An Ontology Based Privacy Protection Model for Third-Party Platform

Haojun Yu[1], Yuqing Sun[1,*], and Jinyan Hu[2]

[1] School of Computer Science and Technology, Shandong University, Jinan250101, China
fengzhengjie2005@163.com,sun_yuqing@sdu.edu.cn
[2] School of Economics, Shandong University, Jinan 250100, China
hwx@sdu.edu.cn

Abstract. A third-party trading platform is a web based system that provides services for sellers and buyers. On such platform, users are required to provide personal information to ensure the authenticity and undeniability of a transaction. In this paper, we propose an ontology based privacy protection model for third-party platform, which allows buyers and sellers to define privacy policies according to their preferences and converts policies into ontology based forms. We introduce the property of *good* sellers who require the minimal information from buyers while satisfying other trading requirements. The proposed policy matching algorithm finds such sellers as candidates for a buyer request. A practical example is given to illustrate our model.

Keywords: third-party platform, privacy policy, ontology.

1 Introduction

A third-party trading platform is a web based system that provides business related services for sellers and buyers. According to a complete business process, services on such platform can be divided into three categories: pre-business services such as information publication and contract negotiation, in-process services such as trade confirmation and payment, post-business services such as logistics, etc. During this process, both buyers and sellers are required to release their personal information so as to ensure the authenticity and undeniability of a transaction, such as name, address, and intentions, etc. Such information includes identity information and business information, which involves buyer's privacy and may face various attacks [12]. Therefore, it is important to protect personal information against misuse.

Privacy protection on third-party platforms has attracted much attention from academia, government and business. Many professionals have devoted to this topic and a lot progresses have been made. From the view of law, the Online Data Privacy Regulation by the European Commission[6]makes many restrictions on web service providers. For example, a provider is only allowed to collect and use buyers' personal

* Corresponding author.

R. Huang et al. (Eds.): AMT 2012, LNCS 7669, pp. 369–376, 2012.

information with their permissions. The regulation also considers buyers' preference as privacy policies, which state the ways a party gathers, uses, and releases personal information it collects. Privacy policies [9] should inform the buyer what specific information is collected, and whether it is kept confidential, shared with partners, or sold to other firms. These regulations provide a legal basis for privacy protection.

In this paper, we propose an ontology based privacy protection model for third-party trading platform according to the current Data Privacy Regulations. In our model, both buyers and sellers are allowed to express their privacy preferences as policies and then the policies can be compared based on domain ontologies. We introduce the property of *good* sellers who require the minimal information from buyers while satisfying other trading requirements. And then we propose a policy matching algorithm to automatically select such sellers. We implement our model using ontology editor Protégé and schema matching tool COMA++, and illustrate our model with a practical example. The rest of the paper is organized as follows. In Section 2, we present the ontology based privacy protection model. In Section 3, we propose the policy parsing and matching method. Then we discuss how the model could be implemented in Section4. In Section 5, we conclude the paper.

2 Related Work

Considering web based privacy protection, many models are proposed. Tunner et al. propose a privacy protection model for web based services[7].It introduces the concept of agents on behalf of buyers to negotiate with sellers on how much personal information should be released. In their model, sellers need to state why they collect personal information and should provide multiple elective information choices for buyers to choose. Bonatti et al.[3]show that the competition between sellers can reduce their information requests through a game-theoretic approach. They present a privacy protection mechanism based on Vickrey auction. Because the auction mechanism is truthful, the proposed mechanism induces sellers to ask for the exactly necessary information from buyers to deliver their service effectively and securely. However, these methods do not provide a way for buyers to define their privacy policies according to their preferences.

Some research works study how to implement the privacy policy interaction between users and systems. Garcia et al. [6] use semantic web technology to present an ontology-based privacy policy definition model. Buyers define some allowed constraints on the operations of their personal information. Hacker et al. [5] propose privacy ontology to enable buyers to understand the content on web so that buyers can match their preferences to the seller's policy and decide whether to use the given service. Gaoet al. [8] propose an ontology based approach for privacy protection in different application scenarios. They specify privacy policies in a semantic way and abstract the policies as trust attributes for privacy ontology, which would be used in evaluating trust in different applications.Hu et al. [13] propose a semantic legal policy definition model, which is in compliance with the laws. Then the policies are enforced automatically at the super-peer to enable Law-as-a-Service. The above methods only consider the case of exact policy match, but ignore the case that sellers and buyers may express the same object in different words.

