A Framework for Privacy-Aware User Data Trading

Johnson Iyilade and Julita Vassileva

Computer Science Department, University of Saskatchewan, 110 Science Place
S7N 5C9 Saskatoon, Canada
{Johnson.Iyilade,Julita.Vassileva}@usask.ca

Abstract. Data about users is rapidly growing, collected by various online applications and databases. The ability to share user data across applications can offer benefits to user in terms of personalized services, but at the same time poses privacy risks of disclosure of personal information. Hence, there is a need to ensure protection of user privacy while enabling user data sharing for desired personalized services. We propose a policy framework for user data sharing based on the purpose of adaptation. The framework is based on the idea of a market, where applications can offer and negotiate user data sharing with other applications according to an explicit user-editable and negotiable privacy policy that defines the purpose, type of data, retention period and price.

Keywords: Privacy, Personalization, User Data Sharing, Policy, Incentives, Trust, Market, Framework.

1 Introduction

The ability to share user data across applications, services and devices has become crucial to personalization recently, with the emergence of cloud-based services and mobile app ecosystems, where many independent applications, services, or devices are interacting with and gathering information about the same user. In the last couple of years we have witnessed unprecedented numbers and dynamics of use of different applications by end users. Users constantly install, use and uninstall, on the fly, apps on their smart phones and tablets. When a user interacts with a new application, she should not have to re-enter the same information. Sharing the information that has already been collected by other installed applications would save the efforts and time of the user [1] and will address the cold-start problem in the personalization of the new application [2]. More information about the user would be available for user modeling (UM), covering more aspects in the aggregated model by both applications, which would allow higher quality personalization.

Sharing user data across applications raises several challenges: (i) the architecture of the user model – centralized (aiming to collect all user data at one place, in a consistent database), or decentralized (aiming to facilitate applications to share user data directly with each other on demand), (ii) ensuring user model (semantic) interoperability, and (iii) respecting user privacy and enabling user control over her data. There has been a lot of work in the field of user modeling that addresses the first

S. Carberry et al. (Eds.): UMAP 2013, LNCS 7899, pp. 310–317, 2013.

two challenges. The existing work on privacy however, is focused only on centralized architectures, where there is one user model maintained for each user on a UM server, using data from and for use of many applications. However, there is no work yet on privacy in decentralized architectures [3], where the user data sharing is taking place directly across applications, or is mediated for semantic interoperability [2].

Decentralized UM architectures can be viewed as marketplaces consisting of many interacting applications, which can be viewed as user data providers, user data consumers, and user data brokers (which facilitate the user data sharing by providing mediation services e.g. semantic mediation and lookup). The user can and should be an equal player in such a marketplace. She can trade her data, and gain benefits from sharing it. This short paper proposes a framework that can enable such a marketplace for sharing user data among applications. In our approach, the users have control on how their data is used by classifying their data based on its relevance to different purposes of use and sharing. We formulate a policy specification language through which user data providers can communicate the purpose and conditions of use under which user data that they have collected can be shared with user data consumers. The users provide their data after adapting and accepting the policy. Thus, the application has a contractual obligation to respect the policy and responsibility can be sought from applications that deviate. The framework ensures flexibility in the contracts through negotiation of some elements of the privacy policy such as *retention time, type of user data* that can be shared *and price*. The aim is to achieve a market, on which user data can be securely exchanged, traded and the user get properly compensated.

2 Related Work

There are several challenges involved in sharing user models gathered across applications. Existing solutions have focused on addressing different aspects of these problems. A key issue is whether the architecture of the system should be centralized or not. In centralized user modelling architectures, the user data is collected from various applications and stored in a database located on a server [4], many servers [5], or on a cloud [6]. According to principles governing databases, the information is kept consistent, secure and available. A centralized representation schema is used to store user data; semantic interoperability across the user data schemas of the feeding and consumer applications is ensured by the server. Therefore, as noted by [7], the system is logically centralized even when the data is stored in a physically distributed way. Most existing frameworks for reusing and sharing user models across applications follow the centralized architecture (e.g. IDIUMS [8], Personis Server [9], UMoWS [10]). Privacy in UM sharing has been addressed in the context of centralized UM server architectures. User data residing on servers can be more easily secured, but it also presents an attractive target of hackers. On a higher level, Wang and Kobsa [10] propose a framework for enforcing privacy in user modeling servers during data collection based on user preference settings and on a combination of the information privacy laws of country where the server is located and where the user resides. Other work addressing privacy in centralized UM servers, emphasizes user control over the accuracy of her data. The Personis Server [9] enables the user to view and control what is stored about them through a scrutable interface. However, the focus is on primary data collection, and the user is not provided means of controlling the secondary use and sharing of their data (it is assumed that secondary use is not happening). But, with the rise of social networks and Web 2.0 applications, where users voluntarily produce massive amounts of data, the challenge is how to protect user privacy not just in primary user data collection but also in secondary sharing and reuse of data. In addition, the dominant technology of centralized UM servers assumes a closed environment, where applications are trusted.

