Technical Enforcement of European Privacy Legislation: An Access Control Approach

Kheira. Bekara¹, Maryline. Laurent ², & Than.Ha. Nguyen ³

¹ Institut Telecom, Telecom SudParis, CNRS Samovar UMR 5157
9 rue Charles Fourier, 91011 Evry, France
{¹Kheira.Bekara, ²Maryline.Laurent, ³Maryline. Thanh_Ha.Nguyen}@it-sudparis.eu

Abstract— Until today, the protection of personal data is mainly left to the legislation by means of guidelines. This paper aims to increase the perceived control by users over their data by helping the user’s agent to check the service requests conformity to the legislation. To do so, it discusses the main concepts involved in the legislative privacy principles, and deduces a privacy semantic information model. The proposed model focuses on the main concepts involved in legislative privacy principles. For proof of concept, we describe our proposed privacy semantic information model by means of privacy ontology. We use OWL as an implementation basis for defining privacy knowledge base, and SQWRL and the Jess rule engine to dynamically interact with that base and enforce legislative requirements as privacy access control rules.

Keywords— Privacy, legislative requirements, access control, ontology.

I. INTRODUCTION

Privacy protection, as a social issue [1], is still left to the legal framework and to Service Providers (SP) self regulation. Privacy policies [2] are defined by SPs, and so far they are displayed to users under a literal form with abstract terms that are difficult to understand by most of the users. As such, to simplify personal data privacy protection enforcement, there is a strong need to bring down legislative requirements into the technological reality and to design new technical solutions. Note that these solutions must be adapted to the transaction context; it must take into consideration all the following elements:

- The type of the service requiring users’ data,
- The type of the required data element,
- The applicable rule to that context.

The Platform for Privacy Preferences (P3P) W3C specification [2] has been the first initiative towards this direction, providing a way for a web site to encode its relevant practices and to communicate them to the visiting users. P3P formalizes SPs’ privacy commitment but is limited to the following aspects: Purpose, Recipient, and Retention. Also P3P does not permit to specify the type of the service requiring users’ personal data, nor the type of the requested personal data item.

The challenge for enforcing privacy requirements has been widely examined. Research and development efforts resulted in several frameworks proposed by HP [3], IBM [4], and OASIS [5]. However, the privacy policies specified in the context of these frameworks cannot be efficiently audited to verify their consistency and legislative compliance. By definition, the expression of privacy policies in these frameworks is not based on the whole privacy legislative requirements. In fact, automating the enforcement of privacy policies is done by applying privacy-aware access control mechanisms, i.e. the traditional Role-Based Access Control (RBAC) models are enriched with additional privacy related aspects [6].

Gandon and Sadeh [7], Rao et al [8], and Jutla et al [9] investigate using the semantic web technologies to support privacy in e-commerce. They choose ontology languages, e.g. OWL and ROWL, to represent users’ privacy preferences and contextual information. Garcia in [10] proposes a privacy ontology to support translation of privacy policies expressed using a P3P vocabulary into assertions that are used to control access to personal data. Although these solutions manage and address the access control issue to personal data by means of privacy policies enforcement, they focus on the user’s privacy preferences specifications and not legislation based privacy policies. Also, the specifications of the privacy policies do not take into account neither the type of the service, nor the type of the requested personal data item.

In the light of above limitations of current approaches, this article presents a new semantic information model formalizing main concepts of legislative frameworks. The main idea behind this semantic information model is to enforce our privacy based access control framework satisfying legislation requirements. That is in a specific transaction context the user agent can check the conformity of the SP’s request to the legislation.

This paper is organized as follows. Section II introduces legislative principles for personal data protection. Section III briefly presents the privacy legislative requirements. Section IV describes our proposed privacy semantic information model for modeling the privacy context associated to each user’s transaction. Section V illustrates a concrete example about the deployment of the privacy semantic information model. Section VI gives conclusions.

