Data Handling: Dependencies between Authorizations and Obligations

Laurent Bussard
European Microsoft Innovation Center Aachen, Germany

Gregory Neven
IBM Research Zürich, Switzerland

Jan Schallaböck
ULD Kiel, Germany

Authorizations and obligations are keystones of data handling. On one hand there are ambiguous links between authorization and obligations. On the other hand a clear separation between both concepts is necessary to improve readability and to avoid inconsistencies.
This position paper focuses on authorizations necessary to enforce obligations. Such authorizations are necessary to prevents overdiligent data controllers from "overdoing" their obligations to the extent that they become a nuisance to the data subject. This problem is discussed from a legal perspective and is addressed in a technical solution that keeps a clear separation between authorizations and obligations.

1 Introduction

Protecting privacy-sensitive information by means of policies involves a combination of *access control* and *usage control* policies. Whereas access control specifies the conditions to be met by a prospective data controller *before* the data is revealed, usage control specifies how the data is to be treated *after* it is revealed. Usage control restrictions can be further divided in authorizations and obligations, where

- an *authorization* is the right to perform a certain action on the data, e.g., forwarding the data to selected business partners. Executing an action that is not explicitly authorized by the policy is a violation of the policy. Not executing an authorized action, however, does not violate the policy.

- an *obligation* is the duty to perform a certain action, e.g., deleting the data after a certain amount of time. Not executing an obligation imposed by the policy is a violation of the policy.

While the above definition may give the impression that authorizations and obligations are separate, orthogonal parts of a policy, in fact, a number of subtle dependencies can arise between them.

1. An obligation to perform a certain action can often be rephrased as the authorization to perform a complementary action. For example, the obligation to delete the data within one month could also be expressed as the authorization to store the data for at most one month. Vice versa, an authorization for a certain action can be seen as the obligation to not perform any complementary actions. For example, the authorization to use data for a specific purpose is equivalent to the obligation not to use the data for any other purposes.

   The PrimeLife Policy Language (PPL) [3, 5] avoids this issue by strictly separating the vocabularies for authorizations and obligations, so that no action defined in one vocabulary has a complement in the other vocabulary. SecPAL for Privacy [1] does not separate vocabularies and requires each obligation to be explicitly authorized.

2. The execution of an authorized action may trigger the execution of an obligation. For example, a policy could state that each access to the data for a particular purpose (authorization) needs to be logged (obligation).

3. Adhering to an obligation requires that the data controller is also authorized to do so. For example, an obligation to notify the data subject when the data is accessed requires that the data controller is actually allowed to contact the data subject.

This position paper focuses on the latter dependency between authorizations and obligations. In particular, we give a legal perspective (see Section 2) and a technical solution (see Section 3) to prevent over-diligent data controllers from "overdoing" their obligations to the extent that they become a nuisance to the data subject or degrade the quality of service. For example, a data subject who insists to be emailed access reports for its data at least once per year may experience daily emails as spam. Worse even, the exaggerated enforcement of the obligation largely defeats its original purpose of giving the data subject an overview of the accesses to her data.

2 Legal Aspects

Legal compliance often requires an authorized entity to meet certain obligations when processing the information, it has been given authorization for. Such requirements can be of contractual nature, but often are also required by law directly. Typical scenarios for such obligations are in the field of processing personal data, where data protection regulation applies. Although the latter is in the focus of this paper, this may also apply in many areas of information regulation such as copyright or protection of business secrets, to name a few.

A standard use case for this could be taken from the area of credit scoring. While the legal framework for credit scoring holistically would be reaching to far for this paper and implementations vary, we will consider the following as giving. In credit scoring a scoring company is provided with a set of relevant information on loans, credit cards and bank accounts in a regular basis by the providers of the latter. Based on this data it calculates probabilities and risks of nonpayment, based on its statistical empirics. This information is then given to a party before entering into a contract with the person the data is collected on (the data subject). It such sustains confidence in the contract for the former party and makes it easier to calculate the risks involved.

European Data protection law (ie. member states implementations of the privacy directive 95/46/EC) prohibits the processing of such personal data unless a specific legal basis is provided. The legal basis for such processing and accessing of information is usually of contractual nature or based on consent by the data subject. In other words the latter consented that the company doing the credit scoring would be provided with the information mentioned beforehand by the parties involved, and equally accepts access to this information for entities considering to enter into a contract with the data subject.

Obviously erroneous data in this data in this data set, can have a high impact on the data subject therefore the correctness is of high relevancy. Due to the fact, however that the data is not collected from the data subject itself, nor is it usually transmitted to her, the system is prone to errors. A recent study indicates that only little more than half of the data subjects scored with the largest German credit scoring agency contain no errors [4].

