Privacy Support and Evaluation on an Ontological Basis

Michael Hecker, Tharam Dillon
eXel Lab[1], University of Technology, Sydney
{mhecker,tharam}@it.uts.edu.au

Abstract

This work is concerned with user perceived privacy and how clients (which we call data subjects here) can be empowered to control their own data consistently with their own interests. To support building and evaluation of privacy-aware applications, we describe a privacy ontology, how the privacy principles relate to that and how they are influenced by the core concepts as well as by each other. We use this influence of the privacy principles to evaluate the level of privacy for a particular transaction, when applying and extending the core concepts for an application domain.

1. Introduction

Privacy in computing and communication has many aspects and issues like legislation, technologies or user perceived privacy, which is the aspect of privacy a user experiences for data related to him or her. The "real" world as opposed to the digital world has to deal with privacy issues as well, but as usually more effort is required to gather and less effort to secure this information. It is not seen as such a big issue compared with the ability to collect, store and process information in the digital world. However, legal implications directly apply to both, the real world and the digital world, hence requiring parties to collect and use data carefully on a need to know basis only.

Although privacy issues are ubiquitous amongst almost every domain, actual instances and concepts as well as their influence on the overall level of privacy for "interactions" are naturally domain specific and must be addressed accordingly.

Firstly however, we elaborate our motivation and issues for the area of privacy in section 2 and describe an overall picture of privacy from a semantic perspective in section 3. This is followed by section 4, showing core concepts of the privacy ontology and their influential aspects towards the privacy principles. Due to lack of space in this paper, we omit our case study, which we will present at a later time in much greater detail and technical depth. Finally, we conclude this paper with section 5, describing issues we are still working on and further work.

2. Motivation and issues

2.1 Privacy

Privacy is considered one of the most important issues nowadays with easy collection, aggregation, linkage and storage facilities available. The Internet provides users with the ability to collect, store and share this kind of information easily, but lacks a cohesive structure making it more difficult to link data together by an automated process. However, other vast data sources (e.g. corporate databases) exist that are much more structured and allow their users to generate much clearer pictures about individuals. Therefore, it becomes more and more difficult to control others access to information about oneself.

The concept of privacy seems to be an endogenous conception, as every person has a different idea about what it means and how it should be implemented to achieve it. Therefore, it is necessary to find some common properties to build a basic foundation. Starting with a common dictionary definition, privacy would be "freedom from unauthorized intrusion" [1]. Similar technically imprecise definitions can be found in other dictionaries such as Oxford English Dictionary or dictionary.com. Therefore, it is necessary to consider the definition of privacy by people experienced in that domain. Naturally, it has evolved over time and started with the expression of "the right to be let alone" [2], expressed by two lawyers in 1890. However, such a definition is also not very precise yet nor very usable nowadays, as one does not necessarily want to be left alone just to "experience" or "have" privacy. A better definition comes from Privacilla [3], a website related to privacy related policies and defines

[1] http://exel.it.uts.edu.au

1-4244-0832-6/07/$20.00 ©2007 IEEE

it as "the subjective condition a person experiences when two factors are in place. First, he or she must have the power to control information about him- or herself. Second, he or she must exercise that control consistent with his or her interests and values". This definition describes privacy in a much better and precise way and it sounds more logical to be in control of information related to oneself than just be left alone. A similar statement has been made by the Privacy Commissioner of Canada, defining privacy as "the right to control access to one's person and to information about oneself" [4]. Trying to make it even clearer, privacy is not about information itself, but the control of that information by a cognitive entity, which is related to it. In order to distinguish such information from other "normal" information, we call this type of information, which is about someone and could potentially identify someone, "Personal Identifiable Information" (PII). Hence, if information cannot be linked to or is about a certain entity, which could have potential interests in controlling it, privacy matters usually do not apply.

