A Privacy Impact Assessment for smart meter data on the cloud

Avaliação de Impacto à Privacidade para dados de medidores inteligentes na nuvem

Matheus Machado
Décio E. do Nascimento
Keiko V.O. Fonseca

UTFPR – Universidade Tecnológica Federal do Paraná
machadom@alunos.utfpr.edu.br
decio@utfpr.edu.br
keiko@utfpr.edu.br

Abstract: Smart Grid (SG) brings advantages to the traditional grid by enabling the bidirectional flow of information and energy. However, detailed energy profiles provided by the Smart Meters pose great privacy risk as they allow the inference of several actions and behaviours thought to be private. The purpose of this paper is to conduct a Privacy Impact Assessment (PIA) on a SG using cloud services. The Privacy Requirements (PR) of the system were taken from Brazilian General Data Protection Law and the PIA process was aided by an Advanced Cloud Privacy Threat Modelling methodology that systematized the enumeration of threats to the PRs and their corresponding countermeasures. Results were then incorporated in the design of the SecureCloud's high-level data model. The process, when taken on the early stages of the project development, can enable systems in line with the principles of Privacy by Design.

Keywords: Privacy Impact Assessment, smart metering, smart grid, threat modelling, privacy, cloud.

Resumo: A Rede Elétrica Inteligente (REI) apresenta vantagens em relação à rede tradicional por habilitar o fluxo bidirecional de informação e energia. Contudo, perfis detalhados de consumo energético fornecidos pelos Medidores Inteligentes apresentam grande risco à privacidade por permitirem a inferência de diversas ações e comportamentos considerados privados. O objetivo deste artigo é conduzir uma Avaliação de Impacto à Privacidade (AIP) em uma REI utilizando computação em nuvem. Os requisitos de privacidade foram extraídos da Lei Geral de Proteção de Dados Pessoais brasileira e o processo de AIP foi auxiliado por uma metodologia avançada de modelo de ameaças na nuvem que sistematizou a enumeração de ameaças aos requisitos de privacidade e as medidas tomadas para mitigá-las. Os resultados foram incorporados ao desenho do modelo genérico de dados do SecureCloud. pode habilitar sistemas alinhados aos princípios de privacidade desde a concepção do projeto.

Palavras-Chave: Avaliação de Impacto à Privacidade; medição inteligente; Rede Elétrica Inteligente, modelo de ameaças, privacidade, nuvem.

1 Introduction

The purpose of this paper is to conduct a high-level Privacy Impact Assessment (PIA) on a modelled Smart Grid (SG) that would use the cloud to operate. The PIA process should consider legal and technical privacy requirements, detailed threat identification enabled by threat modelling and provide recommendations towards a system that observes the Privacy by Design approach.

The transition of the Brazilian Energy Sector to Smart Grid applications is supposed to bring several advantages over the traditional grid by allowing the connection of renewable sources and the bidirectional flow of information and energy [1]. The expected value comes from processing the data generated by the Smart Meters to better operate the grid and to enable value-added services, such as third-party management of electric appliances [2] [3].

Data protection legislation applies to the processing of Personally Identifiable Information (PII) [4] [5] and privacy by design is increasingly being required by law [6]. Smart Meter data may reveal information that is expected to be private [7] [8] [9] [10] and the grid development will rely on the public's trust in the system [11] [12]. Design choices taken from the outset of the system will determine how private it will eventually be and influence how the public balances the expected benefits against the perceived privacy risk [6] [13].

The argument presented in this work is that these privacy issues, when ignored in the early stages of the system development, could lead to costly privacy retrofitting [14], fines from breaching data protection legislation[1], violation of rights and delay of its adoption and deployment [15] [16].

It has been demonstrated that SG applications can take advantage of cloud computing characteristics such as ubiquitous access and easily scalable provisioning to be deployed and managed [3]. However, cloud computing can be provided in different deployment models, such as Private, Public or Hybrid cloud, and service models (Software as a Service – SaaS; Platform as a Service – PaaS; Infrastructure as a Service – IaaS) [17] with varying degrees of access granted to the service provider. When choosing which services better fit its needs, the client should consider the shared responsibility for the security of the system, but

[1] Equifax settled to pay up to $700 million for their data breach https://www.wired.com/story/equifax-fine-not-enough/
British Airways is facing a fine of £183 million for its security breach https://www.bbc.com/news/business-48905907

also how their responsibility is shared in relation to data protection.

Privacy requirements are hard to translate into technical specifications [13] [18]. A PIA is a methodology to assess and address the impacts on privacy that are likely to arise from a project [19] [20]. It is a reproducible process to determine the privacy, reliability and risks in the collection, use and disclosure of information related to individuals, from which a report is produced containing the recommendations to mitigate them [19] [20] [21]. Within this paper, the terms Privacy Impact Assessment (PIA) and Data Protection Impact Assessment (DPIA) are used interchangeably [22].

The soon-to-be enforced data protection law in Brazil was approved on the 14th of August 2018 and applies to the processing of personal data to protect the free development of the person and the fundamental rights of freedom and privacy [4]. It sets principles to be observed by every actor, citizens, companies or government institutions, willing to process personal data. One of the principles that must be observed is prevention (article 6, VIII), or the adoption of measures to avoid the damage from arising when processing personal data. It also defines (article 5, XVII) a data protection impact report as the *"controller's documentation that contains the description of the data processing that could generate risks to civil liberties and fundamental rights, as well as measures, safeguards and risk mitigation mechanisms"* [4]. The law doesn't specify situations where the reports are mandatory, but they can be required by the Data Protection Authority to the data controllers.

Shapiro [23] discusses the limitations of anonymization as a privacy risk control and Narayanan, Huey and Felten [24] show why the risk of ad hoc data de-identification is unknown and unknowable. The overall security of the system depends on its weakest component, so protection cannot be achieved by focusing on one part and neglecting others [51]. This paper takes the approach that data minimization must be the main step to build systems in line with the principles of Privacy by Design [13]. The quality and quantity of data needed to perform each function expected of the Smart Grid varies [2]. For instance, to perform the billing function, the Utility must securely attribute the calculation of the consumption to a specific Smart Meter, but it does not need to have access to this information in such a way that would allow for the discovery of behaviour within the home. At the same time, it might need such high-frequency data for operational reasons, in which case it cannot be tied back to one single Smart Meter and should be analysed in bulk [11].

