P2U: A Privacy Policy Specification Language for Secondary Data Sharing and Usage

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

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

***Abstract* — Within the last decade, there are growing economic/social incentives and opportunities for secondary use of data in many sectors, and strong market forces currently drive the active development of systems that aggregate user data gathered by many sources. This secondary use of data poses privacy threats due to unwanted use of data for the wrong purposes such as discriminating the user for employment, loan and insurance. Traditional privacy policy languages such as the Platform for Privacy Preferences (P3P) are inadequate since they were designed long before many of these technologies were invented and basically focus on enabling user-awareness and control during primary data collection (e.g. by a website). However, with the advent of Web 2.0 and Social Networking Sites, the landscape of privacy is shifting from limiting collection of data by websites to ensuring ethical use of the data after initial collection. To meet the current challenges of privacy protection in secondary context, we propose a privacy policy language, Purpose-to-Use (P2U), aimed at enforcing privacy while enabling secondary user information sharing across applications, devices, and services on the Web.**

***Keywords*—** *Privacy, Secondary Use, Policy Languages, Usage Control*

I. INTRODUCTION

Recently, there is an exponential growth in the amount of data about user available online. This data is shared voluntarily by user or passively collected by applications that analyze users' behaviors, activities and context. For example, many apps running on user's mobile devices (e.g. smartphones and tablets), store information about user's locations, preferences and interests. Wearable sensors, embedded in user's glasses, watches, shoes, and clothes gather data about the physical world environment of the user. With the advent and growing popularity of social networking sites and Web 2.0 applications, users are actively connecting, sharing, and commenting on these sites, thereby producing massive information about their interests, activities and social relationships. Furthermore, as cloud services become commonplace, users are storing, processing and accessing a lot of their personal information online. Although, the information collected about user by these applications and services have been traditionally kept in disparate application silos and databases, there are growing economic and social incentives for connecting and aggregating this data, thereby necessitating secondary sharing and use of the data across systems [1]. In addition, there are numerous benefits (in terms of personalized services) of such "secondary" sharing of user information to the user, the service providers, and society at large. For instance, reuse of existing user information by a new application helps the user avoid the duplication of same information across applications [14]. Moreover, allowing secondary sharing of user information leads to increased breadth and depth in the user model, which results in better personalization of services since more aspects or features of the user are available in an aggregated user model [15][16]. For businesses, the resulting better personalization will lead to better targeted advertisements on mobile devices, which is currently a big challenge for mobile application providers [18]. Finally, the emerging opportunity to aggregate user data from many sources (e.g. for Big Data Analytics) is driving innovation and societal benefits in many areas such as healthcare, national security, law enforcement, education, and home-automation [9]. Despite these benefits, sharing and utilizing user data for secondary purposes pose significant privacy risks to the user. For example, user data could be used for potentially harmful purposes such as surveillance or profiling the user for targeted discrimination with respect to employment, insurance and loans [17]. Hence, there is the need to balance the growing demand of secondary sharing for beneficial purposes with the need to protect user from harms.

One way to achieve this balance is through a privacy policy which allows data owners to set the permissions for a range of allowable usages of data [2]. Currently, there exists a number of privacy policy languages designed to enable both the user and data providers communicate their desired privacy protection. These include: Platform for Privacy Preferences (P3P) [3], A P3P Preference Exchange Language (APPEL)[4], eXtensible Access Control Markup Language (XACML) [6], and Geographic Location / Privacy (GeoPriv) [7], Rei [20], ExPDT [21], AIR[22], etc. Of these languages, P3P has been the most widely used structured privacy policy language on the Web [5].