Previously, privacy protection has been tried to be accomplished by utilising mechanisms that control access to personal identifiable information. However, it is not the data subject, which is the entity the data is related to, to control access to personal identifiable information, but an "authorised entity" controlling or maintaining the system where the data is store. Needless to say that such an "authorised entity" would have a great deal of control over the information, its release and access. Even more, the data subject might not know or even have authorised that entity to regulate access to its data, but just accepted the fact explicitly or implicitly that there is some sort of protection for its data. This can also be seen as an implicit trust in such an entity to do the "right thing" with the information made available to that system by the entity.

2.2 Privacy and policies

As just stated, a person controlling information about others could be of great danger to the privacy of the data subjects it is controlling – remembering that privacy is about the ability of the data subject to control personally identifiable information about themselves. Thus, privacy policies have been established, accepted and are now widely used and are backed up by legislation. This gives users more confidence when providing information about themselves if it is used in a certain way or to inform them at least how it is used. Privacy policies are usually set up and governed by certain rules and regulations that apply in the territory the entity collecting information is located in, leading to different privacy policies in different regions (e.g. privacy legislation in Europe compared to Australia). The problem with privacy policies is their different semantics and their dependency on the domain they are applied to. While privacy policies within the same domain (and possibly region for regulatory reasons) may have similar structures, there is no semantic way of comparing them with each other or even evaluating the level of privacy they try to offer. They may just be (and are often) written in a certain natural language (e.g. English). This obviously creates problems with precision, clarity and interoperability, making it ambiguous for the reader who has to understand it – being a person or software (agent). Different persons would understand a privacy policy differently, depending on the complexity and clarity of the policy and naturally depending on their "knowledge" about privacy and other intrinsic factors (e.g. culture), making it a fuzzy concept, making it hard or even impossible to formulate privacy in a mathematically precise way. Thus, software agents would have even more trouble "understanding" these kinds of privacy policies due to the lack of precision.

2.3 Privacy on the web and P3P

On the web, privacy policies have been established in a structured way with the introduction of the Platform for Privacy Preferences (P3P) [5]. Basically, this formulates certain statements about how resources (that is personal or other information) are used for what purpose, by whom and with what kind of retention. As P3P is a platform designed for websites, it covers mostly web-specific terms and is specific to that domain only (omitting the fact that extensions are possible, but their actual values are not standardised). Considering the initial issue that not all data is on or accessible via the web (in fact, a majority is actually not), other privacy policies that also cover electronic as well as non-electronic records from the same or other domains are necessary and available to be evaluated in a systematic way (by a human being or preferably automatically by a software agent).

2.4 Privacy Ontology

The basic idea behind these thoughts leads to a specified conceptualisation of the terminology "privacy", omitting internal and personal factors as it is difficult to capture them precisely. The terminology used for such a conceptualisation and formalisation is commonly known as "ontology".

Before it is actually possible to think about the specific privacy concepts and issues of a certain domain, e.g. healthcare, it is necessary to begin with a formulation of generic concepts first. These are likely to be domain independent and are abstract enough to support this. Generally, legislative documents provide a solid foundation for those concepts and are usually covered by the individual Privacy Acts of different nations. Privacy notions and concepts are specified by the "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data" [7] and used by us as one of the actual sources, as privacy legislation in the European Union is more advanced (more protective) than in many other countries. However, a more comprehensive and concise guide of those rules has been compiled by the PRIME [8] and PISA [9] projects describing the essential principles involved in the process of privacy. Issues and principles that have been found are: a) Intention and notification, b) Transparency, c) Finality principle, d) Legitimate grounds of processing, e) Quality, f) Data subject's rights, g) Processing by a processor, h) Security, i) Accountability, j) Openness, k) Anonymity and l) Transfer of personal data outside the EU (in general to countries with different privacy protection laws).

3. Ontology overview

This section elaborates how the ontology is built, especially which core concepts we introduce and how we set up the dimensions of the privacy principles. The privacy principles, as stated earlier, are a required component of the ontology to evaluate the privacy preferences at a later stage. Thus, the core concepts of the generic privacy ontology influence the privacy principles in a relative way without specifying absolute values. On the other side, specific extensions of the ontology as well as other extensions from other domains (e.g. legislation) influence the privacy principles directly by asserting absolute – yet fuzzy – values. Those values are evaluated later on to determine the level of privacy one gains from a certain transaction.