The remainder of this paper is structured as follows: section 2 discusses previous work in the domain and situates the expected contributions of this paper; section 3 outlines the methodology employed to take the PIA process; section 4 briefly describes smart metering and its associated issues; section 5 describes cloud computing essential characteristics, deployment models, service models and their associated issues; finally, in section 6, the PIA process is conducted, the SecureCloud model is described, privacy threats are listed and their respective countermeasures are proposed. The concluding section discusses the limitations encountered.

2 Previous work

There is a range of papers surveying the Smart Grid with regard to its technical standards [25], its communication infrastructures [26] [27] and its cybersecurity challenges [28]. This paper assumes cybersecurity is a very important aspect of the SG development in a way that, If a security assumption fails, it can lead to a privacy breach. However, the scope of the process was limited to Smart Grid and Cloud Services' privacy issues (see Methodology) and not their security issues and controls.

Asghar et al. [2] present a survey on Smart Meter Data Privacy, reviewing the different uses of metering data in the smart grid, namely for Billing, Operations and Value-Added Services, alongside the privacy legislation. The authors provide an overview of existing solutions to privacy issues, noting their limitations, and offer recommendations and future research directions.

Some authors [29] propose a privacy-preserving system where the certified outputs of a tamper-evident meter are processed on the user device, through privacy-friendly calculations, allowing for time-based tariff billing, operations and profiling without the disclosure of fine-grained consumption data. It does not require extensive computation to be performed on the meter other than the certification of the outputs. However, it depends on the end-user having the means to perform the privacy-preserving computations on their devices, such as a smartphone or computer. The proposition of this work borrows some assumptions from Rial and Danezis [29], namely, that the Smart Meter has limited computational resources and fine-grained consumption data should not be accessed by parties other than the user. However, privacy-preserving computations are proposed here to be performed in a Trusted Execution Environment (TEE) in the cloud.

M'Rhaoaurh et al. [30] conducted a systematic literature review to identify challenges to cloud computing. The search string: ("cloud computing" OR "cloud" OR "cloud technologies") AND ("issues" OR "challenges" OR "barriers" OR "threats" OR "limits" OR "risks" of cloud computing) was queried on multiple databases, such as IEEE Explore and Science Direct, and the papers were selected based on their titles, abstracts and contents. The authors selected 60 studies to review from the initial 800 obtained from the query and found 23 potential issues of cloud computing. The issues were grouped under "policy and organizational" (e.g. loss of governance and vendor lock-in), "technical" (e.g. data loss or leakage and denial of service) and "legal" (e.g. subpoena and international jurisdiction) according to the three categories offered by the European Union Agency for Network and Information Security (ENISA).

Threat modelling is a known tool to help in the Software Development Life Cycle (SDLC) and has been proven useful to mitigate security risk [31]. Gholami et al. [32] present a Cloud Privacy Threat Modeling (CPTM) methodology that is

an extension of the CPTM presented in 2013 [33] to support the EU Data Protection Directive. In their work, they expand the methodology to encompass other legislation, such as the US Health Insurance Portability and Accountability Act (HIPAA), and to address the privacy issues discovered through the requirements engineering process.

This paper differs from previous papers on Smart Grid and cloud privacy through (i) reviewing the privacy issues in a multidisciplinary way considering the Brazilian Data Protection Law and academic contributions to gather threats and countermeasures; (ii) surveying Smart Grid privacy issues alongside Cloud Services issues to account for the assumption that the grid's transition will be reliant on the cloud; (iii) presenting a data model that leverages recent technology (IntelSGX) to provide trusted code execution in the cloud.

3 Methodology

The goal of this paper is to carry a PIA process that considers legal privacy requirements and privacy threats for a Smart Grid that uses cloud services, which themselves have privacy threats to be considered. To do so, we follow the Information Commissioner's Office (ICO) code of practice on conducting Privacy Impact Assessments [34] supported by Advanced Cloud Privacy Threat Modeling [32].

Privacy Impact Assessments are regarded as a tool to foresee privacy risk and build systems in line with Privacy by Design principles [14]. The ICO, UK's independent body responsible for upholding information rights, has published a handbook on conducting Privacy Impact Assessments [35] that has been revised and expanded into a Code of Practice [34] that provides privacy concepts, overview and guidance on the PIA process. The steps described in the document to produce the PIA report, namely, (0) Identify the need for a PIA; (1) Describe the information flows; (2) Identify the privacy and related risks; (3) Identify and evaluate the privacy solutions; (4) Sign off and record the PIA outcomes, were taken in this paper. The final step in the document, (5) Integrate the outcomes into the project plan [34], is contained in the SecureCloud data model presented.

Some steps of the Advanced Cloud Privacy Threat Modelling (CPTM) as presented [32] [36] were taken to model adversarial behaviour on the cloud domain that could bring repercussions on privacy. Their methodology lists the (0) Privacy Requirements, including regulatory compliance specifications, on the Requirements Engineering stage of the Software Development Life Cycle (SDLC), which are then sent to the Design stage, where the activities of (1) Cloud Environment Specification, (2) Privacy Threat Identification, (3) Risk Evaluation and (4) Threat Mitigation happen. In conducting the threat modelling before the Implementation, Verification and Validation stages of the SDLC, the methodology fosters privacy by design and is more effective than ad hoc solutions. The Advanced CPTM will be used in steps 2 (Identify the privacy and related risks) and 3 (Identify and evaluate the privacy solutions) of the PIA process [34] as a systematic way to elicit privacy requirements and privacy threats on the cloud environment.

