A Privacy by Design Methodology Application in Telecom Domain

Siham Arfaoui [1], Abdelhamid BelmekkI[1] and Abdellatif Mezrioui [1]

[1]National Institute of Posts and Telecommunications, Morocco

Abstract: Telecommunication has been considerably developed over the last decades, notably through optical fiber submarine cables and wireless connections offering voice and data wide range services. Telecommunication infrastructures are the necessary backbone that make possible any voice and data exchange. Unfortunately, these infrastructures are still suffering from various vulnerabilities and continue to be target of specific cyber-attacks. Some of these attacks could lead to service deniability, integrity and privacy loss. Against this fact, it's also established that telecom service providers, as the owner of this infrastructure, can have access to huge data, even personal data related to customer and to their employees. Basically, this personal data is related directly to the customer's and employee's identity, geolocation, interest areas and contact circle, etc., when it comes to the use of this personal data, the privacy concerns become a big challenge for telecom service providers due to heavy impact that can induce. Given the personal data protection criticality in telecom domain, privacy by design PbD should be incorporate. Then, this article aims to apply in telecom service providers ISPM methodology "Information System Privacy Methodology" which focuses on PbD implementation in enterprises architecture, specifically in information systems taking into account all technical and organizational aspects.

Keywords: Privacy by Design PbD, Telecommunication Domain, Information System Privacy Methodology ISPM, Threat Modeling, Personal Data, Togaf, Information Systems, Charging data Records CDR.

1. Introduction

The telecommunications industry is nowadays a complex value chain that includes equipment and platform suppliers, service providers and users... This value chain begins with building blocks such as semiconductor chips and software that are incorporated into equipment purchased by service providers and users. In similar way, service providers, build networks and sell to the end users voice and data telecommunications services. End users could include individuals, or enterprises, who subscribe to various services such as landline, cellular and high-speed Internet access or other added value services. The broad definition of telecommunications means that every layer involved in remote communication becomes, at least partially, part of the telecommunications industry. The wide range of contributors in telecommunications are listed in the following list:

- Network users
- Service providers
- Equipment and platform suppliers
- Software suppliers of equipment and platforms
- Semiconductor manufacturers
- Application developers

However, the relationship between the telecom network provider and the end user is different from that between consumers and equipment and platform suppliers. These last two parts create products that will be used by an unknown user. Service providers, on the other hand, are able to have a more « customized » relationship [1] with their customers according to established agreement. It is for this reason that we focus on service providers personal data protection. In the telecommunications sector, and more specifically among service providers, there are various sources of personal data. Figure 1 below summarizes them [2]:

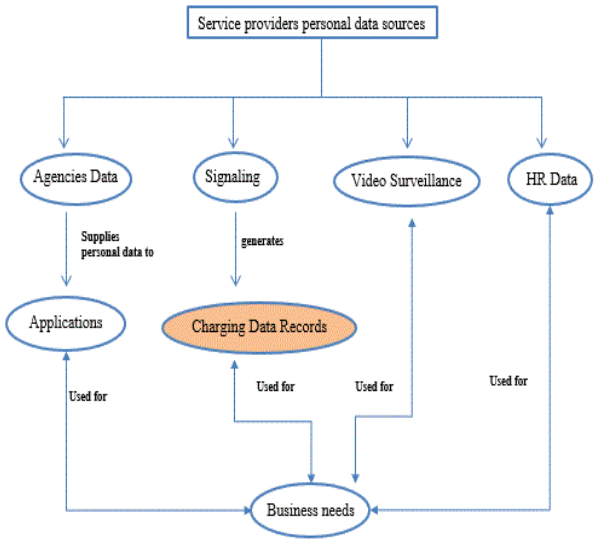


Figure 1. Service providers personal data sources

Personal data is collected from different service providers sources:

- Personal data collected by agencies when drawing up subscription contracts for any services like voice or data services. This personal data is used by different applications serving business needs.
- Personal data from voice and data signaling. A lot of equipment and platforms used to carry voice and data service produce recoverable signals in the form of Charging Data Records CDR. CDRs [2] are a formatted collection of information about a chargeable event (eg, call setup time, call duration, amount of data transfer, etc.)
- HR data: represent personal data gathered by HR department of service provider employees.
- Video surveillance data: which represent all video surveillance stream from all camera deployed within different location.

According to GDPR General Data Protection Regulation personal data are considered as any data that can help to identify a person. Specifically, an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person .

Regarding the premise of this paper, where the focus is more oriented to telecommunication domain, we can draw up list of personal data collected and available in telecom service providers as follows:

- Last name /First Name
- Personal photos
- Identity number
- Date of Birth
- Address
- E-mail
- Curriculum Vitae CV
- Training
- Diplomas
- Professional experiences
- Banking information
- Salary
- Debts
- Union membership
- Health data
- Data relating to offenses and convictions
- Calling party number
- Called party number
- IMSI calling party: International Mobile Subscriber Identity
- IMSI called party: International Mobile Subscriber Identity
- Calling location area code party
- Called location area party
- Calling cell ID : cellule Identity
- Called cell ID
- IMEI: International Mobile Equipment Identity
- URL: Uniform Resource Locator
- IP address
- Website visited

Giving the various personal data sources, the deal with privacy protection issue should not only consider a separate software system but rather the entire information system IS including hardware IT infrastructure. Thus, integrating PbD into enterprise architecture in a comprehensive way requires a coherent approach which must be adapted to the specifications of PbD and must also take into consideration all IS layers such as process, data flow, applications and technical architecture. The available methodologies in literature [3, 4, 5, 6; 7] are dedicated to PbD implementation only in software systems which we consider as limitation. Indeed, software systems are just one of the building blocks of an IS and by the way is not the only block that should be considered. Hence, in this paper, we aim to use ISPM « Information System Privacy Methodology » [8] which aims privacy by design implementation in IS of any domain. In the following sections, the authors present requirements that ISPM must meet. They then synthesize the PbD implementation available in the literature. Thus, the ISPM methodology is synthetised before applying ISPM methodology step by step in detailed way from prerequisites to the solution implementation. We end this work by the result discussion.

2. Background and Related Works

2.1. Privacy by design

PbD or privacy from the design stage is a system engineering approach that considers privacy throughout the process [9]. The close alignment between the “Security by Design” and “Privacy by Design” work was introduced in January 2013 [10]. Privacy by design is based on seven fundamental principles [11]:

- Proactive not reactive: it seeks to anticipate and prevent privacy-invasive events before they happen by not waiting for privacy risks to materialize.
- Privacy as the default setting: it seeks to build privacy measures directly into any given information, communication technology system and business practice by default.
- Privacy embedded into design: it seeks to embed privacy into IS design and the architecture and business practices. It does not bolt it on after the fact.
- Positive-Sum: it seeks to accommodate all legitimate interests and objectives in a positive-sum “win-win” way, not through a zero-sum approach involving unnecessary trade-offs.
- End-to-End security: it seeks to ensure cradle-to-grave secure information lifecycle management, end-to-end.
- Visibility and transparency: seeks to it keep IS component parts and business practices operations visible and transparent to users.
- Respect for user privacy: it seeks to respect and protect individual interests, above all. It keeps it user centric.

2.2. Information System privacy methodology: ISPM

2.2.1. ISPM Requirements

The objectives of ISPM methodology [8] are:

- Identify and census of all personal data that is manipulated in the IS in question.
- Identify the actors who will be able to manipulate personal data and grant them the appropriate rights to do so.
- Prevent other actors from manipulating personal data by establishing control rules and using adequate anonymization techniques.
- Also, IS evolves and therefore the methodological approach must take into account this fact of scalability and be invoked continuously.

ISPM [8] assume that the ISs could be in one of the following states:

- They are already designed and deployed and do not comply with the PbD principles, therefore they must be adapted to comply with this paradigm.

- They are to be designed from scratch respecting the PbD principles.
- They are already designed and comply with PbD principles but they must evolve while respecting the PbD principles.

To summarize, ISPM [8] is:

- PbD driven methodology, which implies taking into account all PbD principles.
- Holistic, so it must take into account the process layer, data flow layer, the application layer and the technical architecture.
- Integrated with the IS design method used in the company and / or proposes generic and standardized models.
- In addition, it should carry out a privacy treats analysis, their impact and occurrence degrees.
- Ultimately, it should propose strategies for resolving threats or at least for reducing them considerably.

