D.3.1 – Privacy Breach Scenarios in SocioPlug

Patricia Serrano-Alvarado

D.3.1 – Privacy Breach Scenarios in SocioPlug

Patricia Serrano Alvarado

[1]*LINA UMR 6241 / Universite de Nantes.*
Email contact: Patricia.Serrano-Alvarado@univ-nantes.fr

ANR-13-INFR-0003
socioplug.univ-nantes.fr

1 Introduction

In SocioPlug, we have particular concerns about data protection. Services proposed by SocioPlug will conform to European regulations [REG08, Com12], during personal data collection and data access. In particular the right to oblivion, collection and access purposes should be explicitly determined by data owners.

SocioPlug's architecture is fully distributed and has no centralized server, thereafter, there is no centralized control about data and applications of users. The goal is to avoid the existence of a "big brother" vigilating every person of the social cloud. Nevertheless, collaboration implies accessing personal data of other users. As services and data will be distributed in a social cloud, participants must be responsible of their data but also of other's data they collect and use. Thus, they must define usage policies for each shared data and people that collects and uses other's data must preserve stated policies.

From application scenarios described in deliverable D.4.1, in this report, we identify some important privacy breach scenarios that may appear in SocioPlug.

2 Breach scenarios

2.1 Breach Scenario 1. Data collection and data usage without stating usage policies

All application scenarios of SocioPlug manage personal data.

- PlugTrack collects user's location from a mobile device and upload it on the personal plug computer of the user. To improve recommendations, PlugTrack collects also other users' location. These information is used for recommendations to people having similar itineraries or interests.
- Collaborative editor produces documents that can be shared and used by people collaborating in the document edition but also by non-editors.
- Smart city application is based on publishing and querying streams of data coming from several user's sensors.
- Collaborative content sharing application allows users to share files with other users. It proposes also to create dynamic communities of users who have shown similar interests. The idea is to disseminate content to a target group of users.

Privacy of users will be transgressed if data used by these applications is collected, used and stored arbitrarily by anyone. Thus, to preserve privacy, it is of utmost importance to define and attach a usage control policy to each data used. Usage policies should contain the context under which data can be used, i.e., which application or user can use it, for how long, for which purposes, in which location, which operations are allowed, what is prohibited, what are the obligations when/after accessing data, when data should be deleted, etc.

SocioPlug users should carefully define their usage policies and applications should be encouraged to use data only if they have valid policies. Applications should take care also of suppressing data when duration time of usage arrives to expiration. Several languages exist to define *usage policies* as PriLoo[AD13], *privacy policies* like ODRL [Ian02], PPO [SP11], or L4LOD [VG13], and *licences* such as CC [Com01], GPL [Pro07], or OpenLicence [Pro11]. Definition of usage policies can use a combination these languages but the singularity of PriLoo is that the usage context is taken into account clearly. So we encourage to use PriLoo in SocioPlug applications.

2.2 Breach Scenario 2. Data combination without combining usage policies

Stating policies for each used data is not enough to deal with privacy. When collaborating, it is frequently necessary to combine several data. Thus usage policies of concerned data should also be combined.

- PlugTrack combines data of different users having similar journeys or itineraries to propose more pertinent recommendations.
- Collaborative editor may fusion several documents in a single document.
- Smart city application may query several streams for personal use or for dissemination.
- Collaborative content sharing application allocates in a same repository, files provided by different users.
- From a non-functional point of view, monitoring activity process, provided by Task 3 of SocioPlug, will merge streams of data, which are not produced by a single user, but by the architecture network activities.

In these cases, which will be the usage policy of the resulting data? What should be possible to do with data resulting of a data merging? To deal with this breach, it is necessary to have an approach to produce a usage policy from the combination of several usage policies. Some works propose solutions for policies combination like PrODUCE [SMSADGM15] or [GRVG13]. [GRVG13] proposes a deontic logic semantics to define the deontic components of licenses and generate a composite license compliant with the licensing items of the composed different licenses. PrODUCE proposes a different approach which is based on semantic web technologies to compose privacy policies. The originality of this approach is that composition rules are based on the data usage context and deduce implicit terms. SocioPlug applications should determine which approach better fits their needs of policies combination. We encourage them to use PrODUCE that is an approach easy to use and based on semantic web technologies (i.e., RDF and deduction rules).

