

Privacy surviving Data Retention in Europe?

W3C Workshop on Languages for Privacy Policy Negotiation and Semantics-Driven
Enforcement

Helena Lind, Johan Hjelm, Mikael Lind

1 Introduction

This paper describes our position regarding standardization of a Privacy Policy Negotiation Language with Semantics-Driven Enforcement.

In this paper we describe a system using such mechanisms that has been designed and implemented within the MobiLife project. We also look at the legislative requirements which we and many of our customers will need to adhere to. We also make some recommendations regarding future work of W3C.

2 The MobiLife Project

The MobiLife project has been performed in the framework of the IST project IST-2004-511 607 MobiLife, which is partly funded by the European Union.. The goal is to develop a virtual social environment for groups of individuals, where the individuals can cooperate, communicate, and share information. A central part of this is a rules engine performing similar functions to what is discussed in this workshop. This allows seamless interaction between different types of devices, cooperation within closed groups for the enterprise case, interoperability between operators, and interoperability between operators and enterprises.

User preferences for privacy management seem to indicate that up until a threshold value (which is entirely subjective), the user does not care. When that threshold is breached, privacy becomes very important. However, it also seems that there is a need to distinguish different data from each other, since the user may be more willing to share certain data types in certain situations. Policies must be adaptable to user constraints based on granularity [MobiLife, IBM].

In the MobiLife project, we have developed a model, the “trust engine”, which is a PEP and PDP which sits in front of the data source, and determines which data should be given out, and to whom.

The Policy Language used by the Trust Engine has to be powerful enough to express, for every entity (user or group):

- **What data** is shared (e.g., location, heart rate, favourite colour, age)
- **With whom** the data can be shared. We support three categories, and combinations thereof:
 - Share data with anybody (e.g. no restrictions)
 - Share data with one or more (named) users or applications
 - Share data with one or more groups
- **Under what condition** the data is shared:
 - Always

TAKING YOU FORWARD

- Never
- After approval

One important aspect is that the query language allows for easy retrieval of (part of) the policies. For example, if a Trust Engine needs to decide whether the location of a user can be shared with the requesting entity, it needs to be able to quickly extract the relevant rules from the policy.

It is also important that the visualization of the policies reflects this. We should not standardize a visualization model, but the language itself will affect how it can be visualized.

Note that there is an additional constraint: That the policies can be overridden by an authorized party, such as an operator (which may be based on the written agreement between the user and the operator).

3 Legislation

The EU Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC, adopted March 15th 2006, poses new severe threats to end-user privacy. Until now, law enforcement agencies have always had the possibility to access telephony and computer log files and other information that is readily available. At a telecom operator site that would typically be billing records, security logs and traffic logs. At an ISP site, HTTP log files are always stored for security reasons. Firewalls, proxies and network intrusion detection systems typically pick up information that the police might benefit from using, be they in time before it is removed.

There is no hard proof that police authorities would benefit from having more than this data other than in extreme cases, but still, their claim for harmonized, long-term storage, raised a few weeks after the tragical Madrid bombing, was convincing enough to make the parliament adopt this directive. To our knowledge, proof of such a need has never been presented, nor a case where killers and abusers were not caught due to the lack of personal data - either when discussing the directive in Brussels, or by national investigations now writing the implementations of the directive. It is therefore of high importance that national implementations do not go further than the directive, requiring that data be retained that is not needed in order to comply with the directive.

The directive is vague - especially from a mobile Internet perspective. The directive distinguishes between telephony services and Internet services, but no distinction is made for mobile Internet services/access.

Here is an example of a white spot in the directive, taken from the first section where the data to be retained is specified:

- (a) data necessary to trace and identify the source of *a communication*:
 - (1) concerning fixed network telephony and mobile telephony:
 - (i) the calling telephone number;
 - (ii) the name and address of the subscriber or registered user;
 - (2) concerning Internet access, Internet e-mail and Internet telephony:
 - (i) the user ID(s) allocated;

TAKING YOU FORWARD

(ii) the user ID and telephone number allocated to any communication entering the public telephone network;

(iii) the name and address of the subscriber or registered user to whom an Internet Protocol (IP) address, user ID or telephone number was allocated at the time of the communication;

You do not even have to use the mobile Internet example to realize that here is a huge question mark: "What is *a communication*? Is it an HTTP request? Is it a GPRS session? When using e.g. a chat service - would it be every chat message?" The term communication is not defined within the directive, and today, police authorities are fighting telecom operators and ISPs for the most favorable interpretation.

Whenever huge masses of personal data are stored at one place, and especially when tied to a system with the intelligence to tailor this data, there is an enormous privacy risk. The idea is that strict access control surround the data. Will that be the case? We can only hope. We see a risk of abuse from corrupted personal and from hackers or other intruders. Also, there is a risk that data be overly interpreted as true, and that end users be wrongly accused. With the ease in accessing and perhaps performing data mining on huge amounts of personal data, the risk that a police investigation might take the wrong turn is much greater than today, where you can ask simple questions, such as "Did Peter Pan call Tinkerbell last Friday?" Apparently this argument can also be used to suggest that more crimes be solved, due to the same new possibilities to play around with the data.

What information will be stored and the issue of how the data is accessed and delivered according to the directive is standardized by [ETSI].

It is possible to partly protect the end user. However, this is risky, since police authorities might consider it a way to help criminals evade the law, so we would not recommend W3C to work in this direction. However, W3C will have to take the various national implementations of this directive into account, whenever dealing with end user privacy. We also anticipate that agents encrypting end to end will be of much more interest to end users.

4 Summary

The EU Directive 2006/24/EC is being implemented by the national states throughout Europe. It is very likely that the implementations will vary between nations, creating a set of laws that may be very hard to match with a Negotiation Language. Thus, it is of great importance to ensure coherent adaptations of this Directive.

For Ericsson, a Privacy Policy Negotiation Language and Semantics-Driven Enforcement standard should support the following

- The implementations of the legislative requirements that will be derived from EU Directive 2006/24/EC.
- Interoperability between different actors and devices.

TAKING YOU FORWARD

- Coherent mechanisms that support both consumers and enterprises and enterprise users.
- The policy language needs to be able to express at least what data is shared; with whom; and under what conditions. The language also needs to be simple, and support retrieval of parts of policies.

5 References

- [ETSI]
http://portal.etsi.org/LI/LI_ToR.asp
- [IBM]
<http://www.acmqueue.org/modules.php?name=Content&pa=showpage&pid=410>
- [MobiLife]
www.ist-mobilife.org