2.2.2. ISPM steps

In what follows, the authors present the ISPM methodology approach [8]. It consists of six steps as shown in Figure 2:

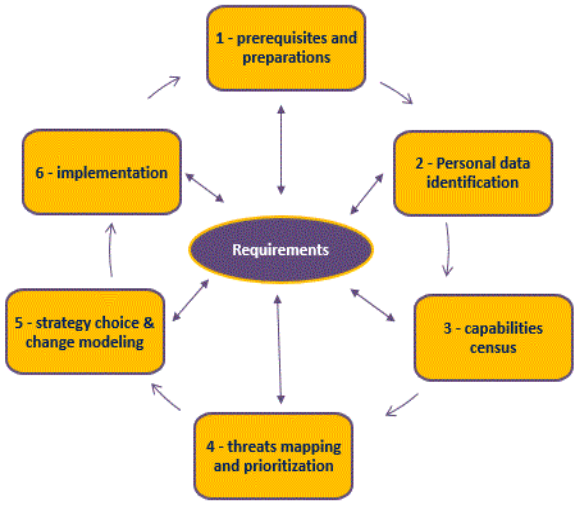


Figure 2. ISPM methodology

The first step objective is to determine the expectations from ISPM implementation. It also serves to define the context and the perimeter of the action field. Thus, the stakeholders are identified, also the constraints are synthesized and the recommendations are established. In step 2, an inventory of all personal data, sensitive personal data and quasi-identifiers is carried out concerning the chosen perimeter or in the entire IS. The input of this step is data classification according to personal data protection laws. As for its output, it represents all of personal data and processes of the chosen scope. The purpose of step 3, for its part, is to identify all the chosen perimeter capabilities. These are the different processes, the data flows, the application layer, the technical architecture and related stakeholders that interact with personal data. The goal is to have an understanding of what exists to know where to act. The input of this third step is existing knowledge based on the paperwork, employees, norms and standards of the company activity field. As for output, it is a set of standardized diagrams that have been carefully chosen by ISPM from the set of diagrams proposed by Togaf. The fourth step aims to rule on all threats of the study scope. These threats prioritization is subsequently developed. The input to this step is the trees proposed by Linddun and completed by the authors adding threats tree. As for output, it represents all the threats to be prioritized and for which the appropriate strategy must be determined to remedy them. It is at step five that the strategy to be followed is decided and also the whole change to operate on the different processes, the data flows, the application layer, the technical architecture and the rights of the stakeholders are ruled. The choice of appropriate privacy enhancing technologies to implement the changes agreed in the different bricks of the perimeter constitutes the sixth and last step of ISPM approach using the matrix made available for this purpose.

2.3. Related Works

The personal data protection concerns by using privacy by design paradigm have been investigated by many authors. The literature focuses on implementing PbD in the following areas:

- Surveillance cameras [12,13,14,15].
- Biometrics [16,17].
- Smart Grid [18,19,20].
- RFID and sensor technology [21,22].
- Health care [23,24,25,26].
- Big data and cloud [27].
- Mobile device and communication [28,29,30,31].

In mobile device and communication, [28,29,30,31] identify telecom stakeholders interacting with personal data. [30] indicates that service providers have the most interaction with personal data. Strategic recommendations to all telecom stakeholders: developers, service providers and mobile device manufactures are mentioned. The focus in articles dealing with privacy by design in the telecom domain is on Wi-Fi localization, IP and video surveillance. For the Wi-Fi localization, the proposed works seek to modify the technology protocol [29]; others encourage adopting privacy by design [31] without giving the way to implement it. For the IP localization, the risk dealt with is related to the applications developers but does not address the risk of insider employee's service providers. With regard to video surveillance and biometrics data, the works carried out [12,13,14,15] are very advanced and propose an end-to-end implementation of Privacy by Design.

In this article, the authors aim to implement PbD principles in the IS of service providers according to ISPM methodology [8], which allow to control insider employees and external skeakholders privacy risks. Indeed, the security problems of telecommunications networks have preoccupied the international community for the past two decades. Telecommunications infrastructures that provide the necessary backbone for the exchange of information such as voice, video, data and Internet connectivity have proven to be particularly vulnerable to various forms of attack. Some of these attacks could lead to a denial of service, loss of integrity and confidentiality of the telecom operator's network services. Protecting these networks against internal or external attacks is therefore an important aspect that cannot be ignored.

So, it has never been more important for telecommunications operators to make data privacy and security a top priority. In fact, telecom operators are sitting on a veritable gold mine of personal data. These relate directly to the identity of the customer, but also to his geolocation, his centers of interest and his circle of relations. A recent survey of digital media trends from Deloitte [32] shows that consumers continue to fear identity theft, financial loss and unauthorized use of their sensitive data, largely because so many between them were directly confronted with these threats. In fact, 23% of US households were victims of cybercrime in 2018. As a result, consumers increasingly demand a reassuring level of control over their personal data. Therefore, telecom companies have to work hard to create digital environments where people feel safe.

3. ISPM Application Step by Step

3.1. Step 1: prerequisites and preparations

In this first step, we have to decide on the points mentioned in figure 3:

Figure 3. points to be decided on stage 1 of ISPM

Regarding the perimeter of the IS on which the authors apply ISPM and given the panoply of personal data sources existing at service provider, the perimeter that the authors have chosen is Charging Data Records. This choice is justified by the fact that the Charging Data Records CDRs provide personal data information on a large number and can be considered as the most important personal data source of telecom provider that inform about : calling party number, called party number, IMSI calling party: international mobile subscriber identity, IMSI called party, calling location area code party, Called location area party, calling cell ID : cellule Identity, Called cell ID, IMEI: International Mobile Equipment Identity, URL: uniform resource locator, IP address and website visited.

CDRs are generated by all telecom equipment and platforms providing voice and data services. The authors will not act on the equipment generating them in this article but rather on all the business processes that consume CDRs. CDRs are a formatted collection of information about a chargeable event (e.g. time of call setup, duration of the call, amount of data transfer, etc.). For each paid event, a CDR is generated. Using the operator's network, whether for a postpaid or prepaid customer, the generated CDRs are then referred to mediation. In a telecom-billing scenario, mediation is the first step after receiving a CDR. Mediation is a process that converts CDRs into a predefined form and is then imported by a specific billing system. Despite the name, all data transferred through mediation platforms is not actually used just for billing purposes. For example, CDRs help to establish statistics. These statistics are then used for capacity planning; CDRs are also used as part of a network monitoring procedure or for any other business intelligence application and for insurance revenue needs.

The authors conclude the processes that interact with CDR Charging Details Records are [2]:

- Mediation
- Billing
- Quality of service
- Income Insurance
- Business intelligence

Regarding the methods and tools to be used in the privacy protection implementation in the above-mentioned IS service provider processes, IS modeling is required. Historically, it was not until the 1980s to see the emergence of the first methods and concepts of enterprise architecture modeling which is Zachmman. Over time, Zachmman framework has inspired other methods. The authors will not detail all approaches of enterprise architecture modeling and their historical evolutions, nor will be able to make a complete list, because at least 80 architectural frameworks have been registered in the world [33]. Some have only one user, their designer only, while others have several tens of thousands of users.

For example, The Open Group, which publishes the TOGAF framework, indicates in real time the number of certified people. [33] Togaf success is due to its need satisfaction of having a common framework facilitating architectural practices capitalization, more specifically, Togaf is positioned as a generic method that brings together a set of techniques centered on the transformation of enterprise architecture [34]. This approach is called ADM architecture development method; it integrates different facets into its process: strategy, business, technique, governance and change planning. How to move from the initial architecture to the target architecture? The answer to this question lies at the heart of the Togaf framework. It offers a set of diagrams and matrices making it possible to target the changes to be adopted and to document them [33]. TOGAF standard reflects a structure of architectural capabilities and a content within a company. Capability is a capacity that an organization, person or system has. Capabilities are typically expressed in broad, high-level terms and typically require a combination of organization, people, process, and technology, such as marketing, customer contact, or outbound telemarketing. Some Togaf diagrams are minutely chosen by the authors in ISPM methodology.

The authors used Modelio as a tool for modeling Togaf artifacts. This tool provides a graphics editor that supports Togaf diagrams and matrices. It ensures the overall models consistency. The metamodel used by the authors is the EAP UML Unified Modeling Language Enterprise Architecture Profile, which extends UML notations to represent Togaf objects set. For the threats modeling trees, Microsoft power point is the tool that the authors used.