An easy approach for reducing such errors would be to regularly inform the data subject on changes in the data set. Depending on the activity of the data subject, these changes could appear quite often, which yields the risks, that the data subject would loose track of the changes and would ignore it, similar to spam. Similar difficulties has been shown for SSL warnings, cf [6]. It may therefore in this case be sensible to attach an obligation to the policy for the credit scoring data, that the data be sent to the data subject on a regular basis, but not too often.

Again this processing would call for a legal basis, as this again is processing of personal data, with another specific purpose. This in turn would not only give the right for such processing to the processing entity, but would usually be a legally binding obligation to this party. One legal basis could be the law, even in the near future, as including a legal obligation for informing data subjects are currently under consideration [2]. Another option could be contractual. In this case a credit scoring company could derive market advantage by ensuring the data subjects be informed about the information stored by them. Thus the information would allow higher quality, while strengthening control and transparency for the data subjects.

A number of problems attached to this problem, however, cannot be discussed herein. For example, what if an individual does not want to be informed? Might she have a "right to not know"? In the case of the results of health exams, such rights have been constructed previously. Another question is the one of identification of the data subject when transmitting the data, cf. [7]. Providing the wrong person with information on the data subject may yield serious consequences for the processing entity, including, but not limited to, liability.

3 Technical Aspects

3.1 The PrimeLife Approach

The PrimeLife Policy Language (PPL) defines a format in which a data controller can specify its data handling policy, describing how it intends to treat personal data item after it is received, and a format in which a data subject can specify her preferences, describing how she expects her personally to be treated after it is transmitted. Moreover, PPL defines an automated matching procedure by means of which the data subject can test whether a proposed data handling policy is compatible with her preferences.

Both a data handling policy *Pol* and preferences *Pref* consist of a set of authorizations *Auths* and a set of obligations *Obls*. Two individual authorizations or individual obligations can be compared against each other by means of a partial order "more permissive than" ($\trianglerighteq$) that is defined over the vocabularies of authorizations and obligations, respectively. Authorizations and obligations can be parameterized. For example, the authorization $\mathit{UseForPurp}(P)$ takes the set of purposes for which the data is intended (in the policy) or is allowed (in the preferences) to be used as a parameter. The obligation $\mathit{DelWithin}(t)$ takes the time within which the data will or has to be deleted as a parameter, while the obligation $\mathit{NotifyAtFreq}(f)$ specifies the minimal frequency at which usage reports have to be or will be sent to the data subject.

Matching individual authorizations or obligations usually involves comparing parameter values. For example, we have that

$$\begin{aligned} \mathit{UseForPurp}(P) \trianglerighteq \mathit{UseForPurp}(P') &\Leftrightarrow P \supseteq P' \\ \mathit{NotifyAtFreq}(f) \trianglerighteq \mathit{NotifyAtFreq}(f') &\Leftrightarrow f \leq f' \\ \mathit{DelWithin}(t) \trianglerighteq \mathit{DelWithin}(t') &\Leftrightarrow t \geq t' \ . \end{aligned}$$

A policy *Pol* is said to *match* preferences *Pref*, denoted $\mathit{Pref} \trianglerighteq \mathit{Pol}$, if and only if for all authorizations in the policy there is a more permissive authorization in the preferences, and for all obligations in the preferences there is a less permissive obligation in the policy, or more formally,

$$\begin{aligned} &\mathit{Pref} \trianglerighteq \mathit{Pol} \ \Leftrightarrow \\ &\quad (\forall\ A' \in \mathit{Pol.Auths} \cdot \exists\ A \in \mathit{Pref.Auths} \cdot A \trianglerighteq A') \\ &\quad \wedge\ (\forall\ O \in \mathit{Pref.Obls} \cdot \exists\ O' \in \mathit{Pol.Obls} \cdot O \trianglerighteq O') \ . \end{aligned} \tag{1}$$

3.2 An Improved Approach

As stated before, PPL avoids some of the dependencies between authorizations and obligations by keeping their vocabularies strictly separate, ensuring that obligations cannot also be expressed as authorizations for a complementary action and vice versa. Moreover, this separation of concerns simplifies policy matching. Also, PPL implicitly assumes that by proposing (imposing) an obligation in the policy (preferences), the data controller (data subject) also implicitly requests (grants) the authorization to perform the action needed to adhere to the obligation. We will make the same assumptions in the approach presented here.

In spite of these assumptions, however, PPL does not solve the specific issue of data controllers becoming a nuisance to data subjects by overdoing their obligations. For example, a policy committing to send the data controller usage reports at least once per month matches preferences requiring it to do so at least once per year, because the proposed behavior is included in the expected behavior. Moreover, the data controller can notify (i.e., spam) the user on a daily basis without violating the policy.