Another challenge is that of ensuring user model interoperability. A comprehensive review of current state-of-art in user model interoperability is provided in [3]. Generally, solutions to the challenge of interoperability have coalesced under two themes: standard-based (e.g. [12]) or mediation-based (e.g. [2]) approaches using frameworks, such as *Mypes* [13] for aggregating user information distributed across social networks, particularly, folksonomy systems. The main features of mediation-based approaches include user account mapping, linkage and aggregation of profiles. Mypes [13] focuses on aggregation of data from many sources and does not address privacy issues. Yet, sharing in decentralized architectures raises serious privacy concerns, related to which applications the data can be shared (trustworthiness), the purpose of the secondary use of data, how long it can be kept and who should be held responsible in case of violations of privacy. The user needs to have means to actively control not only what information is stored by applications about them (for primary use), but also the purpose and conditions of the use of data, the retention of the data, whether and with whom it can be shared. Also the user needs to be given an incentive for sharing data across applications, as currently, apart from personalized service there is no benefit directly for the user if she chooses to allow her data to be kept and shared.

3 Framework for Sharing User Data across Applications

As the ability to collect, share, and aggregate user data becomes crucial for personalized service delivery, we foresee the emergence of a user data sharing marketplace where user data will be securely collected, shared and exchanged in a mutually beneficial manner for all the players. The marketplace involves four active player types: *the users, data providers* (applications that have collected user data and can share it with other applications), *data consumers* (applications that need user data for their own personalization purposes), *data brokers* (applications that ensure semantic operability, monitor the trustworthiness of data providers and consumers, and carry out the negotiation process for each sharing according to policies for user data sharing). To enable interactions among the various players, we present a policy framework for user data sharing across applications, called *Purpose-to-Use (P2U).* We further introduce a negotiation mechanism between the data consumers and data providers for P2U policy elements such as *data, retention time* and *price*. Finally, we discuss how users will be compensated for data sharing in the marketplace.

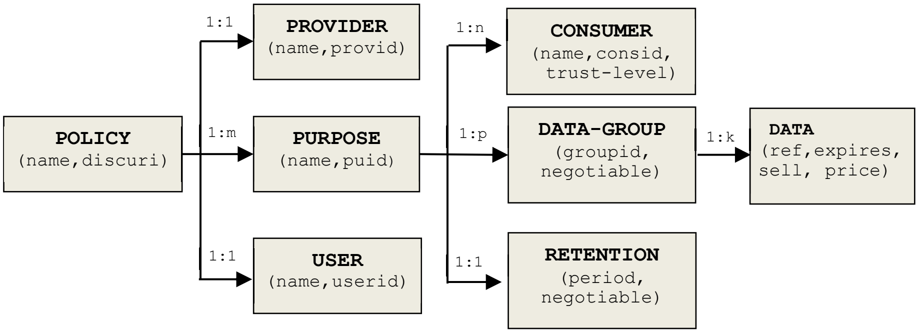


Fig. 1. Main elements of P2U Policy Specification Language

3.1 P2U Privacy Policy

The P2U privacy policy specification is inspired by the W3C's Platform for Privacy Preferences (P3P) [14] but does not follow its principles and syntax. While P3P is focused on limiting user data collection by websites, P2U is focused on enabling secondary information sharing and use. The P2U data sharing policy is based on a *purpose-relevance-sharing principle.* That is, only data relevant to the personalization purpose and context of use is shared. The main elements of P2U, their relationship and cardinality are illustrated in Figure 1. P2U defines eight policy specification elements each of which has some other attributes that further elaborate their usage. User privacy agreements should reflect the P2U privacy policy used by each application and they should be presented to the user allowing her to change the values of the attributes.