II. LEGISLATIVE PERSONAL DATA PROTECTION

Since its acknowledgement as a fundamental human right by the Universal Declaration of Human Right of the United Nation in 1974 [14], the personal data are protected by the relevant legislation in many countries around the world.
The first influential legal framework was the US Privacy Act [13] adopted by the Congress in 1974. Nowadays, the European directive 95/46/EC related to the protection of physical persons and the processing of personal data [14], is the main legislative piece in the European Union in terms of privacy protection. This Directive found its source in the OECD (Organisation for Economic Co-operation and Development) privacy protection guidelines [15] and the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data [16].

Later the European Directive 95/46/EC was completed with the directives 02/58/EC [17], 2006/24/EC [18], which are examples of some sectoral applications of the privacy principles of the 1995 directive (related to the privacy preserving processing of personal data in the electronic communication sector).

Hereafter, there is a summary of the fundamental privacy principles with respect to the lawfulness and fairness of personal data collection and processing. These principles constitute the basis of the functional requirements defined in section III.

1. Fairness and lawfulness of the processed data
The principle of fairness imposes that the data processing cannot be performed with a malicious intent or with the objective to cause harm to the data subject. The data processing should comply with the applicable privacy law, and also with all applicable laws and legislations.

2. Explicitness and specification of purposes
Since the beginning of the data collection, the data processing should be marked with the intended purpose. The data collected and processed for specified, explicit and legitimate purposes may not further be processed for purposes that are incompatible with these for which they have been collected. The purpose principle is aimed at guaranteeing the data subject an effective control over the processing of his data.

3. Necessity of data collection and processing
Article 7 of the Directive 95/46/EC provides the criteria for the legitimate data processing (e.g. collection, recording, storage...). It states that all the personal data must be fairly and lawfully processed. This requirement is amplified by a number of rules prescribing criteria as pre-conditions to legitimate processing. Therefore, processing will be legitimate only if one or more of the following conditions is satisfied:

- The data subject gives consent for the processing,
- The processing is necessary to perform the contract requested by the data subject or to comply with a request of the subject coming later,
- The processing complies with a legal obligation,
- The processing enables to protect the vital interests of the data subject,
- The processing is necessary for the administration of justice,
- The processing is necessary for the legitimate interests of the SP.
- Information, notification and access right of the users

The information requirement is commonly held as the basic requirement for any data processing activity. Article 10 of the directive 95/46/EC specifies a list of information that should be given to the data subject prior to starting the processing activity. Article 11 tackles the case where information is not obtained directly from the data subject but from third parties. The information statement that the system has to give to the data subject must contain at least the following information:

- The identity of the SP,
- The purposes of the processing for which data are requested,
- The recipients or categories of recipients to which data are likely to be delivered;
- Information whether provision of personal data is mandatory or optional, with the consequences when claimed attributes are not delivered.

4. Security and accuracy
Article 17 of the Directive 95/46/EC states that the SP must implement appropriate organizational, physical and technical measures to protect personal data against any unlawful form of processing. Such measures shall ensure the appropriate security level matching the risks represented by the processing and the nature of the data to be protected.

Article 16 of the Directive 95/46/EC deals with third parties providing attributes. Any person acting under the authority of the SP, including any entities which have access to personal data, must process them according to the legal instructions only.

5. Supervision and sanction
An independent Privacy Authority has to be designated and should be responsible for supervising privacy provisions. In that respect, the SP should provide to the Privacy Authority the means for controlling every action of personal data collection and processing.