Moving on to strategic and operational objectives, the authors specified all the strategic and operational objectives that they have set and which motivate the ISPM application. Figure 4 represents an artefact proposed by Togaf that regroups them:

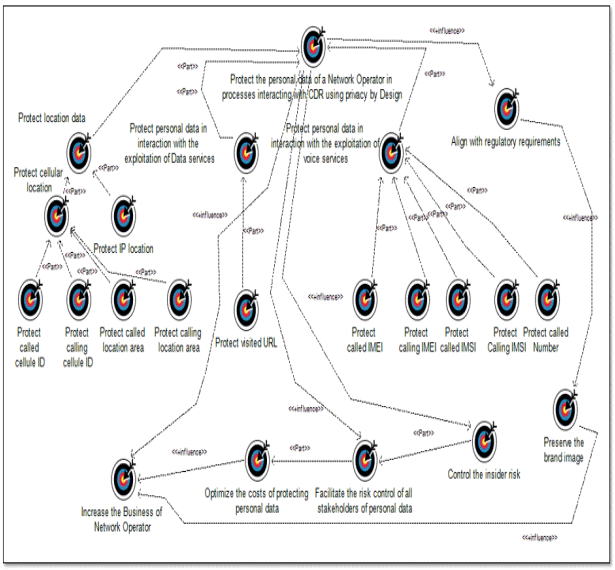


Figure 4. strategic and operational objectives

Concerning the actors, the authors highlighted in table 1 the actors who must be involved in the application of ISPM to protect personal data in the processes in interaction with the CDRs.

Table 1. The actors involved in ISPM application on the processes interacting with the CDRs

Participant	Skill	Decision power	Interest level
CEO: chief executive officer	Goal orientation, decisions	High	High
CIO : chief information officer	Information System IS project orientation, decisions	High	High
Information security officer	Personal data protection guidelines	High	High
Management control officer	Cost control to spend	High	Low
Quality officer	Documentation and standardization of internal processes	Middle	High
Regulatory officer	Control of the personal data protection regulatory requirements	High	High
Business architect	Mastery of the business architecture	Middle	High
Application architect	Mastery of the application architecture	Middle	High
Data architect	Mastery of the data architecture	Middle	High
Technical architect	Mastery of the Technical architecture	Middle	High
Billing manager	Billing process management	Middle	High
Revenue assurance manager	Revenue assurance management	Middle	High
Business intelligence manager	Business intelligence management	Middle	High
Mediation manager	Mediation management	Middle	High
Quality of service QoS manager	Quality of service QoS management	Middle	High
Billing employee	Performing billing tasks	Low	High
Revenue assurance employee	Performing revenue assurance tasks	Low	High
Business intelligence employee	Performing gence tasks	Low	High
Mediation employee	Performing mediation tasks	Low	High
QoS employee	Performing QoS tasks	Low	High
Security employee	Performing security tasks	Low	High
Data base administrator	Administration of the databases	Low	Middle
Systems and network engineer	Mastery of materials, systems and networks	Low	Middle
Support employee	Management of IS support tasks	Low	Middle
Developer	Solution development	Low	High

The governance committee is made up of the members cited in table 2:

Table 2. The governance committee

Participant	Skill	Decision power	Interest level
CEO	Goal orientation, decisions	High	High
CIO	Information System IS project orientation, decisions	High	High
Information security officer	Personal data protection	High	High

	guidelines		
Management control officer	Cost control to spend	High	Low
Quality officer	Documentation and standardization of internal processes	Middle	High
Regulatory officer	Control of the personal data protection regulatory requirements	High	High
Business architect	Mastery of the business architecture	Middle	High

3.2. Step 2: identification of the personal data correlated with the process in question.

At this stage, we must identify the personal data, the sensitive personal data and the Quasi-identifier of the chosen perimeter study in step 1. Remember that personal data are all data that can allow the direct or indirect identification of a person such as surname, first name, photos, social security number, telephone number, etc. Personal data can also be classified as sensitive data such as medical records, biometric records like DNA, fingerprints, iris, retina, voice, etc. In this category we can also include racial or ethnic origin, trade union membership, political opinion, etc. Quasi-identifiers, on the other hand, are non-explicit identifiers such as date of birth, sex, postal code, etc. These quasi-identifiers, if combined, can help infer the identity of the person in question.

In all that follows, the authors will use “personal data” to include personal data, sensitive personal data and Quasi identifiers. To develop the census of personal data, a series of meetings were held with the stakeholders. The results of these interviews are presented in Table 3:

Table 3. Personal data census matrix

Personal Data	Mediation	Billing	Quality of Service	Revenu assurance	Business Intelligence
Calling party number	•	•	•	•	•
Called party number	•	•	•	•	•
Calling party IMSI	•		•	•	•
Called party IMSI	•		•	•	•
Calling location area code party	•		•	•	•
Called location area party	•		•	•	•
Calling Cellule ID	•		•		•
Called Cellule ID	•		•		•
IMEI	•		•	•	•
URL	•		•		•
IP address	•		•		•

After establishing this correlation between the processes in question and personal data, the authors focused on identifying processes capabilities listed in the table 3.

3.3. Step 3: capabilities census

In step 3, the goal is to facilitate the understanding of the information system and allow the exhaustive inventory of the processes, data flows, applications and components of the technical architecture on which it is necessary to act in order to set up privacy by design. This purpose is ensured by the ability to model the processes collected from step 2 based on Togaf. The term "Capabilities" is borrowed from Togaf, it means the capacity that an organization, person or system possesses: "Capabilities" are generally expressed in general and high-level terms and generally require a combination of organization, people, process and technology [35]. Togaf has a panoply of diagrams and matrices called artefacts allowing modeling the capabilities; we have carefully chosen diagrams to ensure a modeling allowing identifying the different perspectives necessary for our goal. The chosen artefacts include a modeling of the processes, data flows, technical applications and architecture.

In what follows, we approached the modeling of diagrams carefully chosen by the ISPM methodology [8] from the range of diagrams proposed by Togaf [34]:The authors in this article constructed the following diagrams: Organizational Diagram, BPMN Business Process Diagram, Application

Communication Diagram, and Platform Decomposition Diagram.

- **Organizational diagram:**

In the following, the authors focused on the different stakeholders and their responsibilities. Figure 5 is dedicated to the specification of the participants and their responsibilities in this mission. The figure shows the actors of Table 1, highlighting the inter-stakeholder relationships in this project.

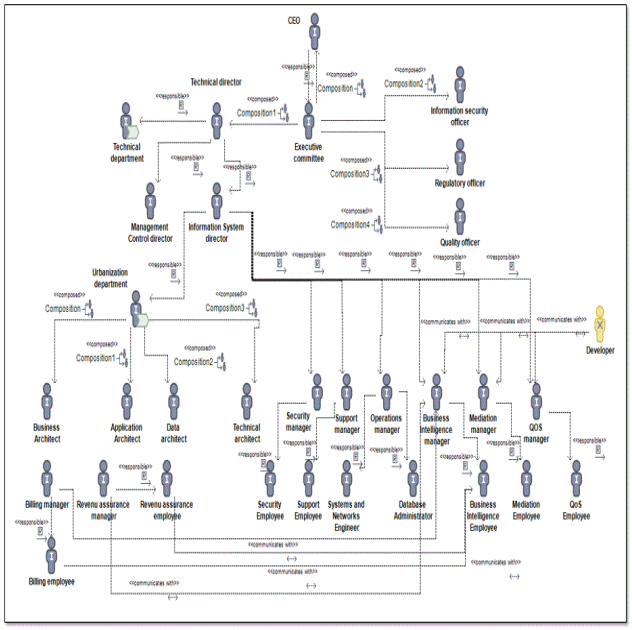


Figure 5. Organizational role diagram

- **BPMN business process diagram**

Togaf offers the possibility of using BPMN Business Process Model and Notation and UML graphs to model complementary perspectives. The authors opted for the BPMN process diagram because it meets the need to provide detailed visibility of the processes and their various interactions. The authors began by modeling the mediation process before it was adapted to the principles of privacy by design. It should be noted that only the use case of mediation in relation to the CDR is dealt with, provisioning is not mentioned.

Mediation is the bottleneck of personal data and it is at this level that personal data from CDRs is routed to the information system processes. Note that in figure 6 the support process has also been treated as we cannot treat the mediation separately from the support process.

As shown in Figure 7, the author dealt with the billing process and its modeling before adapting to the principles of privacy by design. Billing is one of the most important processes of the network operator. The called party, for example, is necessary for the preparation of the bill because the price of the voice communication depends on the destination of the call.