The PIA methodology [34] and the Advanced Cloud Privacy Threat Modelling methodology [32] suggest a consultation with stakeholders, participants of the system, software engineering team and legal experts, both to discover risks and evaluate them, which was not done in this paper. Nevertheless, we draw from academic and legal contributions to conduct our process and believe that it can help practitioners and researchers. Eventually, when a PIA is conducted on an actual system, it is recommended that the different stakeholders are consulted.

3.1 Privacy regulatory compliance

In this step of the Advanced CPTM, the privacy requirements of the system will be listed. Privacy regulatory frameworks can be complex to translate to engineering teams [13] [32]. This work uses the principles that must be observed when processing personal data in Brazil contained in the 6th article of the recently approved Brazilian General Data Protection Law, Law nº 13.709/18 [4], to list and enumerate the Privacy Requirements (PR_i) where $_i$ corresponds to the number of the requirement.

The ten principles contained in the 6th article are: I – Purpose binding; II – Adequacy; III – Necessity; IV – Access; V – Data accuracy; VI – Transparency; VII – Security; VIII – Prevention; IX – Non-discrimination; X – Accountability [4] and have been chosen as the PRs for this paper.

3.2 Cloud environment specification

Cloud service models vary [17] and the attack surface is different for each, for instance, SaaS clouds would benefit from a deeper threat analysis for application-level threats while LaaS requires an emphasis on physical hardware, network and storage [32]. A private cloud limits the physical access threats while moving greater management responsibilities to the cloud consumer, which may or may not be better prepared to deal with them than the cloud provider. The right balance will depend on the mission statement of the project and the tradeoffs intrinsic to each model. In this step the cloud actors are defined, such as the Cloud Consumer and the Cloud Provider, the system is modelled and the assets that need to be protected are listed.

3.3 Privacy threat identification

The privacy threats against the selected Privacy Requirements (PR_i) are then identified and analysed, leading to the identification of threats that will be named $T_{i,j}$, where $_i$ indicates the PR to which the threat is associated and $_j$ is the number of the threat. For instance, $T_{2,3}$ refers to the third threat for the second PR (PR_2).

3.4 Risk evaluation

In this step the participants in the system, such as engineers, lawyers and architects would offer input to grade the threats with regards to their likelihood and effects and offer a risk evaluation matrix that could help prioritize mitigation actions [32].

3.5 Threat mitigation

Countermeasures for the identified threats are offered to meet the goals of the project. Their notation follows the previous scheme, where $C_{2,3}$ refers to the countermeasure for threat $T_{2,3}$.

The Advanced CPTM was applied to support the PIA process and the results were presented following the order described in [34].

4 Smart Meter data issues

The Smart Grid (SG) may be defined as a network that can integrate the actions of all users, consumers and suppliers, to efficiently deliver sustainable energy supply [37]. Its main difference from the traditional grid is the bidirectional flow of information and energy [38]. Traditionally, the supplier received only the sum of the consumption for a given period (e.g. 1 month) from the consumer in order to bill them accordingly. The SG differs from this traditional model by incorporating Smart Meters (SM), devices that have the capacity to record fine-grained consumption for a certain service, allowing for dynamic pricing and inferences about usage patterns. Smart Meters can be used to measure services other than electricity, such as water and heat [39], however, this paper considers only the electricity scenario.

The Smart Meters can allow a party with access to detailed consumption information to expose activities from within the home, once thought to be private [40]. The data from the smart metering system carries privacy issues: the detailed consumption information might be used maliciously to infer the behaviour of a customer (e.g. to determine sleep/wake cycles, vacation time) [39], to determine which appliances are used (brand, model, usage time) [40] or even to enable family members to "spy" on each other (parents checking if their children are up late playing videogames) [41].

A panel that sums some of the concerns is presented in Table 1, adapted from [40] and [42].

Awareness of privacy implications has caused consumers in the Netherlands to oppose and delay Smart Meter rollouts [15] [16], indicating that public acceptance will be a decisive factor in the deployment of the system.

The SG promises to deliver better operation of the system for the Utility through new data provided by the Smart Meters and to foster new services and markets [2] [3]. Privacy issues are studied in some works [40] [43] [44], and suggestions are offered to enforce privacy properties via procedural means, such as agreements between parties to not correlate the data. These solutions assume that price-varying billing requires access to detailed consumption data [29].

The approach to "collect first, process later" is a privacy problem [45]. Other approaches have been studied to enable suppliers to perform time and quantity-based pricing without receiving detailed consumption data [29] [39] [46]. In face of the limited effectiveness of alternatives to remove private information from consumption traces [45], this work assumes that data minimization is the goal when trying to build systems in line with the principles of Privacy by Design [13].

Table 1: Privacy concerns in Smart Metering.

Concern type	Threat examples
Illegal uses	Burglars discovering when the house is empty; Existence and usage of electronic alarm systems; Stalkers tracking their victims.
Commercial uses	Targeted advertisement (When do they watch TV? How often do they eat-out?) Insurance adjusting (How often do they get a full night's sleep v. drive sleep deprived? How often do they leave late, probably driving recklessly? How often do they forget appliances running?)
Uses by law enforcement agencies	Detection of illegal activities (e.g. sweatshops, drug production) Verifying defendant's claim (e.g. claiming they were not home all night)
Uses by other parties for legal purposes	In a custody battle (Do they leave their child home alone?) In a worker's comp hearing (If they were disabled, how could they have performed this set of activities?)
Uses by family members and other co-inhabitants	One householder monitoring another (parents checking for bed-time obedience) Abusive partners controlling each other's behaviour
Medical Uses	Do clinically depressed or bipolar individuals have distinctive energy profiles? Could you tell if someone had not been taking their medication?

5 Cloud characteristics and issues

In this paper, the definition of cloud computing offered by the US National Institute for Standards and Technology (NIST) is adopted [17]:

> "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."