In its current state, we believe the P3P policy language is inadequate in meeting the privacy challenges of user information sharing, particularly in a secondary context due to its focus, underlying privacy principles, inflexibility and lack of formal semantics. First, it is developed with the goal of giving the user a tool to limit information collection by a website in a primary data collection context. However, with the rise of social network and Web 2.0 technologies, users are now active, voluntary providers of massive amounts of information about their activities online. Therefore, the privacy challenge of today is shifting from limiting the collection of user data

DOI 10.1109/SPW.2014.12

towards preventing the unintended secondary sharing and usage of already collected data. Second, P3P assumes that the site collecting the data knows all the purposes of use of data a priori. Hence, it requires organizations to disclose the purpose of use of data at the point of collection and collected data can only be used for this purpose. This assumption fails to acknowledge the possibility of finding new and beneficial use of data for various secondary purposes that may not even be known when the data was collected [9]. Finally, P3P has remained a static language that does not allow negotiations of the elements of the privacy policy, which will be very crucial in emerging marketplaces for sharing and trading user data and when dealing with users with different privacy personality traits and preferences.

In view of the above, this paper proposes *purpose-to-use (P2U)*, a privacy policy specification language inspired by the P3P but adapted to user information sharing in secondary context. P2U is designed to enable emerging marketplaces for sharing and trading user data among applications, so that 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* (of use), *type of data*, *retention period* and *price* (of data).

II. EXISTING PRIVACY POLICY LANGUAGES

Privacy policy languages allow both the user and organizations to express their privacy controls and permissions. They provide a precise, machine-readable approach for specifying a privacy policy and empower applications to elicit user data according to the terms specified in the policy by the website. Within the last two decades, a number of policy languages have been proposed. Examples include: the Platform for Privacy Preferences (P3P) which was proposed by the W3C; A P3P Preference Exchange Language (APPEL) also proposed by W3C; eXtensible Access Control Markup Language (XACML), Extended Privacy Definition Tool (ExPDT), SecPAL4P, Rei, AIR and Geographic Location / Privacy[1]. A distinguishing characteristic of these languages is that they are designed and developed to address emerging privacy management issues in different situations and contexts [8]. For example, the P3P became a W3C recommendation in 2002 to address the growing collection of user data by websites (particularly, e-commerce sites that use the information to learn about user interests and provide personalized recommendations), while the GeoPriv was formulated by IETF in 2001 to manage privacy in the growing number of applications that require geo-location information about the user when providing context-aware services [7]. AIR [23] is focused on data accountability while Rei [20] is targeted at handling conflict resolution across policies in distributed systems. Of all the above languages, P3P has been the most popular and widely used structured privacy policy language on the Web [5].

The aim of P3P is to inform web users about the data-collection practices of Websites [3]. While P3P is targeted at allowing websites to express their data collection practices to the user, the companion W3C language, APPEL, allows the user to specify their privacy preferences to a website. Usually, a P3P user agent then compares the P3P file against APPEL file to discover any mismatch between the policy of a website and the preferences of the user and warn the user of potential privacy violation. Microsoft Internet Explorer 6 and Netscape Navigator 7 were early adopters incorporating P3P as plugins to their browsers.

Although, the P3P language is still deployed in Microsoft Internet Explorer, its adoption has remained stagnant for the past few years. Now it only focuses on cookie-blocking decisions [10]. Previous research and surveys have enumerated the factors responsible for the slow adoption of P3P and some of its weaknesses, including: lack of incentives for organization to adopt it; errors in P3P files on many websites; and lack of clear semantics. In addition, we believe that the P3P language, at present is limited in facilitating cross-system user information sharing and user privacy protection in a secondary data sharing context, which is becoming important for current web, cloud, and mobile applications for the following reasons [9]:

- *Changing role of the user:* P3P is premised on the role of user as data subject that need to be protected from organizations that collects information. With social networks and Web 2.0 technologies, the user's role is changing from being a passive data subject to that of producer of data. The assumption in these settings is that user wants to share. Hence, the emphasis of privacy policy languages also needs to change from limiting primary data collection to preventing unintended secondary use.
- *New purposes of data use after initial collection:* Since its underlying design principle is the traditional "notice and consent", P3P assumes that all the purposes of use of data is known during primary data collection and is expressed by the organization prior to collection [9]. The assumption that user data can only be used for pre-collection purposes is limiting, since it fails to envisage new and potentially very beneficial ways in which the data might be reused after collection to support personalization of services, offers, advertisement, etc.
- *Lack of support for negotiation:* P3P has adopted a static, rigid, and "take-it-or-leave-it" approach to privacy policy: "*an organization or service provider offers a privacy policy; the user has to accept it as a whole or leave it*" [11]. We believe that, as user data market evolves and users become active players in this emerging data market, rigid and inflexible policies will not suffice. Negotiation of some aspects of the policies will be required so as to provide flexibility and adaptability to various privacy personality types, purpose and context of use, as well as give incentives for users to allow sharing of their data across applications [12].

III. PROPOSED P2U POLICY LANGUAGE

As the need for collaboration and secondary sharing of user data across applications and data sources increases, we foresee the emergence of various marketplaces where user data can be

[1] This list is not exhaustive. Interested reader should see [8] and [19] for a full list of existing policy languages.

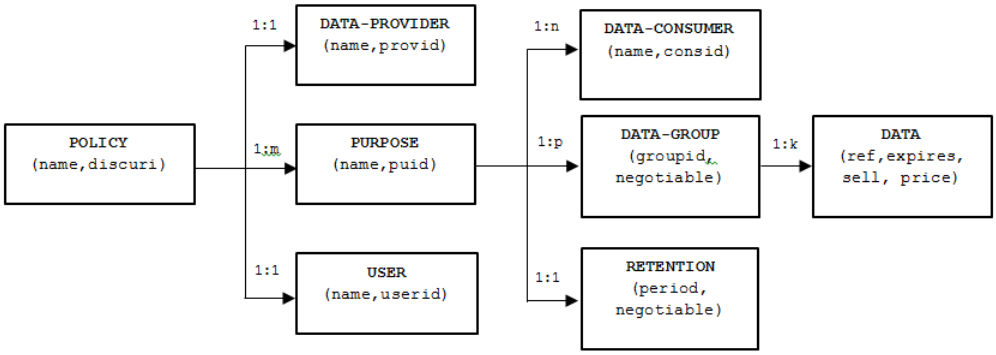

Figure 1. Main elements of P2U Policy Specification Language

shared and traded with the user's awareness and ability to control with whom the information is shared, for what purpose and for how long and for what compensation. We believe such a marketplace involves four active participants: the *user*, whose information is being shared, *data providers* (applications that have collected user data for a primary purpose and can re-share it with other applications for secondary use), *data consumers* (applications that need user data for secondary purposes). Lastly, a *data broker* provides middleware services to ensure semantic interoperability of data, coordination and negotiation with data consumers based on preferences. To capture user and data provider's expectations and facilitate interactions among the market players, we propose a policy language called *Purpose-to-Use (P2U)*. Unlike P3P that is based on the traditional *notice and consent* collection principle, P2U is based on what we referred to as the *purpose-relevance-sharing* principle. That is: *only data that is relevant to a particular purpose and context of use is shared.* Hence, the policy recognizes that data can be reused and shared after collection to fulfill various purposes that are beneficial to the user, business or society. The policy also supports negotiation by data consumers of some policy elements such as *type of data*, *retention period* and *price* (of data).

A. P2U Specification Elements

As shown in Figure 1, P2U defines eight privacy specification elements, each of which has some other attributes that further elaborate on their usage. We hereby describe, in high-level form, the main elements of P2U:

- **POLICY element**: This is the root element in P2U. The Policy element encapsulates every other element in the policy file. A P2U policy file has at least one policy element. The policy element contains one *provider*, one *user* and one or more *purpose(s)*. The policy is created by a provider for a user and with one or more purpose(s) of use. There are two attributes that can be specified in the policy element: (i) *Name* (mandatory) - P2U policy name (ii) *discurl* (optional) - location of human readable version of privacy policy.