In what follows, the authors addressed the rest of the processes in interaction with CDR, which are quality of service, Business intelligence and revenue assurance. What is special about these processes is that they are all based on creating reports, identifying problems and solving them. Each of these three processes deal with specific aspects: QoS deals with reporting in relation to QoS performance. Revenue Assurance takes care of the reporting in relation to the loss of turnover and the business intelligence process is in charge of creating the reporting for these two areas, it is also in charge of the creation of reports for the evolution inventory, sale progress, subscription reports, CDR traffic, calls, roaming and interconnection report.

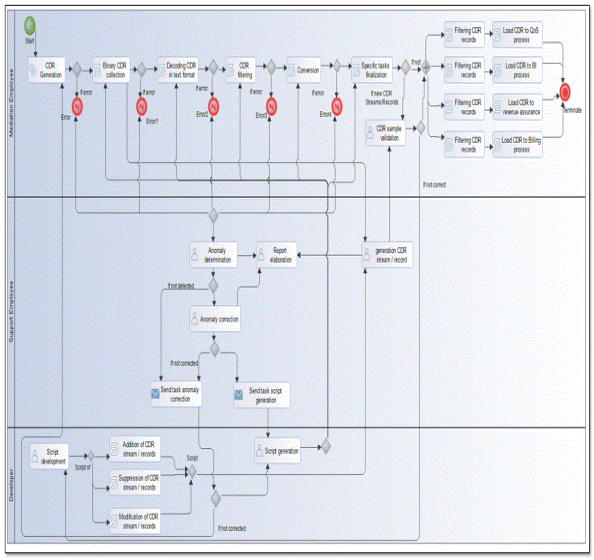


Figure 6. BPMN process diagram of mediation before application of ISPM

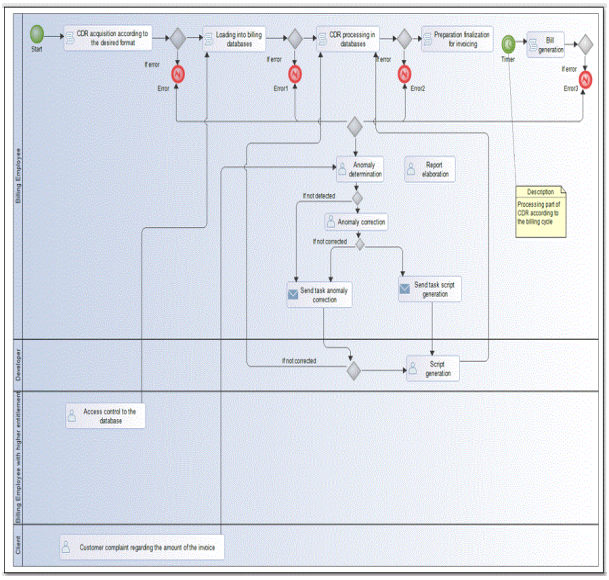


Figure 7. Billing BPMN Process Diagram before before application of ISPM

In order to model these three remaining processes interacting with CDR and seeing how similar they are, the authors modeled them all as a single graph. Therefore, “reporting” in Figure 8 refers to QoS or BI or Revenue Assurance report. User refers to the employee in the process for which the report was created. In the following, we present the BPMN process diagram of the three processes: Quality of Service, Revenue Assurance and Business Intelligence before the desired adaptation:

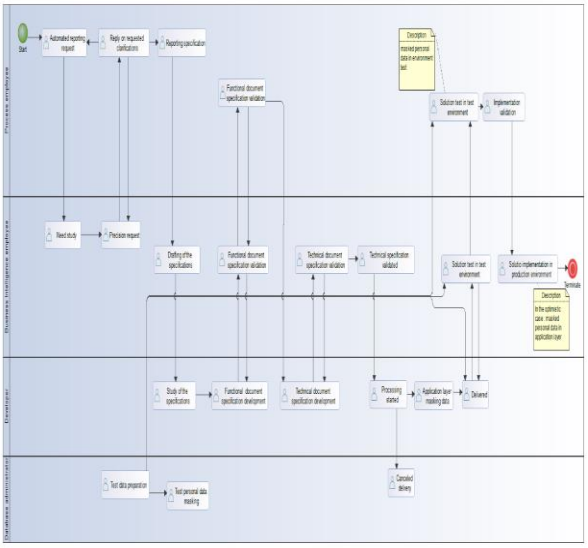


Figure 8. BPMN process diagram of BI, RA and QoSbefore application of ISPM

- **Application communication diagram**

Figure 9 groups together all the applications involved in the different processes of the studied scope. They are as follows: CDR files are routed from telecoms platforms in the form of flat files; they are conveyed to IME, which is a mediation ETL. In turn, IME dispatches the CDRs to SUBEX (revenue assurance application), MyCom (quality of service application), Sage (Business Intelligence application) and finally to Oracle for billing needs.

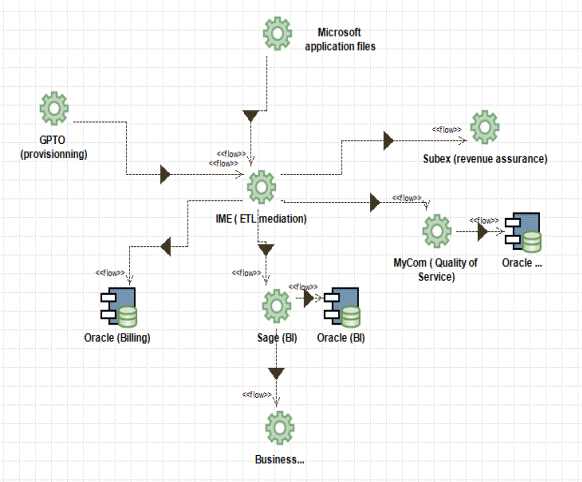


Figure 9. Application communication diagram before application of ISPM

- **Platform decomposition diagram**

Figure 10 shows the decomposition of the technical architecture before the application of ISPM, which summarizes the technical architecture of the applications cited in Figure 9:
Now, visibility all capabilities is possible. The authors did not model the use case diagram due to their mastery of the scope in question and thus the passage through the use case diagram to facilitate the modeling of the BPMN diagram was bypassed. In what follows, we will map the threats based on the capabilities identified.

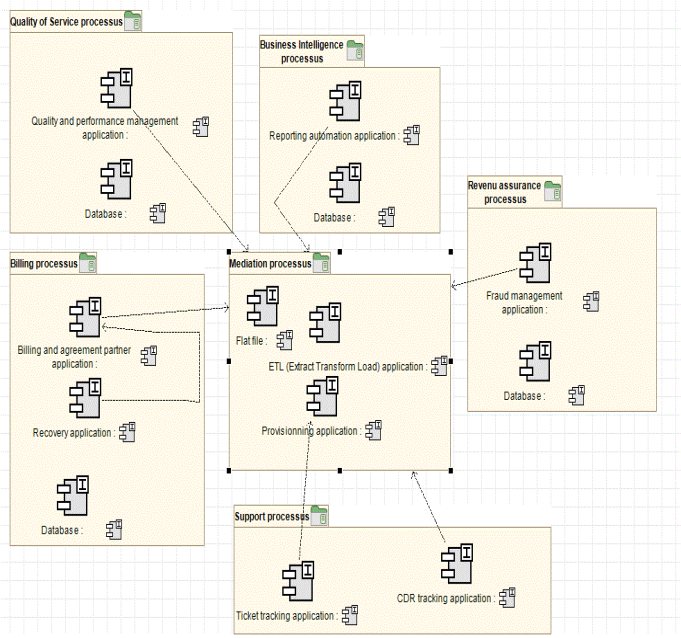


Figure 10. Architecture decomposition diagram before application of ISPM

3.4. Step 4: mapping and prioritizing threats

ISPM considers the same families of threats as Linddun [8]. It should be remembered that Linddun is imbued with the work of Pfitzmann and Hansen [36] for the identification of these families of threats for the categories "Linkability, identifiability and detectability", while "disclosure of information" is imbued with the STRIDE methodology [37]. Non-redudiation, Unawareness and Non-compliance come from the experience of Kim [38] and his team in the field of personal data protection.
Beginning first by reminding the families of the seven threats [8] which the authors will consider:

- Linkability: to be able to distinguish sufficiently whether two elements of interest are related or not, even without knowing the real identity of the subject of the element of interest that can be linked.
- Identifiability: being able to sufficiently identify the subject within a set of subjects.
- Non-redudiation: have irrefutable evidence concerning the occurrence or non-occurrence of an event or action.
- Detectability: when an attacker can sufficiently distinguish whether or not an element of interest exists in the data set.
- Disclosure of Information: disclosure of information to a person not authorized to consult it.
- Unawareness: not understanding the consequences of sharing personal information.
- Non-compliance: non-compliance with data protection legislation.