Cloud definitions are centered on the concept of utility-computing and carry a vision where users consume resources on-demand and benefit from a market of suppliers to optimize costs and access better services [47]. NISTS's definition lists five essential characteristics (On-demand self-service; Broad network access; Resource pooling; Rapid elasticity; Measured service), three service models (Software as a Service – SaaS; Platform as a Service – PaaS; Infrastructure as a Service – IaaS) and four deployment models (Private cloud; Community cloud; Public cloud; Hybrid cloud). The essential characteristics, service models and deployment models are briefly described in the next sections [17]:

5.1 Essential characteristics

On-demand self-service: Consumers can provision computing capabilities according to their needs without the need for provider intervention;

Broad network access: Resources and capabilities can be accessed over the network, allowing for consumers to be independent of location or client platforms (e.g., smartphones or workstations);

Resource pooling: Storage, processing power, memory and network bandwidth are dynamically assigned and reassigned according to consumer demand, creating a multi-tenant model;

Rapid elasticity: Capabilities can scale rapidly, outward and inward depending on demand, in some cases automatically. From the consumer perspective, resources are available at any quantity at any time;

Measured service: Resource usage is monitored and used to bill the customer solely for the utilized service, typically on a pay-per-use or charge-per-use basis;

5.2 Service models

Software as a Service (SaaS): Cloud consumers use the provider's applications running on a cloud infrastructure. Consumers do not manage or control the infrastructure and the application capabilities, except for possible limited user configuration settings;

Platform as a Service (PaaS): Consumers control the application to be deployed onto the cloud infrastructure, but do not manage or control the underlying resources, such as network, servers, operating systems or storage, except for the possibility of configuring settings for the application-hosting environment;

Infrastructure as a Service (IaaS): Consumers can provision processing, storage, networks and other computing resources and are able to run arbitrary software, such as operating systems and applications. Consumers do not manage or control the underlying infrastructure, apart from some select networking components, such as the host firewall.

5.3 Deployment models

Private cloud: The cloud infrastructure is used exclusively by a single organization comprising multiple consumers (business units). The responsibilities for control and management of the infrastructure may be held by the organization itself, a third party or a combination of them and it may be located on or off-premises;

Community cloud: The cloud infrastructure is used exclusively by a specific set of consumers that benefit from grouping to achieve the same goals, such as security requirements, policy and compliance considerations. The operation and ownership may be with one or more of the organizations, with a third party or a combination of them and the infrastructure may be located on or off-premises;

Public cloud: The cloud infrastructure is offered for the general public. The resources are shared by the users through virtualization and are owned, managed and operated by the cloud provider, who may be a business, academic or government organization or a combination of them. The infrastructure is located on the premises of the cloud provider;

Hybrid cloud: The cloud infrastructure is a combination of cloud deployments that are unique but interconnected to enable an integrated workflow.

5.4 Issues

Traditionally, security models start by delimiting the security perimeter within which there is control over the resources and sensitive information may be stored and processed, such as the limit set by a corporate firewall. These boundaries become hard to define in public and hybrid clouds when information gets stored and processed outside trusted areas [47] [48]. While a private cloud deployment solves some of the physical security issues and provides greater confidentiality and control [30], some threats stem from characteristics of all cloud models [49], for instance, the access control procedures must ensure that no unauthorized party is allowed to connect to the private cloud, highlighting the need for strong access control and Intrusion Detection Systems (IDS).

Private clouds are expensive and Public clouds are the most prevalent deployment model for cost-effectiveness. At the same time, the infrastructure and resources are owned, managed and operated by the Cloud Service Provider (CSP), causing privacy concerns when processing personal information outside of the organization's security perimeter [47].

Some privacy issues are consequences of security issues, such as unauthorized data access. However, Pearson and Benameur [48] highlight four privacy issues for cloud computing:

Lack of user control: when using a SaaS model, for instance, the user relies on the service provider to store and process their data in a certain way, with little or no ability to control or check such assumptions;

Unauthorized secondary usage: Some secondary uses of data controlled by the service provider are agreed to on the terms of service, e.g. for advertisement purposes, however, the user has no technical means to enforce only the agreed uses or forbid further secondary usage;

Data proliferation and transborder data flow: It is a characteristic of the cloud model to replicate data, be it to achieve lower latency in content delivery or redundancy in backups. These multiple instances of data may reside in jurisdictions with lower legal protections, facilitated law enforcement access or may not be deleted alongside other copies. It increases the difficulty in complying with data protection legislation as more than one jurisdiction may be involved and increases the need for effective governance processes across organizations;

Dynamic provisioning: In a dynamic environment, it is hard to ensure that security and legal requirements are being observed by the CSP or eventually its subcontractors.

6 Privacy Impact Assessment (PIA)

The PIA process conducted here is described in the Methodology section. There was a combination of the steps presented by UK's Data Protection Authority, ICO, on their Code of Practice [34] and the Advanced Cloud Privacy Threat Modelling methodology, presented in [32].

6.1 Overview

The SecureCloud platform aims at enabling the secure processing of sensitive data in untrusted clouds [50]. If the data security and privacy requirements of Smart Grid applications were met, the project could facilitate the usage of cloud computing while maintaining acceptable levels of trust in the system [46].

The platform leverages cryptographic hardware present on modern Intel CPUs, Intel Secure Guard Extensions (SGX), to access sensitive data only on the Trusted Execution Environment (TEE). It builds enclaves, protected memory areas of code execution, that are secluded from privileged software, shielding data even from the Cloud Service Provider [46] [50].

The data are encrypted at the Smart Meter level and a Role-Based access control occurs at the Meter Data Management System to ensure that only the least amount of information needed to perform the desired function is disclosed. A detailed load signature must be available only to its owner and different computations expected to be performed should happen on different aggregations of SM data [46], providing greater privacy and facilitating the outsourcing of functions without adding new threats.

