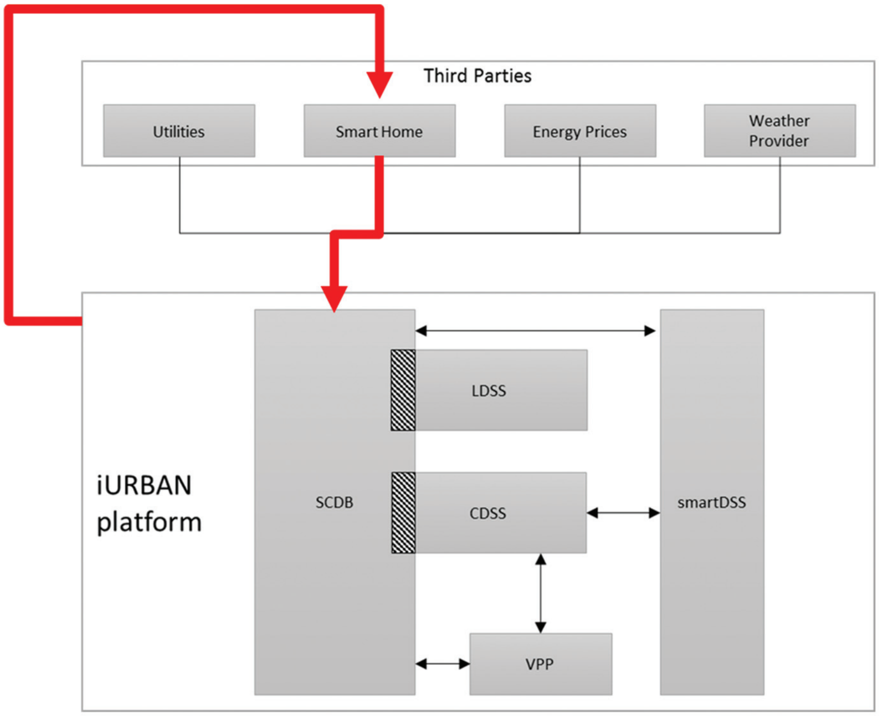

Figure 3.6 Logical architecture for iURBAN platform feedback loop.

- Visibility and transparency—keep it open
- Respect for user privacy—keep it user centric.

As intended by EU directive 95/46/EC, the privacy preservation concept of iURBAN is based upon PET and transparency enhancing. The fundament of these concepts is confidentiality as described in the first part of this section. The circumstance that the guarantee holds even in case of arbitrary background knowledge as well as the composability property makes it the suitable technique for the iURBAN project from privacy protection perspective. The impact on data utility and thus the impact on the energy goals need to be balanced with the required level of privacy protection. For transparency enhancing, we use the opt-in approach intended by the amendment of the data protection directive. An opt-in option is provided to the data producer where he can choose to allow direct access to his energy data without further privacy protection.

3.4 Privacy Enhancing via Transparency

Transparency-enhancing technologies (TETs) aim to inform the user how his data are stored, processed, and used for. The new proposal for data protection regulation demands the following points:

- Individuals shall fully understand how their data are handled.
- Explicit consent of subject to process their personal data which are achievable via TET.

The explicit consent is achieved by providing choice for different levels of data protection. There are several options to provide plaintext access to energy consumption data differing in the level of granularity. Another option is to provide consent for providing their data under the differential privacy guarantee. This choice determines the individual data protection policy. The collected data can be used for all iURBAN platform activities.

Each individual data producer is informed how his data are collected, stored, and processed via local decision support system (LDSS). It achieves this by providing information texts and graphics during the procedure to provide consent.

The user is able to revoke his consent. In this case, the past as well as future data may not be used.

3.5 Privacy Enhancing via Differential Privacy

Following the PPDP model shown in Figure 3.7, SCDB is the data publisher database within the iURBAN architecture. To ensure that data retrieval is only feasible in intended ways, we envision a privacy proxy regulating the access to SCDB and providing answers to request only in a privacy-preserving manner. This privacy proxy encapsulates the SCDB and acts as a trusted third party (TTP). While incoming data by legitimate users are passed to the database directly, all outgoing data are processed to keep privacy of users.