Table 4 impregnated by Linddun and adapted to the ISPM methodology includes an additional layer: the application layer added to the other capabilities identified by Linddun.

Table 4. Threats Correlated to Capabilities – ISPM

ISPM	L	I	N	D	D	U	N
Actor	X	X		X	X	X	X
Data flow	X	X	X	X	X		X
Data store	X	X	X	X	X		X
Process	X	X	X	X	X		X
Application	X	X	X	X	X		X

The mapping in Table 4 should be correlated with the capabilities identified in step 3 to create a custom mapping table that matches the system being analyzed. The resulting mapping table should contain all capabilities generated from the modeling done in step 3. This table should then be used as a checklist throughout the analysis, because each "X" in the table represents a potential threat impacting a specific item generated from the modeling performed in step 3. Therefore, each "X" should be documented as a threat, if applicable a hypothesis should be explicitly written to explain why the threat does not is not maintained and in this case the cell is grayed out.

Table 5 illustrates this mapping concerning the studied perimeter:

Table 5. threat mapping with capabilities

ISPM	Capabilities	L	I		N	D	D	U	N
Actor	Security employee, database administrator, billing employee, developer, mediation employee, CEO, Technical director, Executive committee, information security officer, regulatory officer, quality officer, management control director, business architect, application architect, data architect, technical architect, security manager, support manager, operations manager, business intelligence manager, mediation manager, QoS manager, billing manager, revenue assurance manager, revenue assurance employee, support employee, systems and networks engineer, business intelligence employee, QoS employee, billing manager.	X	X			X	X	X	X
Data flow	CDR flow	X	X		X	X	X		X
Data store	CDR files, Quality of Service DB, Revenu assurance DB, Billing DB, QoS DB, BI DB.	X	X		X	X	X		X
Process	Mediation, BI, Quality of Service, Revenu assurance, Billing, Support.	X	X		X	X	X		X
Application	BI BO, GPTO, Sage, Microsoft application files, Subex, IME, MyCom, Oracle,	X	X		X	X	X		X

For each cell marked in the mapping in Table 5, a threat tree exists. The privacy threat trees are inspired by the Secure Development Lifecycle (SDL) [37] and based on the most recent privacy watches. These threat trees reflect common attack patterns and help people think about privacy conditions in the information system. As threat trees are regularly updated, we refer the reader to the latest version of the trees, available on the Linddun website [39] and to the article [8] for "application" layer threat trees.

The result of this threat decortication phase is a set of threat scenarios that must be documented. For this purpose, the authors presented a documentation sample and mentioned a part of them in table 6:

Table 6. a part of threats census

Catégorie de la menace	Désignation	menace	Désignation
L_e	Linkability of Actor (or Entity)	L_e1	Linkable login using untrusted communication
		L_e2	Linkable login
		L_e3	Untrusted communication
		L_e4	Fixed login re-used
		L_e5	Certificates used that are too specific

		L_e6	Untrustworthy receiver
L_df	Linkability of data flow	L_df1	Linkability of transactional data (transmitted data)
		L_df2	Linkability of contextual data (metadata)
		L_df3	Data available to untrusted party
		L_df4	Non-anonymous communication are linked
		L_df5	Insecure anonymity system deployed
		L_df6	Data flow not fully protected
		L_df7	(Future) receiver untrusted
		L_df8	Based on IP address
		L_df9	Based on computer ID
		L_df10	Based on session ID
		L_df11	Based on behavioral patterns (time, frequency, location)
		L_df12	Traffic analysis possible
		L_df13	Active attacks possible
		L_df14	Passive attacks possible
L_ds	Linkability of data store	L_ds1	Weak access control to data
		L_ds2	Insufficient minimization/ inference
		L_ds3	Data linkable to other DB (external or internal)
		L_ds4	Excessive data available
		L_ds5	Storing data too long
		L_ds6	Storing personal data
L_p	Linkability of a process	L_p1	Different actions can be linked to the same user
	Identifiability at entity	I_e1	Identifiable login using untrusted communication
		I_e2	Identifiable log-in used
I_e		I_e3	Untrusted communication
		I_e4	Identity used as log-in
		I_e5	Pseudo-identity used as log-in
		I_e6	Certificates used as login
		I_e7	E-id used as log-in
		I_e8	Pseudonym used as login
		I_e9	Token used as log-in
		I_e10	Biometrics used as log-in
		I_e11	Too many attributes shared
		I_e12	Username password used as log-in
		I_e13	Link made between account and identity

Before moving forward and looking for appropriate mitigation measures for identified threats, they should be prioritized. Due to time or budget constraints, only the most significant threats will often be considered for inclusion in the specification of the requirements. Risk assessment techniques support this step; the risk is calculated based on the probability of the attack scenario and its impact. The risk value is used to sort the threat roots: the higher the risk, the greater the threat root: Risk = probability x impact. ISPM is independent of the risk assessment technique used. The analyst is free to choose the technique of his choice, for example the risk assessment methodology of OWASP [40], Microsoft's DREAD [41], the NIST special publication 800-30 [42] or OCTAVE by SEI [43]. These techniques exploit the information contained in the threat files, such as the capabilities involved (for the impact), the profile of the attacker as well (for the probability).

The most probable threat and which has the most impact in the article-studied perimeter is linkability in data store. Linkability in a data store occurs when one has access to the data store and when insufficient data minimization is applied. This means that too much data are being stored which enables a large set of information that can be used to look for links. The most obvious consequence of linking lots of information is that more pseudo-identifiers are linked which can result in identifiability, for example: knowing one's city, gender, age or even first name does not reveal an identity, but when combined the anonymity set suddenly becomes a lot smaller and can already lead to identification, depending on the city's population size and the uniqueness of the person's first name. Thus the more data available and linkable based on Quasi-identifiers, the more likely the chance of identification.

Another result of linkability is inference. Instead of linking data that belongs to the same person, data are linked based on certain properties to deduce relationships between them and

generalize them. This can be used in a rather innocent fashion to determine the best way to organize groceries in a grocery store (e.g. people who buy hamburgers usually buy buns at the same time, hence they are stored close to each other). This inference can however also have a more judgmental nature if it is used to discriminate a certain population (e.g. people living in a certain neighborhood have a higher chance of cancer, hence their health insurance fee is higher than the surrounding cities). Inference can thus lead to societal harm.

In what follows, the authors will focus on the strategy to be followed to deal with the "Linkability of Data Store" of the studied perimeter.

3.5. Step 5: strategy choice and changes modeling

In Table 7, the authors focused on the strategies to adopt to deal with the risks associated with the Linkability of Data store, which is choosed from ISPM mitigation strategies [8].

Table 7. Strategy to be adopted regarding "Linkability in Data Store" threats

Root threats	Designation	Threat nodes	Capabilities impacted	Chosen strategy	Personal data concerned
L_ds	Linkability of a data store	L_ds1 :Week access control to data	CDR files, Quality of Service DB, Revenu assurance DB, Billing DB, QoS DB, BI DB.	Authentication and authorization management	Calling party number
		L_ds2: Insufficient minimization/ inference		anonymization and pseudo-anonymization	Called party number
		L_ds3: Data linkable to other DB (external or internal)		anonymization and pseudo-anonymization	Calling party IMSI: International Mobile Subscriber Identity
		L_ds4: Excessive data available		anonymization and pseudo-anonymization	Called party IMSI: International Mobile Subscriber Identity
		L_ds5: Storing data too long		Data retention (policy compliance)	Calling location area code party
		L_ds6: Storing personal data		anonymization and pseudo-anonymization	Called location area party
					Calling Cellule ID
					Called Cellule ID
					IMEI: International Mobile Equipment Identity
					URL: Uniform Resource Locator
					IP address
					Site web visited

The authors started from the assumption that the access control to the data store is mastered and that the historisation of accesses is implemented and is periodically checked in the company in question. In addition, the retention of data follows the legislation of the country in question. The authors will deal in what follows with the threat of "anonymization and pseudo-anonymization" and "Storing personal data". Changes made on the old modeling of steps 2 are added with green color on the chosen modeling tool to highlight them.