3.2 P2U Privacy Policy Negotiation

Policy Negotiation is done by the broker when a data consumer and data provider are ready to enter a contractual agreement for sharing user data. It may or may not require the participation of the user, depending on the degree of conflict in the request by the consumer and the parameters of data sharing set in the policy of the provider. The goal of the negotiation is to create a contract for data sharing that respects the P2U policies of the two applications and the user preferences. While it might be impractical to have negotiation over all the policy elements, some aspects of the privacy policy should be negotiable. For example, the provider, user and purpose for which data is shared may be non-negotiable. However, it should be possible for the data provider and consumer to negotiate the kinds of *data* in the data-group shared for that purpose, the *retention* period and the minimal *level of trust* that the consumer needs to maintain to be able to use the data [15,16]. After the negotiation is completed successfully, a contract between the data consumer and provider is made and the data will be shared according to the conditions of this contract. Compliance with the contract is enforced through a trust mechanism, which is part of the proposed user model sharing framework. Upon violation of any condition, e.g. the consumer uses data for a different purpose, or the retention period of data expires, the trust value of the consumer drops below the set threshold, the contract becomes invalid, and the user data can no longer be shared, under threat of legal action by the user or the data provider.

3.3 User Compensation for Data Sharing

The P2U policy language allows expressing user preferences for negotiating certain parts of their data to be shared in exchange of some compensation. This is indicated in the policy by respectively setting the attribute *negotiable* to TRUE for the data-group, the attributes *sell* and *price* for individual data items to TRUE, and a negotiable price value. The possibility of compensation for user data sharing was explored theoretically in [17] and has also been applied successfully in customer relationship management systems (loyalty reward programs) by credit cards, airlines and grocery stores. For example, the user can earn a certain percentage of the revenue made by the data provider from trading of his/her data, or earn points that are redeemable towards purchases or services.

4 Example Scenario

To illustrate how our data sharing framework might work in practice, we use a hypothetical scenario. Gena is a university graduate and a diabetic patient. She uses various web and mobile applications for her day-to-day activities such as a calendar, email, banking, shopping, photo-sharing, social networks apps. Recently, she installed some new apps on her smartphone: the FoodJournal app keeps a record of her daily food intake; the FitnessApp tracks her daily workout; and MedAdvice App provides her simple health advice; and a DietDataGatering Application by a researcher, who studies the correlation between food intake, exercise, and diseases such as diabetes.

While installing these apps, Gina recognizes that MedAdvice is asking her for the same information that she has already provided to other applications she uses, for example, the FitnessApp, the FoodJournal, the ShoppingApp and the Calendar. Gena does not want to re-enter the same information again and she knows generally, that sharing her data with the application will give her more personalized, and therefore, better services. So, she would like to grant access to the application to reuse the data she has already provided to the other applications. However, she is uncertain whether in this way she won't also grant access to other applications that she doesn't know of which may be harmful to her in some way. She wants to be able to control which applications have access to her data and to know for what purpose they use her data. She does not mind allowing third-party applications to use some of her data, if she is aware of the usage and can get some form of compensation, either monetary or in terms of improved services. However, she cannot trust the websites and applications she uses to protect her data from been generally released to other applications without her knowledge and consent.

Assuming that the P2U policy is adopted by a number of app service providers, Gena configures her data sharing preferences on all the websites and apps she uses. Through a uniform interface that indicates the possible purposes of use of user data, types of data, the duration of storing the data and negotiation preferences, she is able to control the usage and sharing of her data according to her privacy preferences. The information provided by her is converted into P2U policy files (in XML format), which are used by the data providers to control access requests to Gena's data by other applications (user data consumers) and to enter into negotiation on her behalf with them.