6.2 Privacy Requirements

The privacy requirements were selected from the Brazilian General Data Protection Law and are the principles required to be observed when processing personal information in accordance with the country's legislation [4]:

PR_1 – *Purpose binding*: The processing must be performed for legitimate, specific, explicit and informed purposes, with no possibility of secondary processing in a way that is incompatible to those goals;

PR_2 – *Adequacy*: The processing is compatible with the purposes informed to the data subject;

PR_3 – *Necessity*: The processing shall be limited to as little as necessary to achieve its goals, accessing the relevant, proportional and non-excessive data related to its function;

PR_4 – *Access*: To the data subject is warranted the facilitated and free consultation of how and how long the processing is happening, as well as the completeness of their data;

PR_5 – *Data accuracy*: Assurance to the data subjects that their data is accurate, clear and up to date;

PR_6 – *Transparency*: The data subjects are warranted clear, precise and easily accessible information regarding the actors involved and the processing of their data;

PR_7 – *Security*: Technical and administrative measures shall be taken to protect personal data from unauthorized access and accidental or illegal destruction, loss, modification, communication or propagation;

PR_8 – *Prevention*: Adoption of preventative measures to avoid the occurrence of harm from personal data processing;

PR_9 – *Non-discrimination*: Prohibition to perform data processing for discriminatory, illegal or abusive goals;

PR_{10} – *Accountability*: Demonstration by the data processor of the adoption of effective measures, capable of proving the observance and fulfilment of data protection norms, as well as the effectiveness of those measures.

These Privacy Requirements (PR_i) will serve as a basis to identify the Threats (T_i) to the SG system and subsequently propose Countermeasures (C_i) to mitigate them.

6.3 Identify the need for a PIA

The need for a PIA can be integrated into usual business processes, discovered through screening questions [34] or required by law [5].

The screening questions proposed in the PIA process would allow for staff who are not experts on the subject to determine the need for more careful privacy consideration. While most projects would benefit from a systematic analysis of their usage of personal data, projects that deal with sensitive data in more intrusive ways would benefit from full scope PIAs [34].

The guidance offers screening questions to check if the project involves the collection of new information about individuals; if it will compel individuals to provide information about themselves; if this information will be disclosed to organizations who previously did not have access to it; if this information will be used for purposes other than the current; if the project involves using new technology which may be seen as privacy intrusive; if decisions or actions will be taken against individuals based on this information or if the information is sensitive, stating that, even if only one of those conditions is true, a PIA should be taken[34].

Furthermore, PIAs may be required by data protection law. Article 35 of the GDPR states that when a data processing activity *"is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data"* [5]. Brazilian General Data Protection Law defines Data Protection Impact Reports but does not describe situations where it would be mandatory, relying on the Data Protection Authority (DPA) to require them [4].

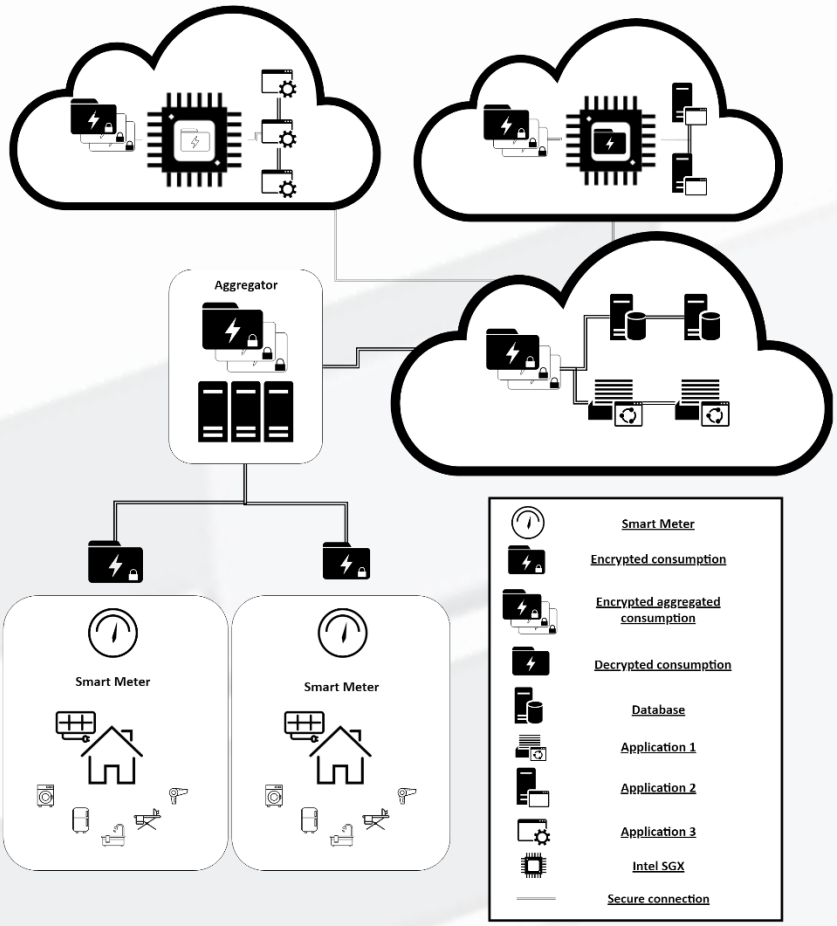

Figure 1: SecureCloud information flow diagram

As stated in section 4, smart metering poses great privacy risk and, in accordance with ICO's screening questions, its deployment would require a PIA.

6.4 Describe the information flows

The SecureCloud high-level data model comprises the following parties (figure 1):

(I) Consumer: The data subject. The Smart Meter is located on the consumer premises and reports information enough to provide the consumer with its correct bill amount;

(II) Tamper-proof Smart Meter: The SM cannot be tampered and can encrypt and sign the consumption information;

(III) Aggregator: The data controller. The system responsible for the Metering Data Collection (MDC) and grid operations that use aggregated data, such as real-time load management. Security, control and management responsibilities may rely on the Utility or may be shared with a trusted third party.

(IV) The Cloud: The data processor. Infrastructure as a Service (IaaS) contracted on public clouds. The service model is a requirement to access the cryptographic properties of the processor. Depending on the desired application, the management and control may rely on the Utility or on a subcontractor with access to a limited set of data, the least amount necessary to perform its functions.