Mediation is the bottleneck of personal data and it is the layer where the personal data in CDRs are routed to the information system processes. In Figure 11, the authors presented how the mediation and support process have been modified to adapt them to the principles of privacy by design more specifically the third principle "privacy embedded into design" by using anonymization. Tasks, events and new roles are represented in green:

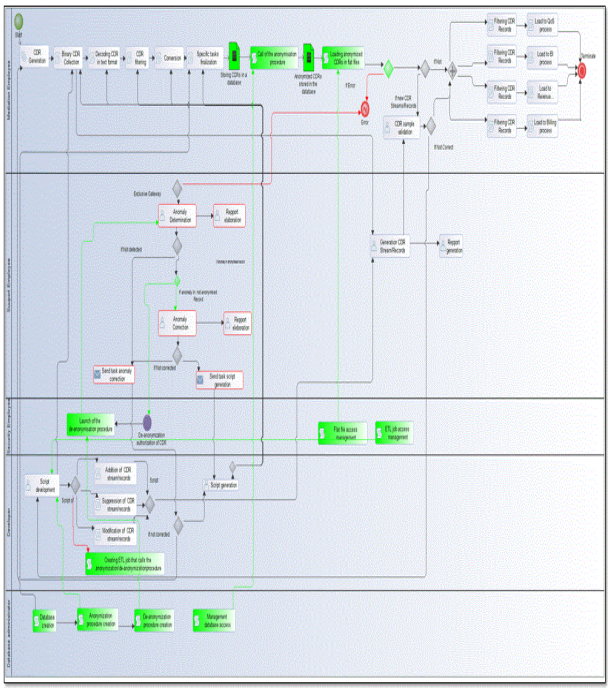


Figure 11. Togaf: BPMN process diagram of mediation after application of ISPM

To be able to implement the privacy by design while respecting the principle 4: positive sum, it is necessary to have a billing employee with higher rights for executing the procedure of de-anonymization. The "Called Party" record is important in the

calculation of the voice and data service use, but the authors made it in clear at the appropriate time of the calculation and not at the reception of the CDRs. Then, they anonymized it just after the calculation. The billing employee with higher rights will also be called to perform the de-anonymization procedure for the handling of anomalies when it is necessary.

In Figure 12, the authors presented the diagram of Billing BPMN process taking into account the privacy by design:

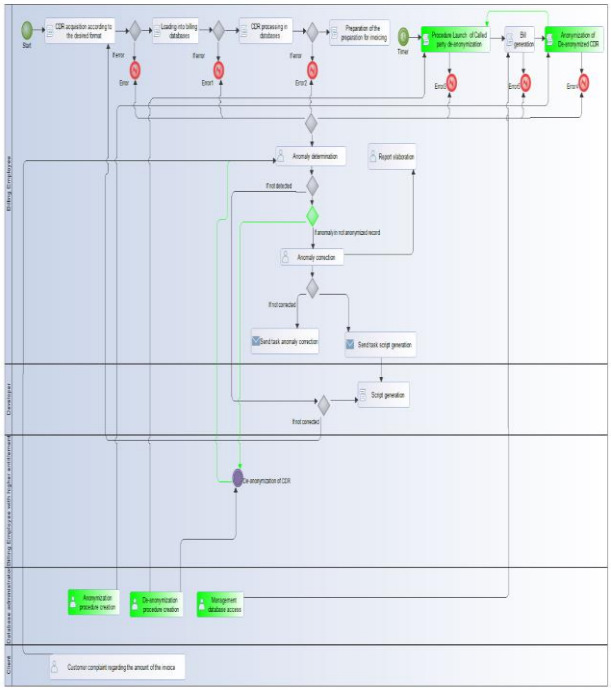


Figure 12. Togaf: Billing BPMN Process Diagram after application of ISPM

In what follows, the authors will treat the BI, RA and QoS diagrams adapted to the principles of privacy by design: the personal data are anonymized because the CDR are received anonymous from the mediation (see figure 6). In some cases, the employee may need to visualize the personal data in clear for performing tests or to proceeding to advanced user service analysis. the authors approached BPMN diagram for the same processes seen in figure 8 after the desired adaptation to PbD taking into account the need for clear access to personal data as shown in figure 13.

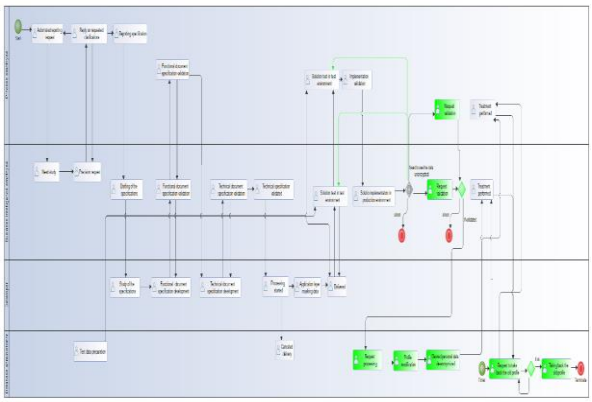


Figure 13. Togaf: BPMN process diagram of BI, RA and QoS after application of ISPM

After having treated the modeling of all the processes in interaction with the CDR before and after application of ISPM, the authors noted that several actors would be authorized to handle the personal data in the clear. Others had to take charge data de-anonymization and anonymization, hence the need to ensure that the integrity of personal data will not be compromised and that their access authorization will be properly controlled. To do this, Togaf proposes a diagram called "Data Security Diagram" in which the authors grouped the actors interacting with the personal data of all processes interacting with the CDRs. The figure 14 specifies the authorizations of the actors in terms of data security.

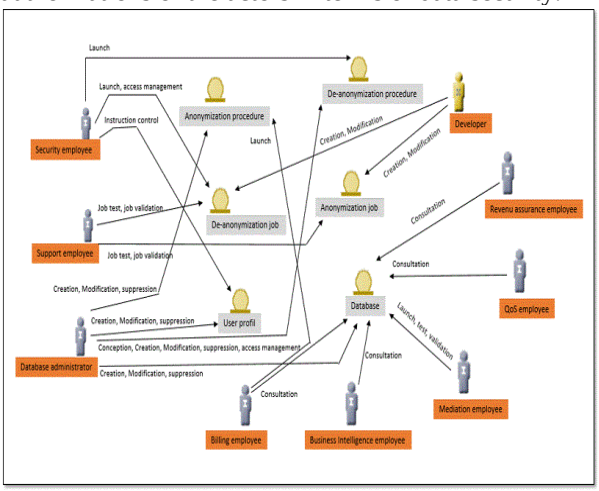


Figure 14. Togaf: Data Security Diagram

In Figure 15, the authors presented the changes made on the technical architecture after the desired adaptation to the principles of privacy by Design. The only change necessary to ensure our goal is to add a database in the technical architecture of mediation process. The universally ETL specification used to ensure mediation makes it possible to integrate anonymization jobs and all the other processes interacting with CDRs already have a database where anonymization and de-anonymization will take place in case of need.

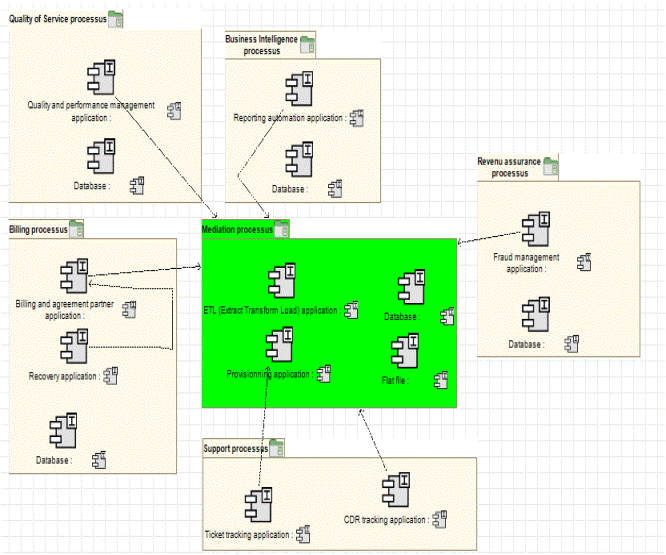


Figure 15. Togaf: platform decomposition diagram after adaptation

In the next section, the authors will implement pseudo-anonymization in the mediation process to measure the execution time of the two operations: pseudo-anonymization and pseudo-anonymization.