3.2.1 Individual obligations.

The key observation here is that the problem arises because of data controllers deviating from the stated obligation parameters on the side of more privacy-friendly, but possibly more annoying values. The solution we propose is therefore that obligations, rather than specifying only the least privacy-friendly permitted value for each parameter, specify the full range of permitted values. A small range can be described simply by exhaustively enumerating its elements. If the range is a contiguous interval, it is most efficiently represented by its endpoints. For example, a data subject's preferences could contain the obligation

$$\begin{aligned} \mathit{NotifyAtFreq}(\{f : \text{once per year } \leq f \leq \\ \text{once per 3 months}\}) \ , \end{aligned}$$

specifying that the user wants to be notified at least once per year, but at most once per three months. To match it against a corresponding obligation in the policy, one has to verify that *all* frequencies allowed by the policy are also allowed by the preferences. For example,

$$\begin{aligned} \mathit{NotifyAtFreq}(\{f : \text{once per year } \leq f \leq \\ \text{once per 6 months}\}) \ , \end{aligned}$$

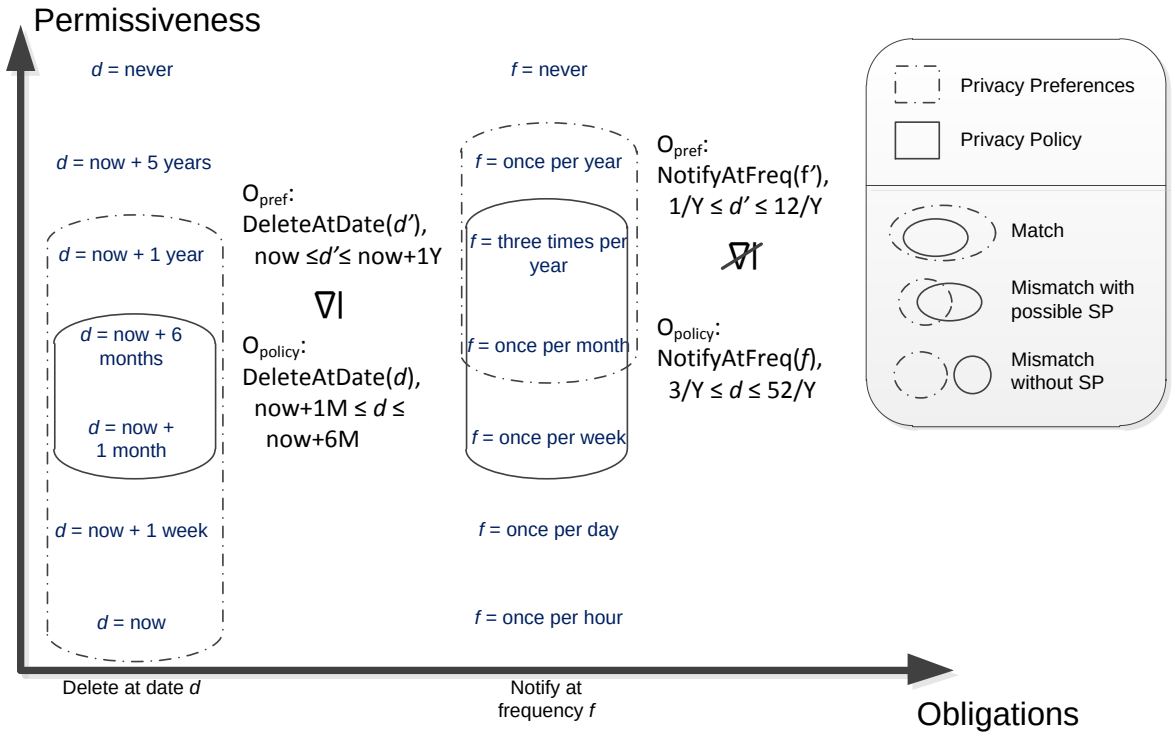

Figure 1: Examples of matching and mismatching obligations

matches the preferences, but

$$\textit{NotifyAtFreq}(\{f : \text{once per 6 months } \leq f \leq \text{once per month}\}) ,$$

does not match, because the data subject does not want to be bothered by monthly reports.