The SM (II) measures the consumption of the Consumer (I) and communicates the necessary information to the Aggregator (III) who may store, process and/or relay this information to the cloud (IV) for further processing, archival or deletion. The detailed consumption is encrypted at the meter level (II) and only decrypted if/when necessary, inside the processor enclave (IV).

6.5 Identify privacy threats

In this step, the threats (T) were identified and related to each of the Privacy Requirements (PR), as explained in the methodology section. A non-exhaustive list of threats gathered from the literature are:

T1 PR1 (*Purpose binding*) – Threats against this privacy requirement are actions that seek to achieve secondary uses of the data.

$T_{1,1}$: The Utility provides access to a third-party to some dataset for marketing purposes;

$T_{1,2}$: The Utility, believing the dataset to be anonymized, does not inform the data subjects of a data-sharing agreement with a third-party, who has other datasets and eventually deanonymizes the provided set;

$T_{1,3}$: The Utility shares data with a subcontractor for operational purposes without enforcing the purpose limitation.

T2 PR2 (*Adequacy*) – Threats against this PR happen when the processing pushes the boundaries of the agreed processing, falling out of the consented scope.

$T_{2,1}$: The Utility claims to process data for operational purposes, doing so through the analysis of identified behavioural patterns.

T3 PR3 (*Necessity*) – Actions that collect or process excessive data in relation to the desired goal.

$T_{3,1}$: The Utility receives, and stores identified detailed energy consumption profiles;

$T_{3,2}$: The Utility attaches location and identity information to usage patterns;

$T_{3,3}$: The Utility processes identified detailed energy consumption to provide the billing.

T4 PR4 (*Access*) – Actions that happen to obstruct the facilitated and free access to self-information.

$T_{4,1}$: The Utility takes a long time to reply to a Data Access Request or delivers an incomplete dataset.

T5 PR5 (*Data accuracy*)

$T_{5,1}$: The Utility does not record a contact information change and discloses private information to someone other than the data subject.

T6 PR6 (*Transparency*)

$T_{6,1}$: The consumer finds that their data is being processed in a previously unknown way;

$T_{6,2}$: The consumer is redirected to multiple entities, previously unknown, when trying to access their data.

T7 PR7 (*Security*)

$T_{7,1}$: An attacker gains low-level credentials and escalates in the system, accessing personal data;

$T_{7,2}$: An attacker impersonates the Utility and phishes the consumer into providing data;

$T_{7,3}$: A third party with access to the utility system escalates privileges to access private information;

$T_{7,4}$: A vulnerability exposes some database to the internet and the data exfiltration goes unnoticed.

T8 PR8 (*Prevention*)

$T_{8,1}$: The Utility waits for an incident to act;

$T_{8,2}$: The Utility deploys systems that "collect first, process later" and implements privacy through add-in solutions, policy, access and control mechanisms.

T9 PR9 (*Non-discrimination*)

$T_{9,1}$: The Utility charges different tariffs based on the neighbourhood where the consumption happens;

T10 PR10 (*Accountability*)

$T_{10,1}$: The dataflow and preventative actions are not documented;

$T_{10,2}$: The systems do not provide logs.

The identified threats (T) to the privacy requirements (PR) were used to inform the countermeasures (C) that aim at mitigating them.

6.6 Identify privacy solutions

Next, countermeasures to each of the identified threats are offered, with their corresponding notation (e.g., $C_{3,2}$ corresponds to the countermeasure for the threat $T_{3,2}$)

$C_{1,1}$: Due to purpose binding obligations, the Utility must process the data following its purposes; those do not include targeted advertisement;

$C_{1,2}$: Truly anonymized datasets are hard to achieve, namely through careful application of differential privacy. Perform adversarial calculations to determine whether the dataset could be shared and seek consent to do so;

$C_{1,3}$: Data sharing agreements must be carefully vetted and only carried if the parties have effective means to enforce the adherence to the terms;

$C_{2,1}$: Establish processes to determine if the processing is happening adequately and as expected;

$C_{3,1}$: Establish procedures for de-identification at source and schedule timely deletion of data;

$C_{3,2}$: Do not correlate the data to obtain personal information;

$C_{3,3}$: Design the system to operate without the need to access detailed consumption profiles;

$C_{4,1}$: Establish processes to comply with user access requests: document where all the data may be and have a system in place to easily organize and provide it to the consumer;

$C_{5,1}$: Establish processes to flag version changes and provide logs;

$C_{6,1}$: Have clear, public and documented procedures for data management;

$C_{6,2}$: Document and publish the entities involved in the data chain. Have a united front to deal with customers and relay the requests if/when necessary;

$C_{7,1}$: Implement strong access controls where a compromised credential does not equate a compromised system;

$C_{7,2}$: Document and publish the default and expected interactions with the customer. Offer a channel for the customer to validate claims that could be made by the Utility (e.g. a phone number to verify if the operator claiming to represent the utility works there);

$C_{7,3}$: Allow access to internal systems only for vetted institutions; Establish Identity Management Systems that default to providing least authority access;

$C_{7,4}$: Have warrant canaries to denounce unexpected behaviour on important systems and have incident response procedures in place to rapidly deal with the situation;

$C_{8,1}$: Take preventative action prior to security incidents. Define most important assets (your "crown jewels") and closely monitor them for an unexpected activity or access (warrant canaries);

$C_{8,2}$: In accordance with PR_3 (Necessity), do not collect any unnecessary data; Privacy by policy is not desirable, even when the policies are enforced if a breach happens, there is harm;

$C_{9,1}$: Build systems that de-identify at source;

$C_{10,1}$: Document and publish the procedures taken to protect privacy;

$C_{10,2}$: Enable logging to facilitate incident response and attribution.

These countermeasures address the uncovered threats according to the listed Privacy Requirements. The assessment conducted early on the system development makes architectural choices easier and cheaper to make, acting in a preventative manner to avoid harm.