The Directive 95/46/EC states that each Member State shall apply its own privacy legislation, resulting from the instantiation of the above listed privacy principles of the Directive 95/46/EC. Data protection legislation worldwide, where available, naturally defines some exceptions, exemptions and restrictions concerning the scope of the aforementioned principles.
III. LEGAL REQUIREMENTS

The provisions specified by the legal frameworks of section II (excluding provisions related to security) lead to defining the following technical privacy requirements:

- Access control: Access to data must be controlled. The decision to give access to data should take into account the context of each privacy session.
- Role of semantics: The semantics is very crucial for characterizing the “privacy context”. Each data element should be handled according to its type, and according to the purpose for which the data are collected and processed.
- Complementary actions: Access to data should be accompanied by certain actions like informing the user, requesting for their explicit consent, notifying the Authorities, automatically enforcing the data retention periods.
- Role of users: The users are granted certain rights, including the right to be informed about the collection or processing of their personal data, and the right for their explicit consent to access their data. Additionally, they should be able to specify their own privacy preferences and to be sure their preferences and the privacy legislation are respected during a transaction.

Therefore, the legislation requires an access control model which: i) relies on the semantics of personal data, and services; ii) suggests the execution of data processing and/or other complementary tasks prior to granting access to data; iii) gives an active role to data subjects and grants them access rights.

IV. SEMANTIC INFORMATION MODEL

During a transaction, when the service provider is asking the user for some personal data, it is of high interest for the user to automatically check whether the service provider request is compliant to the legislation. To do so, this section describes a semantic information model formalizing the legal requirements presented in section III. As such, sub section A designs our privacy ontology. Once the privacy ontology is designed, it is possible to set up the legislation-based access control policies, which are specified thanks to our XPACML [11] vocabulary defined in sub section B.

A. Designing a Privacy Ontology

The selection of the legislative privacy policy to apply in each situation is done based on the three following elements characterizing the privacy context related to each privacy session, i.e. the type of the data element under consideration defined by “DataType”, the service type requiring the data element defined by “ServiceType”, and the legislative rule applying to the [DataType, ServiceType] pair. As-such, we integrate these main elements supporting legislative access decisions to user personal data in an access control framework.

The main idea behind our approach is to model these main concepts using a semantic information model that associates personal data types and service types with explicit legislative rules.

Therefore, to associate the personal data with specific processing tasks, the identification of the particular type of each data item is necessary. Moreover, in order to define the appropriate rules (purpose, recipient, retention...) that will regulate the processing of a personal data item, a similar taxonomy of the services must be present. These taxonomies constitute separate sub-graphs of the privacy ontology. That is, the privacy ontology provides a detailed vocabulary of personal data types and service types, structured in a hierarchical way with well defined inheritance rules that enable the system to associate all privacy related decisions to semantically specified notions.

To that respect, we use the W3C Web Ontology Language (OWL) [19] to implement the sub-graphs related to the data types, the services types and legislative rules. The resulted privacy ontology is shared between the user and the SP as a common information model that contains the vocabulary related to the data types and service types, as much detailed as possible.

Regarding the personal data sub-graph, all the data types are defined as categories using appropriate sub-classes of DataType OWL class. The final data type objects, which are leaves of the personal data sub-graph classes, are defined as OWL instances of DataType sub-classes. Relationships between personal data are defined using the following OWL properties:

1. inheritsFromData
   - This property expresses the inheritance relationship between general data types and specific ones. It is implemented by means of inheritsFromData object OWL property.
2. hasMoreDetailed/hasLessDetailed
   - This property supports different levels of revelation. As such, in case there is a privacy policy conflict between the SP and the user about a data item, it is possible to substitute the data item for another one with a higher level of abstraction.
3. containsType/isContainedToType
   - This property expresses the complexity of a data item (e.g.: Full_Name contains the First_Name, Last_Name and Middle_Name). This relationship is implemented by means of containsType/isContainedToType OWL property, which in essence, defines a tree hierarchy. Figure 1 illustrates part of the personal data sub-graph classes along with some instances.