Figure 1 provides a graphical representation of matching and mismatching obligations with one parameter. In general, our approach suggests to convert any obligation $O(p)$ with parameter $p \in \Pi$ into an obligation $O(P)$ with parameter $P \subseteq \Pi$, where P denotes the range of permitted values. Matching is defined by

$$O(P) \trianglerighteq O(P') \Leftrightarrow P \supseteq P' . \tag{2}$$

This definition naturally extends to obligations with multiple parameters $O(p_1, \ldots, p_n)$ from possibly different domains $\Pi_1, \ldots, \Pi_n$. Namely, our approach converts it into a new obligation $O(P_1, \ldots, P_n)$ where $P_i \subseteq \Pi_i$ specifies the range of permitted values for p_i. The obligation can be matched by checking

$$O(P_1, \ldots, P_n) \trianglerighteq O(P'_1, \ldots, P'_n) \Leftrightarrow P_1 \supseteq P'_1 \wedge \ldots \wedge P_n \supseteq P'_n . \tag{3}$$

In fact, matching definition (3) can be seen as a special case of definition (2) by viewing the vector $p = (p_1, \ldots, p_n)$ as a single parameter and taking the Cartesian product $P = P_1 \times \ldots \times P_n$ as the range of permitted values for p.

3.2.2 Sets of obligations.

Above, we argued that data controllers adhering to stricter obligation parameters than necessary can be experienced as a nuisance by data subjects. We therefore replaced simple parameter values with ranges of values, which affected the matching definition of individual obligations. By the same reasoning, data controllers adhering to *more* obligations than necessary can also be experienced as a nuisance. We propose a similar solution that will affect the overall matching definition (1) of preferences against policies.

The direct analog of our solution for obligation parameters would be that, instead of specifying a set of obligations *Obls*, preferences and policies specify *ranges of sets* of obligations *OBLS*. Matching preferences and policies would then involve checking that for every set of obligations

$Obls' \in Pol.OBLS$ in the policy there exists a set of obligations $Obls \in Pref.OBLS$ in the preferences such that $Obls \trianglerighteq Obls'$.

This approach gets unwieldy for even simple policies, so we propose an alternative mechanism here. Rather than containing a single set of obligations, both the preferences *Pref* and the policy *Pol* specify a set of *mandatory* obligations *MObls* and a set of *optional* obligations *OObls*. The semantics are that the data consumer insists that the obligations in *Pref*.*MObls* are adhered to, and can live with additional obligations *Pref*.*OObls* being adhered to. In the policy, the data controller commits to adhering to obligations *Pol*.*MObls*, and may or may not adhere to obligations *Pref*.*OObls*. Effectively, the sets *MObls* and *OObls* describe an "interval" of permitted sets of obligations that can vary between *MObls* and $MObls \cup OObls$. The overall matching definition is now given by

$$
\begin{aligned}
&Pref \trianglerighteq Pol \;\Leftrightarrow \\
&\quad (\forall\, A' \in Pol.Auths \cdot \exists\, A \in Pref.Auths \cdot A \trianglerighteq A') \\
&\quad \wedge\; (\forall\, O \in Pref.MObls \cdot \exists\, O' \in Pol.MObls \cdot O \trianglerighteq O') \\
&\quad \wedge\; (\forall\, O' \in (Pol.MObls \cup Pol.OObls) \\
&\qquad \cdot\; \exists\, O \in (Pref.MObls \cup Pref.OObls) \cdot O \trianglerighteq O')\ .
\end{aligned}
\tag{4}
$$

4 Conclusion

This position paper shows that implicitly assuming the authorization of enforcing obligations may threaten privacy. We present an approach that solve this issue while keeping a clear border between obligation and authorization vocabularies. Authorizations to enforce obligations are integrated by specifying the set of acceptable values for each obligation parameter.

Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 216483 for the project PrimeLife.

References

[1] M. Y. Becker, A. Malkis, and L. Bussard. S4P: A Generic Language for Specifying Privacy Preferences and Policies. Technical Report MSR-TR-2010-32, Microsoft, April 2010.

[2] K. Biermann. Datenbrief wird ernsthaft beraten. online, April 2010.

[3] L. Bussard, G. Neven, and F.-S. Preiss. Downstream usage control. In *Policies for Distributed Systems and Networks, 2010. POLICY 2010. IEEE International Symposium on*. IEEE, 2010.

[4] D. Korczak and M. Wilken. *Verbraucherinformation Scoring*. Bundesministerium für Ernährung, Landwirtschaft und Verbraucherschutz, 2010.

[5] PrimeLife Project. Draft 2nd Design for Policy Languages and Protocols (Heartbeat: H 5.3.2). Technical report, July 2009.

[6] J. Sunshine, S. Egelman, H. Almuhimedi, N. Atri, and L. F. Cranor. Crying wolf: An empirical study of ssl warning effectiveness. usenix security, 2009.

[7] M. Temmerman, J. Ndinya-Achola, J. Ambani, and P. Piot. The right not to know hiv-test results. *Lancet*, 345(8955):969–70, Apr 1995.