The propositions contained in this section have the overarching goal of enabling systems in line with the principles of Privacy by Design. Focusing especially on the privacy requirements of Necessity (PR_3) and Purpose Binding (PR_1) this work sought to achieve the data minimization principle, considered to be essential for Privacy by Design. The principles of Access (PR_4) and Transparency (PR_6) are important. However, these mechanisms do not mitigate the risks from massive collection and concentration of data, especially on an attractive target such as SM information [13]. Given that reidentification research is a flourishing field and authors have continuously demonstrated the ineffectiveness of anonymization [23] [24], the least data collected the better.

Systems that operate without access to detailed energy consumption are feasible [29] [39] [46], so a naïve implementation of the SG that acquired this information, even claiming to anonymize it later, would disrespect the privacy requirement, and legal principle, of Necessity.

7 Conclusion

The goal of this paper was to conduct a PIA process on a Smart Grid operating with cloud computing to facilitate the design of private systems that prevent harm from happening. An advanced Cloud Privacy Threat Modeling methodology was chosen to help systematize the process and uncover domain-specific threats.

By selecting Privacy Requirements from data protection legislation and conducting the assessment on the early stages of the system, privacy threats were uncovered, and countermeasures were offered before further development of the system, allowing for design choices that favour privacy architectures.

The methodologies used suggest a consultation with domain experts and stakeholders. These actors were not consulted and here lies one of the limitations of this paper. Industry-specific expertise could uncover threats that are not discussed in academic papers and offer a better perspective on the real-life applicability of the proposed countermeasures. Another limitation has to do with the pace of academic publishing (v. the pace of exposed vulnerabilities and attack scenarios discussed in technical conferences). To address this issue, and to account for the fact that privacy and security are dynamic domains, a good posture is to continuously monitor published vulnerabilities (CVEs) against components of the system.

8 Acknowledgements

EU-BR SecureCloud project has been receiving funds granted from the 3rd EU-BR Coordinated Call (Brazilian Ministry of Science, Technology and Innovation, MCTIC/RNP BR grant agreements #2549) and European Union's Horizon 2020 research and innovation programme - EU grant agreement # 690111).

References

[1] European Commission *et al.*, *Smart grid projects in Europe: lessons learned and current developments : 2012 update.* Luxembourg: Publications Office of the European Union, 2013.

[2] M. R. Asghar, G. Dan, D. Miorandi, and I. Chlamtac, "Smart Meter Data Privacy: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2820–2835, 2017.

[3] S. Rusitschka, K. Eger, and C. Gerdes, "Smart Grid Data Cloud: A Model for Utilizing Cloud Computing in the Smart Grid Domain," in *2010 First IEEE International Conference on Smart Grid Communications*, Gaithersburg, MD, USA, 2010, pp. 483–488.

[4] Brasil. Lei Geral de Proteção de Dados Pessoais (LGPD) [Online]. Available: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. [Accessed: 10-Jul-2018].

[5] *EU General Data Protection Regulation (GDPR):* Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and

on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1.

[6] A. E. Waldman, "Privacy's Law of Design," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3263000, Oct. 2018.

[7] H. Y. Lam and G. S. K. Fung, "A Novel Method to Construct Taxonomy of Electrical Appliances Based on Load Signatures," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 2, p. 9, 2007.

[8] E. McKenna, I. Richardson, and M. Thomson, "Smart meter data: Balancing consumer privacy concerns with legitimate applications," *Energy Policy*, vol. 41, pp. 807–814, Feb. 2012.

[9] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "Non-Intrusive Load Monitoring Using Prior Models of General Appliance Types," *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, p. 7, 2012.

[10] J. E. Wynn, "Presentation-Threat Assessment & Remediation Analysis (TARA) Methodology Overview," Feb. 2015.

[11] C. Efthymiou and G. Kalogridis, "Smart Grid Privacy via Anonymization of Smart Metering Data," 2010, pp. 238–243.

[12] G. M. Minamizaki, K. V. O. Fonseca, S. Clauß, E. Franz, S. Köpsell, and H. Lazarek, "Data security issues on metering systems of energy consumption in Brazil," *ESPAÇO ENERGIA*, no. 18, p. 11, 2013.

[13] S. Gurses, C. Troncoso, and C. Diaz, "Engineering Privacy by Design," p. 25, 2015.

[14] Tancock, D., Pearson, S., & Charlesworth, A. (2013). A privacy impact assessment tool for cloud computing. In *Privacy and security for Cloud computing* (pp. 73-123). Springer, London.

[15] C. Cuijpers and B.-J. Koops, "Smart Metering and Privacy in Europe: Lessons from the Dutch Case," in *European Data Protection: Coming of Age*, S. Gutwirth, R. Leenes, P. de Hert, and Y. Poullet, Eds. Dordrecht: Springer Netherlands, 2013, pp. 269–293.

[16] R. Hoenkamp, G. Huitema, and A. de Moor-van Vugt, "The Neglected Consumer: The Case of the Smart Meter Rollout in the Netherlands," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2061668, Nov. 2011.

[17] P. Mell and T. Grance, "The NIST Definition of Cloud Computing," National Institute of Standards and Technology, NIST Special Publication (SP) 800-145, Sep. 2011.

[18] I. Rubinstein and N. Good, "Privacy by Design: A Counterfactual Analysis of Google and Facebook Privacy Incidents," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2128146, Aug. 2012.

[19] B. Stewart, "Privacy impact assessment: some approaches, issues and examples," *Assistant Commissioners Office of the Privacy Commissioner, New Zealand*, 2002.

[20] D. Wright, "The state of the art in privacy impact assessment," *Computer Law & Security Review*, vol. 28, no. 1, pp. 54–61, Feb. 2012.

[21] C. Greer *et al.*, "NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 3.0," National Institute of Standards and Technology, NIST SP 1108r3, Oct. 2014.

[22] P. De Hert, "A Human Rights Perspective on Privacy and Data Protection Impact Assessments," in *Privacy Impact Assessment*, D. Wright and P. De Hert, Eds. Dordrecht: Springer Netherlands, 2012, pp. 33–76.

[23] S. S. Shapiro, "Situating Anonymization Within a Privacy Risk Model," Sep. 2013.