3.6. Step 6: implementation

In this section, the authors focused in the performance aspect of adding a layer of pseudo- anonymization and its impact on the mediation process capability. The performance parameter that is studied is the run time. The CDR sample the authors worked on contains 110 records separated by "|" [44]. The tool used for the tests is Microsoft Visual Studio, on a 2GB memory and four CPU 1.8 Ghz. The algorithms they applied for the anonymization job is a symmetric one: AES-256 whose private key is protected by a certificate. The personal data concerned in the tests are:

- Calling party number
- Called party number
- Calling party IMSI: International Mobile Subscriber Identity
- Called party IMSI: International Mobile Subscriber Identity
- Calling location area code party
- Called location area party
- Calling Cellule ID
- Called Cellule ID
- IMEI: International Mobile Equipment Identity

Figure 16 shows the job that the authors performed for anonymization:

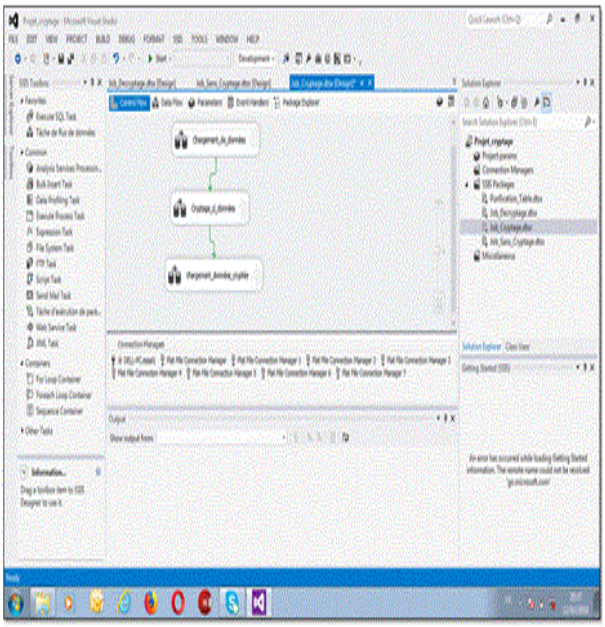
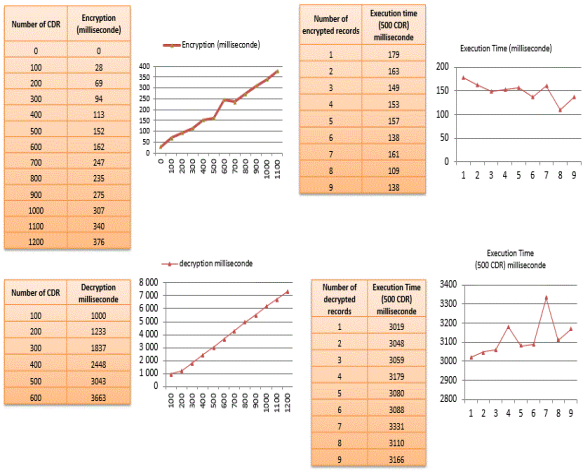


Figure 16. Pseudo-Anonymization job

The authors studied four varieties of tests: varying the number of CDR encrypted, varying the number of encrypted records simultaneously, varying the number of CDR decrypted and varying the number of decrypted records simultaneously.
The results of the tests are shown in Figure 17:



Number of CDR	Encryption (milliseconde)
0	0
100	28
200	69
300	94
400	113
500	152
600	162
700	247
800	235
900	275
1000	307
1100	340
1200	376

Number of encrypted records	Execution time (500 CDR) milliseconde
1	179
2	163
3	149
4	153
5	157
6	138
7	161
8	109
9	138

Number of CDR	Decryption milliseconde
100	1000
200	1233
300	1837
400	2448
500	3043
600	3663

Number of decrypted records	Execution Time (500 CDR) milliseconde
1	3019
2	3048
3	3059
4	3179
5	3080
6	3088
7	3331
8	3110
9	3166

Figure 17. Performance of the implementation

- Study of encryption by varying the number of encrypted CDRs and setting the number of records encrypted to one value. The execution time is in the order of milliseconds, so it can be used in production environment.
- Study of encryption by varying the number of encrypted records simultaneously and fixing the number of CDRs at 500. The execution time is in the order of milliseconds, so it can be used in production environment.
- Study of decryption by varying the number of decrypted CDRs and setting the number of records decrypted to one value. The execution time is in the order of milliseconds, so it can be used in production environment.
- Study of decryption by varying the number of decrypted records simultaneously and fixing the number of CDRs at 500. The execution time is in the order of milliseconds, so it can be used in production environment.

4. Discussion

According to the analysis of the necessary prerequisites of ISPM implementation in the processes in interaction with the CDR, it appears that in terms of modification on the technical architecture, only the addition of a database in the architecture of the process of mediation is sufficient. The tool of mediation allows universally creating jobs that can integrate anonymization and de-anonymization. The other processes already contain databases, thus the execution of the procedures of anonymization and de-anonymization is possible without having to add another tool or application. The reason for adding the database in the mediation is that the CDRs are in flat file format and the other processes use already the content of the CDRs in the databases. So if there is a need to anonymize the CDRs in the mediation for the downstream processes, it would be necessary to de-anonymize the data when necessary, which requires that the anonymization and de-anonymization key be shared in an intrinsic way. This can only be assured if the anonymization and de-anonymization is made in the same type of application.

In what follows the authors discussed the taking into account of the privacy by design principles in the proposed solution:

- Proactive not reactive: the authors anticipated and prevented privacy-invasive events before they happen. Having the strategic thinking of adopting the privacy by design principles and implementing them allows not to wait for privacy risks to materialize.
- Privacy as the default setting: the authors built privacy measures directly into the appropriate processes interacting with CDR, by default. They make sure that information systems are configured to be as secure as possible by default rather than having users do it one by one by configuring the system: least privilege, Need-to-know, least trust, mandatory access control, segregation of duties. They anonymized the personal data to be under cover of any configuration error.
- Privacy Embedded into Design: the authors embedded privacy into the design and architecture of processes interacting with CDR and business practices. They do not bolt it on after the fact. They embedded security into the design by anonymizing data by default.
- Positive-Sum: the authors accommodated a positive-sum "win-win" way, not a zero-sum approach, involving unnecessary trade-offs. They retained the utility of data by allowing the de-anonymization of data when is needed for business needs.
- End-to-End security: the authors ensured cradle-to-grave, secured lifecycle management of information, end-to-end. They generalized principles 1, 2 and 3 on all the processes in interaction with CDR and for all the insiders.
- Visibility and Transparency: the authors kept component parts of process interacting with CDR and operations of business practices visible and transparent to users. To do this, they used known standards in the scientific community such as the ADM method of Togaf. In addition, they have adopted known security principles, evaluated and validated.
- Respect for User Privacy: the authors respected and protected interests of the service provider clients above all. They kept it user centric by respecting the individual privacy rights.

The application of ISPM methodology done is also:

- Holistic, so it takes into account the process layer, data flow layer, the application layer and the technical architecture.
- Integrates with the IS design method used in the company and proposes generic and standardized models.

In addition, it carries out a privacy treats analysis, their impact and occurrence degrees. Ultimately, it proposes strategies for resolving threats and reducing them considerably.

5. Conclusions

In this article, the authors began by presenting the reasons that prompted them to choose implementing privacy by design on processes interacting with CDR for telecom service provider according to ISPM methodology. Then, the authors approached the related works on privacy by design Implementation. They followed their work by presenting the implementation of personal data protection according to the ISPM methodology. Subsequently, they evaluated and discussed the proposed solution. The authors ended by highlighting the level of alignment of their solution with the requirement of ISPM methodology.

References

[1] A. Cavoukian, M. Prosch. “The Roadmap for Privacy by Design in Mobile Communications: A Practical Tool for Developers, Service Providers, and Users,” December 2010. Toronto, Ontario, 2010.

[2] S. Arfaoui, A. Belmekki & A. Mezrioui. “Privacy increase on telecommunication processes”, Published in 2018 International Conference on Advanced Communication Technologies and Networking (CommNet). 10.1109/COMMNET.2018.8360266, 2018

[3] S. Miyazaki, N. Mead & J. Zhan. “Computer-Aided Privacy Requirements Elicitation Technique,” published in IEEE Asia-Pacific Services Computing Conference, 2008. DOI 10.1109/APSCC.2008.263.

[4] N. R. Mead, S. Miyazaki & J. Zhan. “Integrating Privacy Requirements Considerations into a Security Requirements Engineering Method and Tool,” published in Information Privacy, Security and Integrity journal, Vol. 1, No. 1, 2011.

[5] C. Kalloniatis, E. Kavakli & Stefanos Gritzalis. “Addressing privacy requirements in system design: The PriS method,” DOI: 10.1007/s00766-008-0067-3 · Source: DBLP, 2008.