Role-based access control is applied to ensure that only legitimate data providers are allowed to put data into the database. Via interfaces and APIs, the privacy proxy receives queries. By access control, it is ensured that only legitimate users can successfully query. For each query, the privacy proxy (PrP) checks whether it is within the scope of the iURBAN privacy policy, respectively checks whether the budget is not exceeded or the request contains data producers who opted-in and those who did not. The proxy retrieves the data from SCDB and processes it to protect the privacy according to policy.

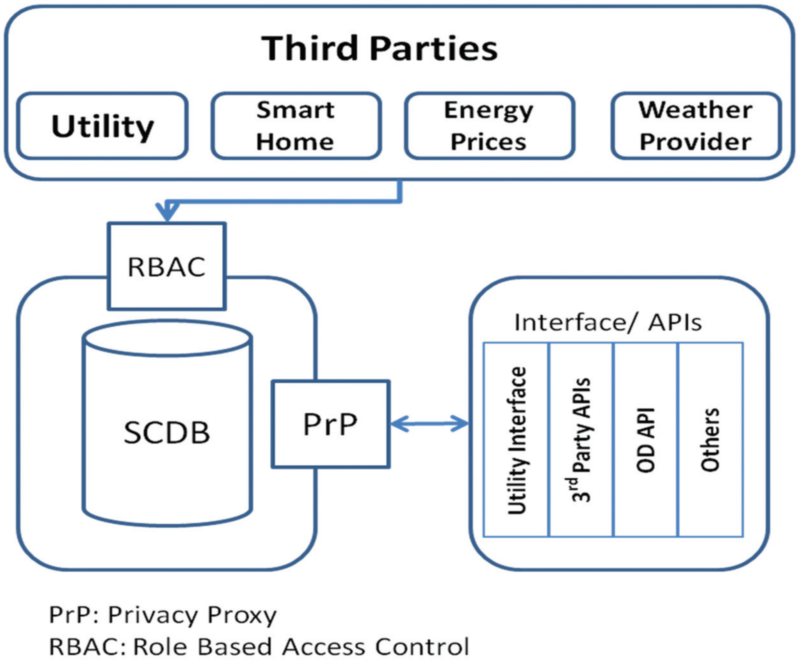

Figure 3.7 Logical view database privacy proxy.

Only this privacy protected answer is provided to the inquirer. Note that the only way to retrieve information from SCDB must be via the privacy proxy.

3.5.1 Privacy-Enhancing Technologies Based on Privacy Protection

The privacy proxy applies two kinds of privacy protection. The first one is temporal aggregation of consumption values of single data producers if they opt-in accordingly. Thus, if the customer stated he allows only access to energy data for a whole hour, the proxy sums up only complete four quarter hour measures.

The second option to protect privacy is to hide values of single customers within a privacy user group, if they did not grant access to plain energy values. This is achieved by using techniques providing the guarantee of differential privacy. Note that it is not possible to request data from producers who opt-in for plaintext access combined with data producers who did not under the differential privacy protection [7] (Figure 3.8).

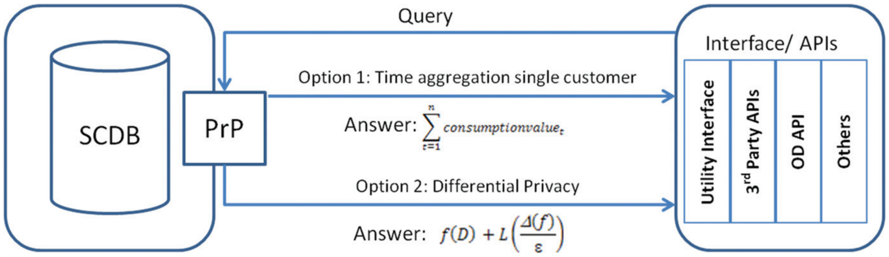
Query
SCDB
PrP
Option 1: Time aggregation single customer
Answer: $\sum_{t=1}^{n} consumptionvalue_t$
Option 2: Differential Privacy
Answer: $f(D) + L\left(\frac{\Delta(f)}{\varepsilon}\right)$
Interface/ APIs
Utility Interface
3rd Party APIs
OD API
Others

Figure 3.8 Logical view privacy protected query answers.