- **DATA-PROVIDER element**: Gives information about the data service provider that issued this privacy policy. There are two attributes that can be in the provider element: (i) *Name* (optional) – i.e. name of the provider, (ii) *Provid* (mandatory) - a unique identifier for the provider

- **USER element**: The user element specifies the user for whom the privacy policy is about. The user element can have two attributes: i) *Name* - the username of the user on the provider's network, ii) *Userid* - unique identifier of the user.

- **PURPOSE element**: This specifies the data sharing purpose, with whom it was shared, for how long it can be retained, and the kinds of data that is relevant for that purpose. There can be one or more purpose elements in a P2U policy file. The purpose elements have two key attributes: i) *name* (mandatory): name indicating the purpose of sharing one or more data type, ii) *puid:* a unique identifier for this purpose. In addition, the purpose element is further subdivided into three sub-elements which are *consumer*, *retention*, and *data-group*.

  - **DATA-CONSUMER element**: The consumer element indicates the third-party application with whom this policy was created. However, the same data can be shared with more than one consumer by indicating that it is public. CONSUMER element has two attributes: (i) *Name* - specifies the name of the consumer. When the name "public" is used here, it means that the data is shared with any third party application; (ii) *Consid:* a unique identifier for the consumer.

  - **RETENTION element:** This specifies the time period (in days) for which data can be retained for the specified

```
<POLICY discuri=http://mydatawebsite.com/privacy.html name= "ShoppingPolicy">
<PROVIDER name = "FoodIntakeApp" provid="p6528m2" />
<USER name ="Jerry" userid ="u1030050503050" />
<PURPOSE name="Shopping Recommendations" puid="102">
        <CONSUMER name="MyShopApp" consid="c10023" />
        <RETENTION period="180" />
      <DATA-GROUP groupid="g090353" negotiable="TRUE">
           <DATA ref="#dailyfoodintake.food" sell="FALSE" />
           <DATA ref="#dailyfoodintake.quantity" sell="FALSE" />
            <DATA ref="#dailyfoodintake.hungerscale" sell="FALSE" />
      </DATA-GROUP>
</PURPOSE>
</POLICY>
```

Figure 2: Simplified example of P2U policy file in XML format.

purpose. The attribute *period* (in days) is mandatory. Also, an optional attribute *negotiable,* which is either TRUE or FALSE, indicates whether the retention period for the data is negotiable with the data consumer. If the value is not stated, the default value is FALSE.

- **DATA-GROUP element**: This element describes the group of data that can be shared for this purpose. There can be one or more variants of the data-group and each with different sharing constraints. Providing these variants allows the consumer and data providers to be able to negotiate the data options that best meet the needs of the consumer and not compromise the usage policy of the data provider. Each variant is identified by a unique *groupid* attribute for the data group. In addition, the data-group also contains a Boolean *negotiable* attribute which indicates whether the data in the data-group can be negotiated or not. Finally, the data group comprises one or more *data* elements.

    (i) **DATA element** – The DATA element describes the individual data within a data group. The data element has other attributes about the data that constraint how each individual data within the group can be used. These include: (i) *Ref (mandatory)* - a unique reference name for the data; (ii) *Expires (optional)*: specifies an expiry period for a particular data. It performs a similar function with the retention element. However, this attribute only affects individual data while retention affects the data-group. Where the *expires* attribute is specified, it overrides the timestamp indicated in the retention element for the particular data; (iii) *Sell* – this is a boolean attribute that indicate whether the user is willing to sell the data or not. The default value for the attribute is "FALSE". (iii) *Price:* if the attribute sell is TRUE, then the price attribute must be set by the user to indicate an initial price for the data which the parties can use during negotiation.

Figure 2 shows a simplified example P2U policy file in XML format for secondary user information sharing by an hypothetical mobile app, *FoodIntakeApp* (data provider) with another app, *MyShopApp* (data consumer) on behalf of a user named *"Jerry"* and for the purpose of the user receiving *"Shopping Recommendations"*. In natural language, the following permissions are allowed on user data: *"retain the data for 180days"*; *"You can access the following data about user for the purpose of Shopping Recommendation – food, quantity, and hungerscale"*. In addition, if the data consumer app requires more data than is specified in the "ShoppingPolicy", it can negotiate with the provider since the user sets the negotiable flag for data group to TRUE.

