COMPUTERS & SECURITY 67 (2017) 122–141

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

Privacy policies verification in composite services using OWL

Assadarat Khurat [a,b,*], *Boontawee Suntisrivaraporn* [c,d], *Dieter Gollmann* [a]

[a] *SVA, Hamburg University of Technology, Hamburg, Germany*
[b] *ICT, Mahidol University, Bangkok, Thailand*
[c] *SIIT, Thammasart University, Bangkok, Thailand*
[d] *Customer Insights Department, Marketing Group, dtac, Thailand*

ARTICLE INFO

Article history:
Received 21 August 2016
Received in revised form 14 December 2016
Accepted 23 February 2017
Available online 2 March 2017

Keywords:
P3P policy
Privacy policy
OWL ontology
Composite service
Reasoning

ABSTRACT

Privacy has been an important issue for online services collecting customer data. P3P is a privacy policy language with a fixed vocabulary to express privacy practices of online services. The matching between the privacy practices (P3P policies) and users' privacy preferences facilitates the users to be aware of services' usage of their data. However, the change from single to composite online services raises more privacy concern due to the increasing amount of user data being collected, stored and shared. This change impacts on P3P since it was designed from a single service perspective. In addition, P3P allows the specification of policies containing semantic inconsistencies. In this paper, we extend P3P to be suitable for composite services and propose a formal semantics for P3P using OWL to facilitate reasoning about semantic ambiguities in P3P policies. The constraints defined in our ontology are used to verify potential semantic inconsistencies and to check for conflicts occurring from P3P policies of service members. We have implemented a P3P verification tool and verified five hundred P3P policies collected from actual websites. The verification result shows that more than half of these P3P policies contain conflicts.

1. Introduction

In the beginning of the online services era, most services were single and independent, employing and developing proprietary technology to serve their customers. Hence, the main communication messages for service provisioning of these online services were between users and service providers. Nowadays, there is a strong competition in the online market to increase and expand the number of customers. This is an incentive for developing new and better services, which should be fast enough and better serve user demands. Service providers, therefore, attempt to combine their services with other existing services to obtain a new service, a so-called composite service. For example, a weather forecast service may combine with a text-to-voice service to become a new service that can provide weather forecasts in voice. Therefore, the main communication messages for service provisioning now also include the communication between online services.

In order to provide a service, most of the online services collect and store personal data of their users such as name, addresses, photos, telephone numbers, etc. For instance, online shopping services require name and address of their customers for product delivery. A user normally registers with more than one online service and is likely to increase the usage which reveals more of the user's data. These collected data include

[*] *Corresponding author.*
E-mail address: assadarat.khu@mahidol.ac.th (A. Khurat).

http://dx.doi.org/10.1016/j.cose.2017.02.015

personal data. Hence it is essential to consider what services will do with the data to provide privacy protection. This has been a concern for some time as seen from the enactment of privacy legislation.

In order to comply with legal requirements, a technical implementation of privacy protection is by enforcing appropriate policies. In general, a privacy policy is used to express what services will do with the collected data. Among existing privacy policies e.g. S4P (SecPAL for Privacy) (Becker et al., 2010) and PPL (PrimeLife Policy Language) (Ardagna et al., 2009), P3P (Platform for privacy preferences) is the only one being used in real applications. It describes the data practices of websites in a machine-readable format which enables users to decide whether to use the service or not.

The emergence of the composite service paradigm though providing better and more convenient ways for service provisioning introduces some problems at the same time to P3P. This is because P3P was designed from a single service perspective. For instance, some terms of its predefined vocabulary used for describing policies, e.g. *Recipient*, can represent recipient types from the perspective of a single entity, but not in a composite service assembled on the fly. We thus extend the vocabulary so that P3P can be employed in composite service scenarios. In addition, the flexible syntax of the P3P policy language and some combinations of terms in the vocabulary may cause ambiguous and conflicting meanings as observed in (Cranor, 2003; Karjoth et al., 2003; Li et al., 2003; Schunter et al., 2002; Yu et al., 2004).

In a nutshell, P3P has a taxonomy (vocabulary with structure) but not yet an ontology (taxonomy with relationships, constraints and rules). Therefore, we propose to use a semantic web ontology, i.e. OWL (Web Ontology Language), as our solution (Khurat and Suntisrivaraporn, 2011) by interpreting P3P policies with a data–purpose centric view (Khurat et al., 2010) to mitigate ambiguities and to check for conflicts. Moreover, when combining several services together, their privacy policies may not be compatible. We also employ the OWL ontology to check for incompatibilities.