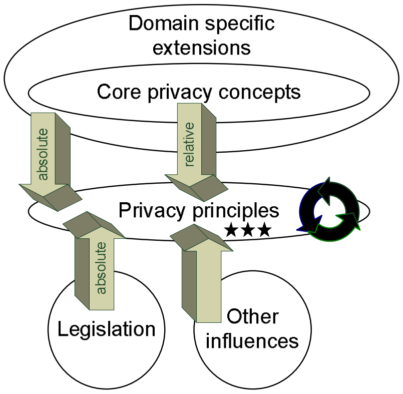

Figure 1: High level overview

From a very high level and abstract point of view, the core privacy concepts are the heart of the ontology. These concepts are domain independent to a great extent to be as general as possible. They include concepts which set up the basic ideas behind privacy, namely entities whereas the data subject is the most important one, resources, which are about a data subject and the resource users accessing the data subject's resources.

By definition, privacy is about empowering data subjects to regulate access to their data and therefore the core ontology must include necessary concepts to support this. They include concepts such as abstract security and consent mechanisms as well as privacy policies and support for legislative systems. The core ontology however does not specify any absolute values with regards to privacy levels gained. It only specifies relative and fuzzy values that will be taken in consideration by an evaluation engine, when the level of privacy is evaluated. The second part of the ontology is comprised of a conceptual view of the different privacy principles. Naturally, as we want and need to support privacy at its best, we derive the privacy principles from the OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data [6] as well as the Directive 95/46/EC of the European Parliament and the Council of the European Union on the protection of individuals with regard to the processing of personal data and on the free movement of such data [7]. The later one is a more specific and extended specification of the former one and includes more details and definitions about the different concepts of privacy and their users. We acknowledge that other definitions and directives of privacy in other countries or jurisdictions exists, but this one is the most comprehensive to our knowledge and many of the other ones use parts of them or can be mapped to these ones.

The concepts that build up the privacy principles are an essential part of the privacy ontology, as they are the actual parts that make it possible to determine the level of privacy. The privacy principles can be distinguished by twelve different concepts, whereas each concept has different dimensions that are influenced by the core privacy concepts as well as by each other. For example, privacy principle a) might influence privacy principle b) in a certain way. That also means that the privacy principles are not orthogonal, mainly due to the endogenous conception of privacy. However, the core privacy principles are not just influenced by each other, but also by the domain of application and related external factors.

Important components of the core privacy ontology are conceptual processes that specify certain actions with regards to the usage of personal identifiable information. We call the main concept of these processes "Privacy Process", as they are all governed by privacy policies, which have been set up by the data subject to protect his/her own information respecting his/her own values and goals. They may have been established in conjunction with the controller (which is the concept that will use the data later – see following sections for more information).

As we have previously mentioned, privacy cannot be achieved on a technological basis only, but requires strong legislative support. This is especially true in non-digital environments where security mechanisms such as encryption for example do not apply. As we design the ontology to be useful in digital and real-world environments, we have to take this into account. Legislative support as defined by the laws and regulations in certain jurisdictions. The core privacy principles associate a certain jurisdiction with every entity. When the privacy is evaluated, this link is followed and the evaluation would check the actual jurisdiction the receiver of personal information resides in and what legislation applies. As legislation and regulations can be very fuzzy and even conflicting with each other when multiple laws apply to single jurisdictions (e.g. state and federal legislation) and priorities have not been determined, it is difficult to actually compute how privacy is influenced. For the sake of simplicity in this paper, we assume that privacy should always be as strong as possible and in best interest of the data subject, even if this might not always be true in reality. Therefore, when we provide actual concepts for legislative support, we only use the ones that enhance or support the privacy of the data subject in the best way. Needless to say that the actual concepts of legislation for European countries are very close to the privacy principles, as we derived them from Europe's privacy legislation. It is interesting to see however, what other kind of privacy protection laws apply to other countries (e.g. India that has very weak privacy protection laws).