2.3 Breach Scenario 3. Usage control

In a fully distributed architecture, as the one used in SocioPlug, participants are peers playing the roles of users and data providers. As there is no centralized entity, data providers participate in distributed query processing without knowing they actually do. Only the requester query engine knows the federated query it process. Thus, federated queries exist but, apart from federated query engines, nobody knows which data sources are joined. This is a usage control breach because data providers have no control about the distributed usage of their data.

- PlugTrack may produce a distributed query to simultaneously collect information of other PlugTrack users. But, each user does not know she participates in a distributed query, thus she does not know her information is combined with others' information.
- Collaborative editor may query other plugs about documents concerning a particular topic. But each plug owner does not know she participates in a distributed query.
- Smart city may produce also distributed queries to simultaneously query several users' streams. But, each stream owner does not know her stream is combined with other streams.

SocioPlug application	Breach scenario 1	Breach scenario 2	Breach scenario 3
PlugTrack	x	x	x
Collaborative editing	x	x	x
Smart city	x	x	x
Collaborative content	x		
Monitoring activities (Task3)	x		

Table 1: Table showing which breach scenarios occur in which application

Thus, how to know if a data provider is participating in a distributed query? And how to discover the concerned distributed query if query users (or query engines) do not publish them? This is an open issue but some works are emerging. FETA [NSAMD15], proposes a federated query tracking approach in the context of the linked data. Based on a set of heuristics, original federated Basic Graph Patterns are extracted from a shared log maintained by a federation of data providers.

2.4 Summary

In this report we presented some privacy breach scenarios that may appear in SocioPlug. For each on we mention some possible solutions to adopt/adapt. Table 1 shows these scenarios and the application they appear.

References

[AD13] Patricia Serrano Alvarado and Emmanuel Desmontils. Privacy-Lookout. `http://www.privacy-lookout.net/pluxml/`, 2013. Online; accessed 5-January-2016.

[Com01] Creative Commons. Creative Commons. `http://creativecommons.org/`, 2001. Online; accessed 5-January-2016.

[Com12] European Commision. Proposal for a regulation of the european parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (general data protection regulation). *COM (2012) 11 final, 2012/0011 (COD), Brussels, 25 January 2012*, 2012. Online; accessed 5-January-2016.

[GRVG13] Guido Governatori, Antonino Rotolo, Serena Villata, and Fabien Gandon. One License to Compose Them All. In *International Semantic Web Conference (ISWC)*, 2013.

[Ian02] Renato Iannella. Open digital rights language (ODRL) version 1.1. *W3c Note*, 2002. Online; accessed 5-January-2016.

[NSAMD15] Georges Nassopoulos, Patricia Serrano-Alvarado, Pascal Molli, and Emmanuel Desmontils. Tracking federated queries in the linked data. *arXiv preprint arXiv:1508.06098*, 2015.

[Pro07] GNU Projet. GPL General Public License, Version 3. `http://www.gnu.org/licenses/gpl-3.0.fr.html`, 2007. Online; accessed 5-January-2016.

[Pro11] Etalab Projet. Licence Ouverte, Open Licence. `https://www.etalab.gouv.fr/licence-ouverte-open-licence`, 2011. Online; accessed 5-January-2016.

[REG08] HAVING REGARD. Recommendation of the council concerning guidelines governing the protection of privacy and transborder flows of personal data. `http://www.oecd.org/sti/ieconomy/2013-oecd-privacy-guidelines.pdf`, 2008. Online; accessed 5-January-2016.

[SMSADGM15] Valeria Soto-Mendoza, Patricia Serrano-Alvarado, Emmanuel Desmontils, and Jose Antonio Garcia-Macias. Policies Composition Based on Data Usage Context. In *Sixth International Workshop on Consuming Linked Data (COLD) at ISWC*, 2015.

[SP11] Owen Sacco and Alexandre Passant. A Privacy Preference Ontology (PPO) for Linked Data. In *Workshop on Linked Data on the Web (LDOW)*, 2011. Online; accessed 5-January-2016.

[VG13] Serena Villata and Fabien Gandon. L4LOD Vocabulary Specification 0.2. `https://ns.inria.fr/l4lod/v2/`, 2013. Online; accessed 5-January-2016.