A similar pattern is adopted for the services sub-graph. The various types of services are defined as subclasses of the ServiceType class, and are organized as a hierarchy that defines the inheritance of characteristics. The final service type objects, which are leaves of the services sub-graph classes are defined as OWL instances of ServiceType sub-classes. In accordance with the DataType sub-graph, properties (1) and (3) are implemented for services classes. Figure 2 illustrates part of the services sub-graph.
The access control rules needed for regulating the collection and disclosure of personal data are part of a third sub-graph of the privacy ontology having the class Rule as root. Subclasses and instances of Rule class express legislative requirements like data retention period, user notification, and user consent requirements. Every rule is associated to \(\{\text{DataType}, \text{ServiceType}\}\) pair, using Access class. Therefore, the reasoner can infer the required legislation based policies that should apply when a specific service requests a specific data type. Figure 4 demonstrates the rule classes and instances. It is important to note that two instances are assigned to each element of the Policy sub-class of Rule class. One instance defines the minimalist privacy policy and the other defines the largest privacy policy allowed for a specific ServiceType requiring a specific DataType. This is done thanks to our ordered classification of each policy element instances according to their increasing risk level regarding user privacy. Figure 3, presents the DataPurpose instances ordered classification.

Therefore, we use the prefix Policy1 and Policy2 (example: purpose1 and purpose2) to define the minimum and maximum privacy element instances allowed for a given SP.

DataRecipient and DataRetention instances are organized following the same ordered classification approach.

The binding between the three sub-graphs is performed using the instances of the Access class. Instantiated AccessObjects link the services instances, and personal data type instances with appropriate rule instances, using “refersToService”, “refersToData”, “refersToRule” OWL object properties. Therefore they act primary as access right policies. That is, each AccessObject instance points to a set of instantiated PolicyObjects, which should apply when a specific service type requests a specific data type element. This concept is illustrated in figure 5.
B. Legislation-based Privacy Rules with XPACML

With the privacy semantic information model, it is possible to formally express the legislation-based access control rules that govern the issues of data collection and processing. After instantiation of the different privacy semantic information model classes and properties, relationships are defined between classes. The obtained legislation-based privacy rules can be stored or exchanged over the Internet thanks to our XACML-based language, namely the eXtensible Privacy Access Control Markup Language (XPACML) [11]. With the use of OWL annotation properties, every rule contains the following information:

- **DATA_TYPE**: Expresses the type of the data under consideration; its values come from the personal data ontology. It can also take the value ALL, covering all the types of data.
- **SERVICE_TYPE**: Expresses the service type for which some data are requested for disclosure or processing; its values come from the service Ontology. It can take the value ALL, covering all the services.
- **Effect**: Determines whether the data of type DATA_TYPE for the service of type SERVICE_TYPE should be disclosed to the provider or not.
- **PURPOSE**: This P3P element specifies the purposes for which the data of type DATA_TYPE are requested by the SP.
- **RECIPIENTS**: This P3P element lists the entities intended to collect the data of type DATA_TYPE.
- **RETENTION_PERIOD**: This P3P element specifies the retention period for the data of type DATA_TYPE.
- **ACTION**: Specifies the list of actions that can be performed by the SP (read, collect, share) that are permitted by the legislation.
- **Abstraction_LEVEL**: Determines the level of precision for the data of type DATA_TYPE for the service of type SERVICE_TYPE.

While the above sub-elements define the core of the rule, additional properties specify some complementary actions that might be executed:

- **MODIFICATION_PERMISSION**: whether the service provider has modification privileges over data of type DATA_TYPE during the provision of a SERVICE_TYPE service.
- **NOTIFICATION**: whether the user should be notified for some action on his DATA_TYPE data for the SERVICE_TYPE service.
- **CONSENT**: whether the user should be asked for his consent for some actions on his DATA_TYPE data for the SERVICE_TYPE service.
- **DATA_TYPE_DESCENDANTS**: whether the defined rule is applicable by inheritance to the descendants of the specified DATA_TYPE in the class hierarchy of the personal data ontology.
- **SERVICE_TYPE_DESCENDANTS**: whether the defined rule is applicable by inheritance to the descendants of the specified SERVICE_TYPE in the class hierarchy of the service ontology.