[6] F. S. Gürses. “Multilateral Privacy Requirements Analysis in Online Social Network Services,” Dissertation presented in partial fulfillment of the requirements for the degree of Doctor in Engineering in the Arenberg School, 2010.

[7] K. Beckers, S. Faßbender, M. Heisel & R. A. Meis “Problem-based Approach for Computer Aided Privacy Threat Identification. In Privacy Technologies and Policy,” volume 8319 of LNCS, pages 1–16. Springer, 2014.

[8] S. Arfaoui, A. Mezrioui & A. Belmekki. “A Methodology for Assuring Privacy by Design in Information Systems,” International Journal of Communication Networks and Information Security (IJCNIS), Vol. 12, No. 3, December 2020.

[9] M. Prosch. “The Roadmap for Privacy by Design in Mobile Communications: A Practical Tool for Developers, Service Providers and Users,” http://www.ontla.on.ca/library/repository/mon/25003/308516.pdf, December 2010.

[10] A. Cavoukian. “Privacy and Security By Design: a convergence of Paradigms,” published jointly by an team composed of the Information and Privacy Commissioner of Ontario, Ann Cavoukian et Oracle Corporation,2013 http://www.discoveringidentity.com/2013/03/21/privacy-and-security-by-design-a-convergence-of-paradigms/.

[11] A. Cavoukian & M. Dixon. Privacy and Security by Design: An Enterprise Architecture Approach, http://www.discoveringidentity.com/2013/09/23/privacy-and-security-by-design-an-enterprise-architecture-approach/, 2013.

[12] A. Cavoukian. “Information and Privacy commissioner of Ontario, Guidelines for the Use of Video Surveillance Cameras in Public Places,” 2007. https://files.townlife.com/public/uploads/documents/7032/Appendix_F_-_Video_Surveillance_System.pdf

[13] A. Cavoukian. “Information and Privacy Commissioner of Ontario: White Paper: Anonymous Video Analytics (AVA) technology and privacy,” April 2011. https://www.ipc.on.ca/wp-content/uploads/Resources/AVAwhite6.pdf

[14] A. Cavoukian. “Privacy and Video Surveillance in Mass Transit Systems: A Special Investigation Report– Privacy Investigation Report MC07-68,” 2008.
[15] K. Marti & B. Inc. “Video Surveillance Cameras: An Innovative Privacy-Enhancing Approach,” 2008.
[16] A. Cavoukian & A. Stoianov. “Biometric Encryption: A Positive-Sum Technology that Achieves Strong Authentication, Security AND Privacy,” 2007.
[17] T. Marinelli. “Privacy-Protective Facial Recognition: Biometric Encryption Proof of Concept,” 2010.
[18] K. Anderson, M. Winters, S. Ferencie, N. Kelly & J. Meyers. “Operationalizing Privacy by Design: The Ontario Smart Grid Case Study,” 2011.
[19] A. Cavoukian & K. Kursawe. “Implementing Privacy by Design: The Smart Meter Case,” published in the IEEE International Conference on Smart Grid Engineering (SGE’12) Oshawa, Ontario, 2012.
[20] A. Cavoukian. “Shaping Privacy on the Smart Grid. You Can Make a Difference: A Roadmap for Data Protection Commissioners and Privacy Regulators,” 2010.
[21] A. Cavoukian. “Adding an On/Off Device to Activate the RFID in Enhanced Driver’s Licences: Pioneering a Made-in-Ontario Transformative Technology that Delivers Both Privacy and Security,” 2009.
[22] A. Cavoukian & F. Carter.” Privacy Guidelines for RFID Information Systems (RFID Privacy Guidelines),” 2006. Downloaded from https://www.ipc.on.ca/wp-content/uploads/resources/rfid-guides&tips.pdf.
[23] C. Hertzman, N. Meagher & K. McGrail. “Privacy by Design at Population Data BC: a case study describing the technical, administrative, and physical controls for privacy-sensitive secondary use of personal information for research in the public interest,” 2012.
[24] K. El Emam. “A Positive-Sum Paradigm in Action in the Health Sector,” 2010. Downloaded from https://www.ipc.on.ca/wp-content/uploads/2010/03/pbd-positive-sum-paradigm.pdf
[25] R. Alvarez. “Embedding Privacy into the Design of EHRs to Enable Multiple Functionalities Win/Win,” 2012. Downloaded from https://www.ipc.on.ca/wp-content/uploads/Resources/pbd-ehr-e_1.pdf
[26] J. Curtis & N. Jolly. “Encryption by Default and Circles of Trust,” 2012. Downloaded from http://www.ontla.on.ca/library/repository/mon/26012/320323.pdf
[27] J. Jonas. “Privacy by Design in the Age of Big Data,” 2012. Downloaded from https://jeffjonas.typepad.com/Privacy-by-Design-in-the-Era-of-Big-Data.pdf
[28] M. Ho. “Redesigning IP Geolocation: Privacy by Design and Online Targeted Advertising,” 2010. Downloaded from https://www.ipc.on.ca/wp-content/uploads/resources/pbd-ip-geo.pdf
[29] K. Cameron. “Wi-Fi Positioning Systems: Beware of Unintended Consequences,” 2011. Downloaded from https://www.identityblog.com/wp-content/images/2011/06/beware_of_unintended_consequences.pdf
[30] K. Rannenberg. “Privacy by Design in Mobile Applications and Location Based Services,” 2010. Paper presented at “Privacy by Design: The Gold Standard,” Toronto, Ontario.
[31] A. Cavoukian. “Mobile Near Field Communications (NFC) “Tap ‘n Go” – Keep it Secure & Private,” 2013. Downloaded from https://webcache.googleusercontent.com/search?q=cache:pDSYGaRRAMsJ:https://www.rfidjournal.com/whitepaper/mobile-near-field-communication-nfc-tap-n-go-keep-it-secure-and-private+&cd=10&hl=fr&ct=clnk&gl=ma
[32] « 2020 Telecommunications Industry Outlook - CIO Journal-WSJ ». https://deloitte.wsj.com/cio/2020/01/22/2020-telecommunications-industry-outlook/ (seen on February. 17, 2021).
[33] J. Pepin. “Enterprise Architecture: mapping alignment business and applications of the information system,” 2018. DOI: 10.1109/rICT-ICeVT.2013.6741505
[34] The Open Group. “Togaf version 9.1 Entreprise Edition: Sample Catalogs, Matrices and Diagrams v3,” 2011. Download from http://www.opengroup.org/bookstore/catalog/i093.htm,
[35] P. Desfray & G. Raymond. “Togaf in practice, Entreprise Architecture Models,” Dunod, Paris, 2014. ISBN 978-2-10-071451-3.
[36] A. Pfitzmann & M. A. Hansen. “Terminology for talking about privacy by data minimization: anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management (Version 0.33 April 2010), technical report,” TU Dresden and ULD Kiel, http://dud.inf.tu-dresden.de/Anon_Terminology.shtml
[37] M. Howard & S. Lipner. “The Security Development Lifecycle: A Process for Developing Demonstrably More Secure Software. Published by Microsoft Press a Division of Microsoft Corporation,” 2006.
[38] M. Deng, K. Wuyts, R. Scandariato, B. Preneel. & W. Joosen. “A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements,” Requirements Engineering Journal , volume 16, issue 1, pages 3-32, Springer-Verlag (This is the original LINDDUN paper), 2011.
[39] K. Wuyts. “Treat tree catalogue,” 2020. https://distrinet.cs.kuleuven.be/software/linddun/catalog.php
[40] M. E. Ramadlan. “OWASP, "Risk rating methodology,” 2013 https://owasp.org/www-pdf-archive/Riskratingmanagement-170615172835.pdf.
[41] S. Chilton, M. J. Lee, F. Kiraly, H. Metcalf, & W. Pang. “Dread Risks, Journal of risk and uncertainty,” volume 33, pages 165-182, 2006.
[42] G. Stoneburner, A. Goguen & A. Feringa. "Risk management guide for information technology systems, special publication 800-30," Available: http://csrc.nist.gov/publications/nistpubs/800-30/sp800-30.pdf, 2002.
[43] R. Krikken. “The OCTAVE Risk Assessment Methodologies, Gartner Research ID: G00203984,” 2010.
[44] Technical Specification 32.298. “Telecommunication management, charging management, Charging Data Record (CDR) parameter description, 3GPP, version 15.3.0,” 2018.