Table 1. FoodJournal Application User Model

Name	Zip code	Age	Nationality	Condition
Gena	S7N8C0	22	Canadian	Diabetic

Name	Date	Time	Food	Quantity	Hunger Scale (1-10)	Mood	Comment
Gena	13/01/2012	13:00	Chicken	100g	7	Motivated	None
Gena	13/01/2012	19:00	Pasta	200g	4	Tired	Sleepy
Gena	14/01/2012	07:00	Veggies	100g	8	Energetic	Woke up strong

```
Policy: FoodTrack
PURPOSE: To generate
health advice
RETENTION PERIOD:
unlimited
DATA TO SHARE (data,
sell, expires):
(food consumed,
false),
(quantity, false),
(hunger scale, false)
```

(a)

```
Policy: ShopRec
PURPOSE: To provide
shopping recommendations
RETENTION PERIOD:
unlimited
DATA TO SHARE (data,
sell, expires):
(food consumed, true,
180days),
(zip code, true, 180
days)
```

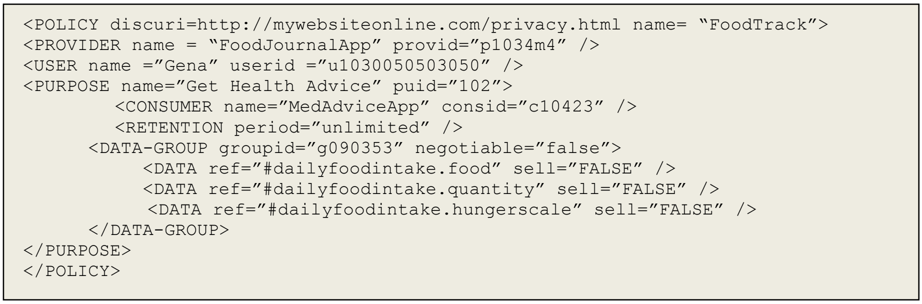
(b)

```
Policy: FoodRes
PURPOSE:   To   benefit
health research
RETENTION: 180 days
DATA  TO  SHARE  (data,
sell, expires):
(age, false),
(nationality, false),
(condition, false)
(food, false)
```

(c)

Fig. 2. Data sharing policies specified by Gena for three different purposes that can be used by (a) MedAdvice App, (b) Shopping App (c) DietDataGatering App

Assume the FoodJournal App user model contains two records, whose structure is depicted in Table 1 and Gena wants to allow the FoodJournal app to share the following information: with the MedAdviceApp → food consumed, quantity, and hunger scale; with the ShoppingApp → food consumed and zip code, with the DietDataGateringApp → age, nationality, condition and food consumed. Figure 2 shows three sample preference settings for sharing data that Gena has entered in the privacy settings of the FoodJJournal Application for three different purposes (the purposes are established by the developers of the FoodJournal App based on how they envisage user data collected by their application may be reused by other applications).



```
<POLICY discuri=http://mywebsiteonline.com/privacy.html name= ‶FoodTrack″>
<PROVIDER name = ‶FoodJournalApp″ provid=″p1034m4″ />
<USER name =″Gena″ userid =″u1030050503050″ />
<PURPOSE name=″Get Health Advice″ puid=″102″>
        <CONSUMER name=″MedAdviceApp″ consid=″c10423″ />
        <RETENTION period=″unlimited″ />
      <DATA-GROUP groupid=″g090353″ negotiable=″false″>
           <DATA ref=″#dailyfoodintake.food″ sell=″FALSE″ />
           <DATA ref=″#dailyfoodintake.quantity″ sell=″FALSE″ />
            <DATA ref=″#dailyfoodintake.hungerscale″ sell=″FALSE″ />
      </DATA-GROUP>
</PURPOSE>
</POLICY>
```

Fig. 3. Gena’s *FoodTrack* policy translated to *FoodTrack* Contract

Each data consumer application can only access the data specified by Gena in the preference settings of the respective purpose for which it request the data. Figure 3 shows the contract established between MedAdvice App and FoodJournal App after the negotiation phase based on the *FoodTrack* policy.

5 Conclusion

Sharing user data for purposes other than the one for which the data was collected poses a threat of violating user privacy through secondary use. This paper proposes a decentralized framework for user data sharing based on purpose of adaptation, which allows for flexible negotiation of various policy elements such as *type of data, retention period, trust level of consumer,* and *price*. We present Purpose-to-Use (P2U) policy specification language which allows the creation of different purposes and specification of relevant data to the purposes. The framework addresses the important issue of providing incentives for users to participate in the specification of their privacy policies and to allow sharing of their data.