As mentioned above, the core privacy concepts are domain independent and at a high level of abstraction and therefore influence the privacy principles in a particular domain. To be able to be used in that domain, it will be necessary to elaborate and / or specialise these core concepts for that domain, while the specialisations will commit to the core concepts. When we want to use the privacy ontology in a certain domain (e.g. health care or financial area), we have to create an extension of the core privacy principles, by utilising ontologies of that target domain. A domain expert would be required to create such an extension for the very first time it is used in that area and it might be subject to change over time as knowledge and understanding of the domain changes. Obviously, such an extension is highly dependent on the experience and skill level of the domain expert, but the eventual outcome should not differ too much if different "experts" create different extensions for the same domain. This should be particularly true in essential questions about privacy levels. This is also another reason while the influential parts of the ontology are not based on interval or ratio scales but rather rely on ordinal and perhaps fuzzy characteristics as opposed to strict numerical values. Privacy is still an endogenous conception and some people may put more value or importance on certain concepts than other people.

To conclude the overview of the privacy ontology, it is important for us to mention that we try to use established concepts from other domains, especially concepts from other ontologies such as the standard upper ontologies (e.g. [10]) or at least provide a mapping for them.

4. Privacy concepts

This section will now elaborate the core privacy ontology and the concepts involved. Furthermore, we will show how the concepts influence the privacy principles (1) Intention and Notification, 2) Transparency, 3) Finality, 4) Legitimate Grounds of Processing, 5) Data Quality, 6) Data Subject's Rights / Individual Participation, 7) Safeguards, 8) Processing by a Processor, 9) Transfer of data to a different jurisdiction, 10) Accountability, 11) Openness and 12) Anonymity) as elaborated in [8].

4.1. Core privacy concepts

Going back to the definition of privacy, the general idea is to control the access and use of personal identifiable information by the data subject or an entity authorised by them. Therefore, the ontology is based around the concepts of "Data Subject", "Resource" and "ResourceUser" that accesses this "Resource". Every concept of the generic privacy ontology has certain attributes, which may or may not contain actual values that describe how the concept influences the overall level of privacy by influencing the privacy principles. Absolute values may and are usually not be possible, as it is domain unspecific and therefore not necessarily known how big the actual impact may be.

Figure 2 depicts the main associations related to the data subject. It states that the data subject is an entity and has or controls resources, which are obviously about him or her. The controller is also an entity that wants to access the resources of the data subject. The remainder of the ontology will be based around this simple set of concepts, while we guide the reader through the extensions we are describing in this paper.