[24] A. Narayanan, J. Huey, and E. W. Felten, "A Precautionary Approach to Big Data Privacy," in *Data Protection on the Move*, vol. 24, S. Gutwirth, R. Leenes, and P. De Hert, Eds. Dordrecht: Springer Netherlands, 2016, pp. 357–385.

[25] R. Leszczyna, "Cybersecurity and privacy in standards for smart grids – A comprehensive survey," *Computer Standards & Interfaces*, vol. 56, pp. 62–73, Feb. 2018.

[26] M. H. F. Wen, K.-C. Leung, V. O. K. Li, X. He, and C.-C. J. Kuo, "A survey on smart grid communication system," *APSIPA Transactions on Signal and Information Processing*, vol. 4, 2015.

[27] Y. Yan, Y. Qian, H. Sharif, and D. Tipper, "A Survey on Smart Grid Communication Infrastructures: Motivations, Requirements and Challenges," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 5–20, 2013.

[28] D. A. D. Schlichting, "Assessment of Operational Energy System Cybersecurity Vulnerabilities," May 2018.

[29] A. Rial and G. Danezis, "Privacy-Preserving Smart Metering," *Proceedings of the 2011 ACM Workshop on Privacy in the Electronic Society*, p. 12, 2011.

[30] I. M'rhaoaurh, C. Okar, A. Namir, and N. Chafiq, "Challenges of cloud computing use: A systematic literature review," *MATEC Web of Conferences*, vol. 200, p. 00007, Jan. 2018.

[31] M. Deng, K. Wuyts, R. Scandariato, B. Preneel, and W. Joosen, "A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements," *Requirements Eng*, vol. 16, no. 1, pp. 3–32, Mar. 2011.

[32] A. Gholami, A.-S. Lind, J. Reichel, J.-E. Litton, A. Edlund, and E. Laure, "Design and Implementation of the Advanced Cloud Privacy Threat Modeling," *International Journal of Network Security & Its Applications*, vol. 8, no. 2, pp. 103–122, Mar. 2016.

[33] A. Gholami, A.-S. Lind, J. Reichel, J.-E. Litton, A. Edlund, and E. Laure, "Privacy Threat Modeling for Emerging BiobankClouds," *Procedia Computer Science*, vol. 37, pp. 489–496, 2014.

[34] Information Commissioner's Office (ICO), *Conducting privacy impact assessments code of practice*. v. 1.0. Technical report, Information Commissioner's Office (ICO), UK, 2014.

[35] Information Commissioner's Office (ICO), Privacy Impact Assessment Handbook, Wilmslow, Cheshire, December 2007, Version 2.0, June 2009.

[36] A. Gholami and E. Laure, "Advanced Cloud Privacy Threat Modeling," in *Computer Science & Information Technology (CS & IT)*, 2016, pp. 229–239.

[37] ETP Smart Grids. "The SmartGrids European Technology Platform". 2006. Available: http://www.smartgrids.eu/ETPSmartGrids

[38] European Commission *et al.*, *Smart grid projects in Europe: lessons learned and current developments : 2012 update.* Luxembourg: Publications Office of the European Union, 2013.

[39] M. Jawurek, M. Johns, and F. Kerschbaum, "Plug-in privacy for Smart Metering billing," *arXiv:1012.2248 [cs]*, Dec. 2010.

[40] E. L. Quinn, "Privacy and the New Energy Infrastructure," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 1370731, Feb. 2009.

[41] Hargreaves, T., Nye, M., Burgess, J., 2010, Making energy visible: A qualitative field study of how householders interact with feedback from smart energy monitors. Energy Policy, 38, 6111-6119

[42] E. McKenna, I. Richardson, and M. Thomson, "Smart meter data: Balancing consumer privacy concerns with legitimate applications," *Energy Policy*, vol. 41, pp. 807–814, Feb. 2012.

[43] P. McDaniel and S. McLaughlin, "Security and Privacy Challenges in the Smart Grid," *IEEE Security & Privacy Magazine*, vol. 7, no. 3, pp. 75–77, May 2009.

[44] A. Cavoukian, J. Polonetsky, and C. Wolf, "SmartPrivacy for the Smart Grid: embedding privacy into the design of electricity conservation," *IDIS*, vol. 3, no. 2, pp. 275–294, Aug. 2010.

[45] M. Jawurek, "Privacy in smart grids," 2013. Available: https://opus4.kobv.de/opus4-fau/files/3645/Dissertation_MarekJawurek_fuer_finalen_Druck_1.0.pdf

[46] R. J. Riella *et al.*, "Securing Smart Metering applications in Untrusted Clouds with the SecureCloud Platform," in *Proceedings of the 1st Workshop on Privacy by Design in Distributed Systems - W-P2DS'18*, Porto, Portugal, 2018, pp. 1–6.

[47] Creese, S., Goldsmith, M., & Hopkins, P. (2013). Inadequacies of current risk controls for the cloud. In *Privacy and Security for Cloud Computing* (pp. 235-255). Springer, London.

[48] S. Pearson and A. Benameur, "Privacy, Security and Trust Issues Arising from Cloud Computing," presented at the Proceedings - 2nd IEEE International Conference on Cloud Computing Technology and Science, CloudCom 2010, 2011, pp. 693–702.

[49] B. Grobauer, T. Walloschek, and E. Stocker, "Understanding Cloud Computing Vulnerabilities," *IEEE Security Privacy*, vol. 9, no. 2, pp. 50–57, Mar. 2011.

[50] F. Kelbert *et al.*, "SecureCloud: Secure big data processing in untrusted clouds," in *Design, Automation Test in Europe Conference Exhibition (DATE), 2017*, 2017, pp. 282–285.

[51] Habibzadeh, H., Nussbaum, B. H., Anjomshoa, F., Kantarci, B., & Soyata, T. (2019). A survey on cybersecurity, data privacy, and policy issues in cyber-physical system deployments in smart cities. Sustainable Cities and Society, 50, 101660. Available: https://doi.org/10.1016/J.SCS.2019.101660