From the above analysis, we find that quite a few shortcomings exist in current privacy protection and should be improved. First is the privacy preferences definition. Since people have various considerations on their personal information [4] and the required information in different business are not the same, it is necessary to allow buyers to define privacy policy according to their own preferences. Second is the automatic matching requirement. For the same information, sellers and buyers may express their understanding in different ways, such as home address and family address. So, there should be a semantic-based method of policy matching so as to get rid of ambiguity. Third point is the minimization principle. Consider the case that there are multiple candidate sellers with different requirements on buyers' personal information, the third-party platform should choose the *good* sellers who require minimal information while satisfying buyers' preferences.

3 The Ontology Based Privacy Protection Model

3.1 Basic Concepts

According to the latest Online Data Privacy Regulation [6], privacy policies on a third-party platform should include three aspects: the personal privacy items, such as name, contact information, the purpose of sellers' using information, such as credit card for purchase, and the process mode of personal information, such as the ID numbers must be deleted after transactions. We introduce the following basic notions.

Definition 1 (Item). An item is an attribute of a buyer and its associated sensitivity. Let PRI_ITM be a set of all attributes in a given system. An item is denoted as a pair of t=<a, s>, where a∈PRI_ITM, s∈[0..1], denotes an attribute and its sensitivity.

The attribute sensitivity shows how an attribute is important for a buyer. For a given item *t*=<*a*, *s*>, we use the notion *t.a* and *t.s* to denote an attribute *a* and its sensitivity *s*. The higher sensitivity is, the more important the attribute is for the buyer.

Privacy policy for third party platform involves three aspects. Privacy information is denoted as a finite set of items that a buyer releases in a transaction. Purpose is a finite set of ways that how sellers use privacy information. Process mode is denoted as a set of ways how sellers process privacy information after a transaction. Let INFO, PURPOSE, and PROCESS respectively denote the set of all privacy information, purposes of information, and process modes of information in a given system.

Definition 2 (Privacy Rule, PR for short). A privacy rule specifies the purpose and process mode of a set of privacy items. It is defined as a triple PR= (PIn, Pur, Pro), where PIn⊑INFO, is a set of privacy items, Pur⊑PURPOSE is a set of purposes, and Pro⊑PROCESS is a set of process modes.

For example, the privacy rule *PR*= ({<*name*, *0.4*>, <*telephone*, *0.6*>}, {*notification*}, {*retain*}) declares that the name and telephone are allowed to use in a transaction for the purpose of notification. The sensitivities of these attributes are 0.4 and 0.6, respectively. After the transaction, the seller should retain this information.

Definition 3 (Privacy Policy, PP for short). A privacy policy is a set of rules specified by a buyer or a seller before a transaction, which specifies how their information could be used in the periods of this transaction (e.g. registration, payment, logistics). It is denoted as PP= $\{PR_1, PR_2 \ldots PR_k\}$, where k is an integer, PR_i, $i \in [1..k]$, $PR_i \in PR$, is a privacy rule. Let PSET be the set of all privacy policies used in a system.

3.2 The Ontology Based Privacy Protection Model

In this section, we describe the proposed ontology based privacy protection model in details. Our model (see Figure 1) contains three aspects.1) *Policy Definition* allows buyers and sellers to define privacy policies according to their preferences, which are based on the proposed definition.2) *Semantic Parse* converts policies into ontology-basedforms.3) *Policy Match* enforces an automatic policy matching to find the good service for buyers according to the policy matching algorithm.