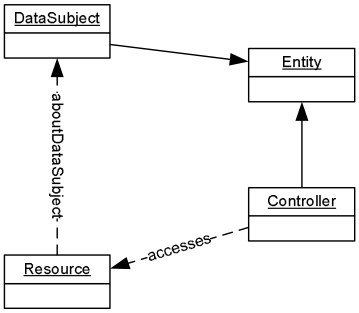

Figure 2: Resource, Data Subject, Controller

In order to distinguish between different types of entities and the participants of the system with their respective functionalities and properties, we will now show the hierarchy of entities (Figure 3). An entity is the most general type of cognitive agent (as specified by upper ontologies). The concept of entity represents a "Representative", which is the generalisation of either an individual or a group. As every group consists of one or more entities, it is possible to have entities that represent groups, individuals or any kind of mixtures between them. We need this kind of conceptualisation to support more specialised concepts of entities, which may be actually more than *one* real or legal person. Traversing down the entity specialisation tree, the next concept is a "ResourceAccessor", which is an entity that knows about a certain resource, but does not have the actual information. The concept following is the "ResourceUser", which can obtain resources and is regarded to be the recipient of resources in general. Nevertheless, this concept does not have any permission to actually access and read, alter or delete resources or part of them. Another more specialised concept of "ResourceUser" is the so called "ResourceAuthoriser", which can grant or revoke access to certain resources. Implicitly, a data subject is such a resource authoriser for their own resources. However, we need to distinguish a bit further here, as a data subject can be alive or non-alive. With alive and non-alive we literally mean living or dead (with regards to human beings), because it seems to be difficult for a dead person to authorise someone else to access their resources. A more important point has to be taken into account: legislation in some countries (e.g. Australia) terminates the protection of privacy once a person has deceased. A data subject is naturally also a resource modifier, which is a composition of concepts, incorporating reading, altering and deleting abilities. Finally, another concept is the "ResourceHandler", which is a "ResourceReader" and a bit more. It can not just read the data but also "transform" then into a different format without loss of semantics. For example, this could be an interpreter translating information from one language into another.

In order to support identities, anonymity and similar concepts, which are essential for certain privacy principles, we need to introduce identities on an entity level. We limit ourselves to a few categories at a high level of granularity for this paper only. Figure 4 shows how we proceed. Every entity has an abstract concept of identity, which in turn can be either a non-identifiable identity or an identifiable one. The non-identifiable identity can be broken down into three different more specialised concepts. The most anonymous one with regards to privacy protection is the "NoIdentity" one, followed by the anonymous identity. In most cases, the "NoIdentity" one will be equal to the anonymous one, but it might not be true universally. The last one is the pseudo anonymous identity, which entities might use when they don't want to reveal their true identity, e.g. by using nicknames. An identity by itself however does not make any assumption about the real identity of the entity. Resources can usually be classified to either identify an identity (not an entity!) or not. Hence, as we have different types of identities, we also need different types of resources that match those identities. In reality however, it might be very difficult sometime to classify if a resource element identifies a certain identity.

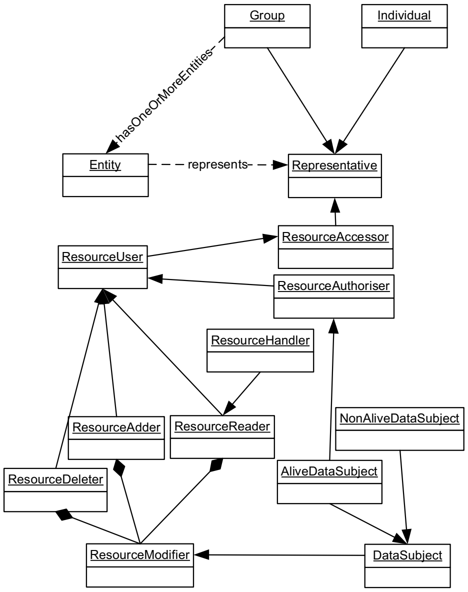

Figure 3: Hierarchy of Entities

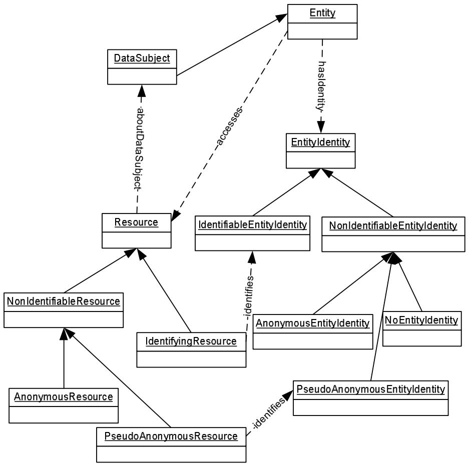

Figure 4: Identities & Resources

This hierarchy of concepts already creates some influences in the privacy principles. One can easily see that anonymous identities influence privacy principle 12 in a positive way, hence protection of privacy. That means that transactions that do not require identifiable identities (both, resources and actual identities – ie for authentication purposes) automatically yield a higher rating in terms of level of privacy achieved.