Figure 6 gives a small example of our XPACML policy statements restricted to the first seven elements of the rule core.

```xml
<xpacml:PolicySet xmlns:xpacml="urn:xpacml:policy">
  <xpacml:Policy PolicyId="eCommerce">
    <xpacml:Subject>
      <xpacml:Actions>
        <xpacml:read />
        <xpacml:collect />
        <xpacml:share />
      </xpacml:Actions>
    </xpacml:Subject>
    <xpacml:Resource ResourceId="urn:xpacml:resource">
      <xpacml:Target>
        <xpacml:Service_Type ApplyToDescendent="No">
          <xpacml:Description>
            One rule describes a policy for a specific category</xpacml:Description>
          </xpacml:Service_Type>
        </xpacml:Target>
      </xpacml:Resource>
    </xpacml:PolicySet>
  </xpacml:Policy>
</xpacml:PolicySet>
```

Fig. 5 Access class

Fig. 6 An XPACML-based legislation privacy rule

First the privacy ontology must be deployed on the user side, and then legislation-originated rules are translated internally into a set of concrete XPACML policies prior to be compared with SP’s privacy policy requesting a user personal
data. The objective is to detect, for a given SERVICE_TYPE, whether the SP making a DATA_TYPE personal data request, if privacy practices of its privacy policy intended as a future processing for the requested personal data type (e.g: Purposes, Recipients, Retention, …) is legal or not.

V. ILLUSTRATION OF THE ONTOLOGY

This section presents a concrete example about the deployment of the privacy semantic information model presented in section IV.

A. Specification of Legislation-based Policies

As mentioned in sub section A of section IV, we got the reference links between each Access object with DataType, Service Type and Rule objects. These links are represented under Protégé tool [20]. Figure 7 presents part of the concerned legislation Rule object (the purpose policy element).

Fig.7 Policy elements for one pair {DataType, ServiceType} under Protégé

Policy elements are stored as values of P3P tags [21] – purpose, recipient and retention - in “referToRule” section in figure 7.

To get a simple proof of concept, we set manually the values of instances for DataType and ServiceType objects using Protégé.

B. Extraction of Privacy Information from the Ontology

To extract information from the semantic information model, we use SQWRL (Semantic Query-Enhanced Web Rule Language) [20] and Jess rule engine [22]. Based on SQWRL queries, it is possible to get any information contained in the privacy ontology, like:

- For a given service, the personal data types which access is permitted
- The legislative policies that apply for a given {DataType, ServiceType} pair.

Jess rule engine, is used to run SQWRL from our Java application.

Considering the following query determining the policy that associates CCNumber (Credit Card Number) DataType with the eCommerce ServiceType: what is the purpose, recipient, retention for {CCNumber, eCommerce} pair?

It was mentioned in sub section A of section V that the Access objects are linked semantically to the DataType, ServiceType and Rules objects. Therefore, we developed our application by extracting directly the information from the Access objects. Figure 8 presents the policy elements related to the {CCNumber, eCommerce} pair extracted from the ontology.

Fig.8 Extraction of information from the privacy ontology

Once the type of service is set, all DataType objects allowed by the legislative framework for this ServiceType are displayed. After, the selection of one of these DataType objects, the related privacy policy is displayed thanks to the ordered classification of P3P elements instances defined in sub section A of section IV.

The resulted privacy policy is then exported in an XPACML format (figure 6) to be compared with the SP’s privacy policy.

VI. CONCLUSION

In this paper, a semantic information model for supporting personal data protection is presented under the form of privacy ontology. This work is based on the European legislation framework, and as such, it provides the means for the enforcement of the legislative provisions.

Our approach provides an innovative feature by introducing a formal modelling of the personal data protection legislation requirements in terms of concepts of privacy ontology. This ontology can be considered as a powerful tool to enforce privacy based access control satisfying legislation requirements. That is in a specific transaction context the user agent can check the conformity of the SP’s request to the legislation.

The privacy ontology might be specified and maintained by Data Protection Authorities, so that either the SPs or users can load the privacy semantic information

REFERENCES