IV. CONCLUSION AND FUTURE WORK

This paper briefly reviews existing privacy languages and enumerated the limitations of the P3P specification in secondary information sharing. The paper then proposes a new policy language, called *purpose-to-use (P2U)*. Although, P2U was inspired by P3P, it is not based on its principles. P2U supports information sharing across applications based on the principle of purpose of use. To support emerging market for user data sharing and provide some incentive to user for sharing her data, the policy supports negotiation of some elements of the privacy policy to allow flexibility and adaptability to the need of users and diverse contexts of use.

As future work, we hope to formulate a more detailed use-case for the policy and design a negotiation protocol for the interaction. We will also implement a prototype P2U policy to support secondary user data sharing among mobile applications as proof of concept. There are also many issues to be addressed in the future. One is the semantics of the language, especially for allowable purposes. We are working on defining a set of constraints [23] or rules [24] on data retention period and data consumer that could access the data based on sensitivity of data group and different purpose of use. Another issue that is not addressed yet is the influence of *context* of use the policy. For example, releasing user data for law enforcement or emergency purposes when explicit user preference is not available or desirable and the system has to decide based on the situation what is best for the user or

society. Also, since we aim to enable user awareness and control in setting preferences and permissions for secondary sharing of their data, it is also important that we ensure users understand the policy, are not overwhelmed with too much information and are able to set their preferences [13]. The user will be asked to review and set privacy preferences if the policy expires, if a new potential purpose for sharing emerges, and if the user has set a flag to be notified for negotiation with user data consumers. Another issue is how to enforce compliance and ensure that data consumers use data in accordance with the contractual agreement for sharing the data. The problem is similar to that of enforcing copyright agreements (digital rights management). Multi-dimensional solutions to this challenge are in the works [9], involving technical, legal, economic and social measures. We believe one solution for enforcing compliance by data consumers with the privacy policies is through the use of trust and reputation mechanism. Trust and reputation (TR) mechanisms present a compelling approach to detect violators based on feedback from others. TR mechanisms have been successfully applied in managing interactions and mitigating misbehaviors in different areas such as e-commerce, peer-to-peer networks, and mobile ad-hoc networks [25]. For example, we can allow an independent monitoring and enforcement agent within the market to track complaints on misuse of data by a data consumer. This agent computes trustworthiness of data consumers based on violation their data use contract. In so doing, the framework allows the data-market community to police data usage itself and report potential violations to the framework management. The trust value will then form an important factor to consider when determining what data to share with a data consumer and for the price the data consumer will have to pay to acquire the data. If the consumer application's trust level drops below a defined threshold, the consumer will be unable to participate in the user data market.

REFERENCES

[1] Tene O and Polonetsky J. (2012). Privacy in the Age of Big Data: A Time for Big Decisions. Standford Law Review. Online at:http://www.stanfordlawreview.org/sites/default/files/online/topics/64-SLRO-63_1.pdf

[2] Spiekermann, S. and Cranor, L. F., (2009). Engineering Privacy. IEEE Transactions on Software Engineering, Vol. 35, Issue No.1, pp.67-82

[3] W3C P3P Specification. Online at: http://www.w3.org/TR/P3P11/

[4] Cranor, L., Langheinrich, M., and Marchiori, M. A P3P Preference Exchange Language 1.0 (APPEL 1.0). Tech. report, World Wide Web Consortium, Retrieved December 12, 2012. Online: http://www.w3.org/TR/P3P-preferences/.

[5] Dong, L., Mu, Y., Susilo, W.; Wang P. and Yan, I (2011), "A Privacy Policy Framework for service Aggregation with P3P," in Proceedings of 6th International Conference on Internet and Web Applications Services, Thinkmind, St. Maarten, pp. 171-177

[6] Moses, T. eXtensible Access Control Markup Language (XACML) Version 2.0. Tech. rep., Oasis, Retrieved June 17, 2012, http://xml.coverpages.org/XACMLv20CDCoreSpec.pdf, 2004.