To mitigate the privacy threats for energy values, requests for differential privacy protected data shall be allowed as long as the following privacy requirements holds:

A user of the iURBAN platform shall not be able to learn an exact energy consumption value of a 15-minute interval for a single data producer.

To keep this requirement, a privacy budget for requests per data consumer ($\varepsilon - \mathrm{budget}$) must be settled. As soon as $\varepsilon - \mathrm{budget}$ is used up, no further request shall be allowed. SCDB is continuously updated by the data producers. Differential privacy as a worst case guarantee associates this budget to the whole lifetime of the database with the consequence that in case of a used up budget, future requests of future energy values would be blocked. To prevent such a situation, $\varepsilon - \mathrm{budget}$ is associated with time frames. Only requests for this frame are blocked, if the budget of the frame is used up. A suitable $\varepsilon - \mathrm{budget}$ as well as an appropriate time frame needs to be evaluated during the course of the project.

3.5.2 Privacy Protection Implementation

The implementation of the privacy proxy approach is based upon the PINQ differential privacy framework [8]. It is written in C# and follows the privacy budget concept. The framework enforces differential privacy but does not define the access policy to the data base. For enforcing differential privacy, PINQ implements as aggregation operations NoisySum as well as NoisyCount achieved via Laplace-based noise mechanisms and NoisyMedian plus NoisyAvg achieved via an exponential noise mechanism.

The privacy proxy enforces that a differential privacy-preserved request does not contain data of customers who opted-in for providing their precise data. For enforcing access control, the privacy proxy relies on access control mechanisms of the underlying database.

3.6 Conclusions

This chapter has provided information about the approach implemented within iURBAN with respect to privacy and confidentiality of energy information being captured and stored from the smart grid.

As iURBAN stores public (public buildings) and private (households) energy information, its APIs and interfaces have been built to allow the data to flow transparently or biased depending on the level of privacy that the user would like to maintain.

A proxy has been designed and applied over that data captured by iURBAN for residential buildings. The proxy induces some perturbation on the original data in order to maintain privacy, which depends on the time span of the request and the type of query, meaning that depending on the target of the use of data queried, the corresponding use case can be or not achieved; this implies that this approach can jeopardize the possibility to launch services that can provide benefits the household (which holds the ownership of the data).

References

[1] Machanavajjhala, A., Gehrke, J., Kifer, D., and Venkitasubramaniam, M. (2006). l-diversity: Privacy beyond k-anonymity. In *Proceedings of the 22nd IEEE International Conference on Data Engineering (ICDE)*, 2006.

[2] Chen, R., Fung, B., Wang, K., and Yu, P. (2010). Privacy-preserving data publishing: A survey of recent developments. *ACM Comput. Surv. (CSUR)*, 42 (4), 2010.
Confidentiality, Integrity and availability:
http://security.blogoverflow.com/2012/08/confidentiality-integrity-availability-the-three-components-of-the-cia-triad/

[3] Albert, A., and Rajagopal, R. (2013). Smart meter driven segmentation: What your consumption says about you. *IEEE Trans. Power Syst.*, 28 (4), November 2013.

[4] AlAbdulkarim, L., and Lukszo, Z. (2011). Impact of privacy concerns on consumers' acceptance of smart metering in The Netherlands, international conference on networking, sensing and control, Delft Netherlands, 2011.

[5] Karwe, M, and Strüker, J. (2014). Privacy in residential demand side management applications. *Smart grid security: Second international workshop, SmartGridSec2014*, Munich Germany, 2014.

[6] Jawurek, M., Kerschbaum, F., and Orlandi, C. (2013). Zero-knowledge using garbled circuits: How to prove non-algebraic statements efficiently. *ACM Conference on Computer and Communications Security 2013*.

[7] Dwork, C., and Smith, A. (2009). Differential privacy for statistics: What we know and what we want to learn. *J. Privacy Confident. 2009*.

[8] McSherry, F. (2009). Privacy integrated queries an extensible platform for privacy-preserving data analysis. *SIGMOND* 2009.