References

[1] Heckmann, D., Schwartz, T., Brandherm, B., Kröner, A.: Decentralized User Modeling with UserML and GUMO. In: Dolog, P., Vassileva, J. (eds.) Proceedings of the Workshop on Decentralized, Agent Based and Social Approaches to User Modeling, DASUM 2005, at UM 2005, Edinburgh, Scotland, pp. 61–66 (July 2005)

[2] Berkovsky, S., Kuflik, T., Ricci, F.: Mediation of User Models for Enhanced Personalization in Recommender Systems. User Modeling and User-Adapted Interaction 18(3), 245–286 (2007)

[3] Carmagnola, F., Cena, F., Gena, C.: User Model Interoperability: a Survey. User Model User-Adapted Interaction 21(3), 285–331 (2011)

[4] Kobsa, A.: Generic User Modeling Systems. User Modeling and User-Adapted Interaction 11, 49–63 (2001)

[5] Fink, J., Kobsa, A.: A Review and Analysis of Commercial User Modeling Servers for Personalization on the World Wide Web. User Modeling and User-Adapted Interaction 10, 209–249 (2000)

[6] Dolog, P., Kay, J., Kummerfeld, B.: Personal Lifelong User Model Clouds. In: Proceeding of the Lifelong User Modeling Workshop at UMAP 2009, Trento, Italy, pp. 1–8 (June 2009)

[7] Vassileva, J., McCalla, G., Greer, J.: Multi-Agent Multi-User Modeling. User Modeling and User-Adapted Interaction 13(1), 179–210 (2003)

[8] Prince, R., Davis, H.: IDIUMS: sharing user models through application attributes. Poster presentation. in Proc. User Modeling, Adaptation and Personalization, UMAP 2011, Girona, Spain, pp 40–42. Springer, Heidelberg (2011) ISBN: 978-3-642-22362-4

[9] Kay, J., Kummerfeld, B., Lauder, P.: Personis: A server for user models. In: De Bra, P., Brusilovsky, P., Conejo, R. (eds.) AH 2002. LNCS, vol. 2347, pp. 203–212. Springer, Heidelberg (2002)

[10] Bielikova, M., Kuruc, J.: Sharing User Models for Adaptive Hypermedia Applications. In: Proc. 5th Int. Conf. Intelligent Systems Design and Applications, Washington DC, USA, pp. 506–513 (2005)

[11] Wang, Y., Kobsa, A.: Respecting users' individual privacy constraints in web personalization. In: Conati, C., McCoy, K., Paliouras, G. (eds.) UM 2007. LNCS (LNAI), vol. 4511, pp. 157–166. Springer, Heidelberg (2007)

[12] Heckmann, D., Schwartz, T., Brandherm, B., Schmitz, M., von Wilamowitz-Moellendorff, M.: GUMO – The General User Model Ontology. In: Ardissono, L., Brna, P., Mitrović, A. (eds.) UM 2005. LNCS (LNAI), vol. 3538, pp. 428–432. Springer, Heidelberg (2005)

[13] Abel, F., Henze, N., Herder, E., Krause, D.: Linkage, Aggregation, Alignment and Enrichment of Public User Profiles with Mypes. In: Proc. 6th Int. Conf. Semantic Systems (I-SEMANTICS), Graz, Austria, Article No. 11 (September 2010) ISBN: 978-1-4503-0014-8

[14] W3C P3P Specification, `http://www.w3.org/TR/P3P11/`

[15] Buffett, S., Jia, K., Liu, S., Spencer, B., Wang, F.: Negotiating exchanges of P3P-Labeled information for compensation. Computational Intelligence 20(4), 663–677 (2004)

[16] Preibusch, S.: Privacy Negotiations with P3P. In: W3C Workshop on Languages for Privacy Policy Negotiation and Semantics-Driven Enforcement, October 17-18 (2006), Ispra, Italy (2006), `http://www.w3.org/2006/07/privacy-ws/papers/24-preibusch-negotiation-p3p/` (last accessed: March 14th, 2013)

[17] Aperjis, C., Huberman, B.A.: A Market for Unbiased Private Data: Paying Individuals According to their Prvacy Attitudes (2012), `http://www.hpl.hp.com/research/scl/papers/datamarket/datamarket.pdf` (last accessed: January 10, 2013)