The organization of this work is as follows. The background of the P3P Policy Language is shown in Section 2. We analyze P3P with regards to its semantics and when employed in a composite service environment in Sections 3.1 and 3.2 respectively. The extension of P3P is proposed in Section 4. How we interpret P3P and what constraints we define for policy verification are presented in Section 5. We also describe the background of Web Ontology Language (OWL) in Section 6. Section 7 illustrates our proposed P3P Ontology which is implemented and tested as shown in Section 8. Work related to this research is discussed in Section 9. Finally, we conclude this work in Section 10.

2. P3P policy language

The Platform for Privacy Preferences (P3P) (Cranor et al., 2002, 2006; P3PToolbox Introduction, 2013) is an XML-based policy language specified by W3C to enable websites to describe their privacy practices in a standard format. It was specified based on principles laid down in the OECD Privacy Protection Guidelines (OECD, 1980) and the EU Directive 95/46/EC. User consent and purpose of collection are two principles at the core of this approach to privacy protection. The main content of the P3P policy language describes what websites may do with the data collected from users such as the purpose of data usage and how long they will keep the data. In practice, users can retrieve the P3P policies of a website via the HTTP protocol. A P3P user agent embedded at the client side compares the obtained policy files with the user preferences and notifies the user about the comparison result. This may help the user deciding whether to use the website or not.

The P3P specification defines the P3P policy language (syntax and predefined vocabulary), mechanisms for associating policies with web resources, e.g. by HTTP, and a standard set of data elements in a hierarchy which all P3P user agents must understand. P3P was designed for the single service paradigm. The first version of the P3P specification is 1.0 (W3C Recommendation) (Cranor et al., 2002) and the current version is P3P 1.1 (Working Group Note) (Cranor et al., 2006).

We describe the syntax of the P3P policy language and the predefined vocabulary in the following subsections.

2.1. P3P syntax

The XML schema for P3P policy documents is shown in Listing 1. The main elements of a P3P policy are the *Entity*, *Access*, *Disputes-Group*, and *Statement* elements. The *Entity* element identifies a legal entity, i.e. the service or website issuing this policy. The *Access* element indicates the ability of users to access their data. The *Disputes-Group* describes the resolution procedure when disputes about these privacy practices occur.

Listing 2. P3P Policy of Walmart.com with Proposed Extensions Elements

```
Pol{Entity(#business.name):walmart.com,...,
  S1{Pur:(current,contact[opt-in]),
     Rec:(ours,ours-list:(urn:service:id:Walmart)),
     Ret:(indefinitely),
     Dat:(#user.login,#user.home-info)}
  S2{Pur:(current,develop[opt-in],contact[opt-in]),
     Rec:(ours,ours-list:(urn:service:id:Walmart)),
     Ret:(stated-purpose,current-duration:(PT10M),
          develop-duration:(P3DT10H30M),
          contact-duration:(P2Y4M)),
     Dat:(#user.name,#user.login,#user.home-info)}
  S3{Pur:(current,develop,admin),
     Rec:(ours,delivery,ours-list:(urn:service:id:Walmart),
          delivery-list:(DeliveryConcepts)),
     Ret:(stated-purpose,current-duration:(PT10M),
          develop-duration:(P3DT10H30M),
          admin-duration:(P6M)),
     Dat:(#user.name,#user.home-info,#thirdparty.name,
          #thirdparty.home-info,#dynamic.interactionrecord,
          #dynamic.cookies
             {Cat:(uniqueId,computer,navigation,state,
                   purchase)})}
  S4{Pur:(current,develop,admin,contact[opt-in]),
     Rec:(ours,delivery,ours-list:(urn:service:id:Walmart),
          delivery-list:(DeliveryConcepts)),
     Ret:(stated-purpose,current-duration:(PT10M),
          develop-duration:(P3DT10H30M),
          admin-duration:(P6M),contact-duration:(P2Y4M)),
     Dat:(#user.name,#user.home-info,#user.login,
          #thirdparty.name,#thirdparty.home-info,
          #dynamic.interactionrecord,#dynamic.http,
          #dynamic.searchtext,#dynamic.clickstream,
          #dynamic.cookies
             {Cat:(uniqueId,computer,state,navigation,
                   purchase)})}
  S5{Pur:(current,develop,admin),
     Rec:(ours,ours-list:(urn:service:id:Walmart)),
     Ret:(indefinitely),
     Dat:(#dynamic.interactionrecord,#dynamic.http,
          #dynamic.clickstream,#dynamic.searchtext,
          #dynamic.cookies
             {Cat:(uniqueId,computer)})}
  S6{Pur:(current),
     Rec:(prospective-composite-service),
     Ret:(stated-purpose,current-duration:(PT10M)),
     Dat:(#user.home-info)}
 }
```