[7] Schulzrinne, H. A Document Format for Expressing Privacy Preferences. Tech. rep., The Internet Engineering Task Force, Retrieved June 12, 2005, http://www.ietf.org/internetdrafts/draft-ietf-geopriv-common-policy-04.txt.

[8] Kumaraguru, P., Cranor, L., Lobo, J., and Calo, S. A survey of privacy policy languages. Workshop on Usable IT Security Management (USM 07). In SOUPS '07: Proceedings of the 3rd symposium on Usable privacy and security (New York, NY, USA, March 2007), Online at: http://cups.cs.cmu.edu/soups/2007/workshop/Privacy_Policy_Languages.pdf

[9] World Economic Forum Report (October 2012). Unlocking the Economic Value of Personal Data: Balancing Growth and Protection. Online at: http://www3.weforum.org/docs/WEF_IT_UnlockingValueData_BalancingGrowthProtection_SessionSummary.pdf

[10] Goldman E. (2011). The Economics of Privacy. Online at: http://blog.ericgoldman.org/archives/2011/12/economics_of_pr.htm

[11] Preibusch S. (2006). Privacy Negotiations with P3P. Online at: http://www.w3.org/2006/07/privacy-ws/papers/24-preibusch-negotiation-p3p/

[12] Iyilade J. and Vassileva J. (2013). A Framework for Privacy-Aware User Data Trading. Proceeding of UMAP 2013 conference, June 10-14, Rome, Italy. pp 310-317.

[13] Carroll, J and Rosson, MB, 1987, "The paradox of the active user" in JM Carroll (ed.) Interfacing Thought: Cognitive Aspects of Human-Computer Interaction MIT Press.

[14] Heckmann, D. (2005). Ubiquitous User Modeling, Ph.D. thesis, Computer Science Department, Saarland University, Germany

[15] Heckmann, D., Schwartz, T., Brandherm, B., Kröner, A. (2005): 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-05, at UM2005, July, Edinburgh, Scotland, pp. 61–66.

[16] Berkovsky S. (2006). Decentralized mediation of user models for a better personalization. Adaptive Hypermedia and Adaptive Web-Based Systems. Springer Berlin Heidelberg. Pp. 404-408.

[17] Toch E., Wang Y., Cranor L.F. (2012) Personalization and Privacy: A Survey of Privacy Risks and Remedies in Personalization-based Systems. User Model User-Adapted Interactions. 22:203–220.

[18] Dredge, F. (2013): Facebook IPO filing reveals mobile risks and opportunities. Available online at: http://www.guardian.co.uk/technology/2012/feb/02/facebook-ipo-mobile-risks. Last Accessed: July 15, 2013.

[19] W3C (2009). Policy Language Review Wiki. Accessed Online at: http://www.w3.org/Policy/pling/wiki/PolicyLangReview. Last accessed date: 22-March-2014

[20] UMBC equity research (2005). Rei : A Policy Specification Language. Online at: http://rei.umbc.edu/

[21] Kahmer M and Gilliot M. Extended Privacy Definition Tool. Online at: http://ceur-ws.org/Vol-328/paper12.pdf. Last accessed date: 16-March-2014

[22] AIR Policy Language. Online at: http://dig.csail.mit.edu/TAMI/2007/AIR/.

[23] [24] Li N., Yu T., Anton A. (2003). A Semantics-Based Approach to Privacy Languages. Online at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.131.4832&rep=rep1&type=pdf. Last accessed date: 17-March, 2014

[24] Denker G. and Martin D. (2004). Using Rules to define the semantics of privacy policies. Online at: http://www.w3.org/2004/12/rules-ws/paper/76/

[25] Jsang, A., Ismail, R., and Boyd, C. (2007). A survey of trust and reputation systems for online service provision. Journal of Decision. Support System, 43:618-644.