Every entity belongs to a certain jurisdiction in general and hence the laws and regulations of that particular jurisdiction apply. Inherently, the jurisdiction has a direct influence on privacy principle 9 and its related dimensions. Our figures in this section omit the existence of the jurisdiction and the laws applicable, but will be described in an actual example in a later section.

The few concepts that we have just described are not shown with all their details and associations between each other due clarity issues and lack of space in this paper. The full and regularly updated version can be obtained from the authors' webpage [11].

4.2. Privacy processes

The second part of the core concepts consists of the privacy processes, which are concepts that are related to any kind of usage, processing or sharing of personal information by entities other than the data subject.

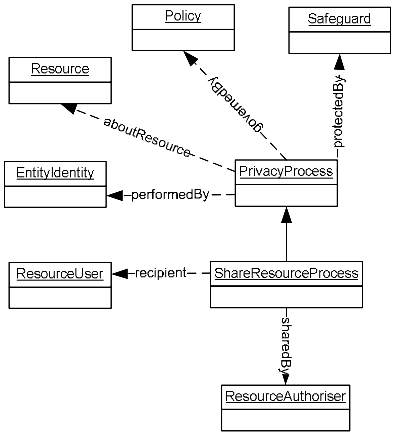

Figure 5: Privacy Process

As Figure 5 depicts, every privacy process involves a certain resource and is performed by some entity. Furthermore, it is governed by some sort of privacy policy and protected by safeguards (while no safeguard is regarded as a safeguard as well). The recipient of such a process is not necessarily the same entity as the one that performs it. One of the actual resource processes is the "ShareResourceProcess", which dimply states that a resource is shared by a certain entity (a "ResourceAuthoriser") with a particular recipient. From the previous elaboration of concepts, this already tells us that the recipient may not be able to actually read the data, but this is dependent on the policies and permissions of the data subject or legal regulations if any.

5. Conclusion and further work

In this paper, we explained the privacy principles and looked at the way these influence the level of privacy. These are used as the basis for deriving the key concepts and relationships for a generic ontology for privacy. We provide an illustrated example based on an actual website for obtaining ringtones for mobile phones.

A number of issues will be the subject of future research and these include:

- Identification of key dimensions in each privacy situation and their assessment
- Privacy preferences as expressed by the user and their impact on privacy principles when navigating through different application domains
- The use of CCCI metrics [13] as a basis for privacy calculations

6. References

[1] Merriam-Webster Online Dictionary: “Privacy”, accessed 08/2006.

[2] S. D. Warren and L. D. Brandeis, "The Right To Privacy," Harvard law review, vol. 4, pp. 193-220, 1890.

[3] Privacilla, "Privacy Fundamentals: Privacilla's Two-Part Definition of Privacy," 2003.

[4] G. Radwanski, Privacy Commissioner of Canada, “Patient Privacy in the Information Age”, E-Health 2001: The Future of Health Care in Canada, May 29, 2001

[5] World Wide Web-Consortium, "Platform for Privacy Preferences (P3P) Project," 2004.

[6] OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. Council of the OECD. 23rd September 1980. Available at: www.oecd.org/document/18/0,2340,en_2649_34255_1815186_1_1_1_1,00.html

[7] “Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data” http://www.cdt.org/privacy/eudirective/EU_Directive_.html

[8] PRIME, "Privacy and Identity Management for Europe – PRIME White Paper," 2005

[9] J. Huizenga, Handbook of Privacy and Privacy-Enhancing Technologies: The case of Intelligent Software Agents. The Hague: College bescherming persoonsgegevens, 2003.

[10] I. Niles and A. Pease, “Towards a Standard Upper Ontology”, Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001), Ogunquit, Maine, October 17-19, 2001

[11] http://exel.it.uts.edu.au

[12] Robert Frances Group, “The Challenge of Global Privacy Regulation Differences”, 2001, http://www.rfgonline.com/subsforum/archive/daily/090401/090701nt.html.

[13] E. Chang, T. S. Dillon, F. Hussain “Trust and Reputation in Service Oriented Environments”, John Wiley and Sons Ltd, Chichester, 2006