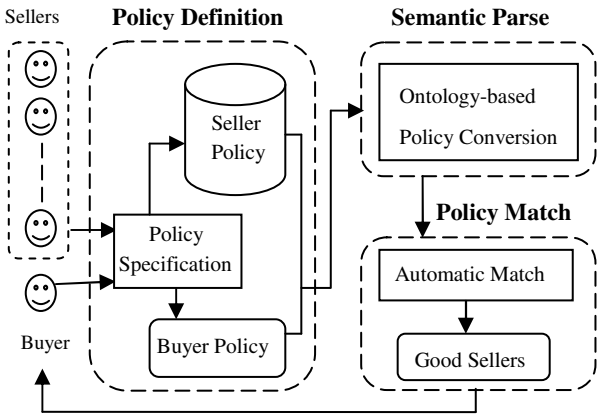

Fig. 1. An Ontology Based Privacy Protection Model

The details of a privacy policy matching process are given below. First, before transactions, sellers publish their services on a third-party trading platform with the specifications of privacy policies, namely the required information for buyers. These policies are stored in the seller policy repository. When a buyer invokes a transaction, he/she defines the privacy policies according to his/her preferences. Then, the trading system on the third-party platform semantically parses sellers' policies and the buyer's policy by converting them into ontology forms. Finally, according to policy matching, the platform finds the *good* seller candidate policy for the buyer and returns policy and the related services to the buyer. Policy matching is a group of matches between a buyer's policy and a set of sellers' policies, and is used to find the *good* seller policy. We present the policy matching problem as follows.

Definition 4 (Policy Matching Problem, PMP for short). Given a buyer privacy policy *pol*, a set of sellers privacy policies $p_{set} = \{pol_1, pol_2 \ldots pol_k\}$, where k is an in-

teger, pol_i∈PSET, i∈[1..k].The policy matching problem is to find the policy in p_{set} satisfying one of the following constraints:

1) *exact match*: pol_i exactly matches *pol* if and only if the following formula holds

pol. PIn⊇pol_i*. PIn* ∧ *pol. Pur*=pol_i*. Pur* ∧ *pol.Pro*=pol_i*. Pro*∧minSizeof(pol_i*.PIn*) ∧ minimal(pol_i*.s*),where$pol_i.s$=$\sum_{t_j \in pol_i.PIn} t_j.s$, t_j*.s* is the sensitivity defines by the user, pol_i*.s* is the sum of all sensitivities in the policy pol_i.

2) *fuzzy match*: pol_i fuzzy matches *pol* if and only if the following formula holds

!(*pol.PIn* ⊇ pol_i*.PIn* ∧ *pol.Pur*=pol_i*.Pur* ∧ *pol.Pro*=pol_i*.Pro*) ∧ *maximum*(pol_i*.d*), where

$pol_i.d$=$\sum_{t_j \in pol_i.PIn, t_k \in pol.PIn} similarity(t_j.a, t_k.a)$, is the *fuzzy match degree* of a seller policy pol_i.

Exact match refers to the case that a seller's required attributes, the purpose and process mode of this information satisfies the buyer's privacy preference. For policies satisfying *exact match*, we find the *good* seller policy with the minimal information requirement, i.e. minimal size of pol_i*. PIn*, and the minimal sum of sensitivities. If there is not any seller's policy *exact match* with the buyer's policy, we calculate the *fuzzy match degree* for all seller policies. To find which seller policy satisfies the buyer's preference as much as possible, we give top priority to the attributes with higher similarities. So we calculate the sum of similarities and find the *good* seller policy with the maximum *fuzzy match degree*. So, the above constraints are suitable for different cases for the buyer to choose the *good* seller.

4 Ontology-Based Policy Parsing and Matching

In this section, we parse privacy policies to eliminate semantic ambiguity. Ontology represents [11] knowledge as a set of concepts and the relationships between those concepts. OWL is the web ontology language[1], which provides abundant semantic expression and supports reasoning between concepts. By means of converting policies into OWL, the system can realize ontology-based policy expression and policy matching, and then it can find the good policy.

We propose a policy matching algorithm *Policy Match* to solve the PMP problem. We introduce a threshold α to judge if the corresponding elements in two policies refer to the same object. If the *similarity* between two elements is greater than α, it can be considered that they express the same object. The value of α is set by buyers. In this paper, we suppose the value of α be 0.8 by experience. According to the definition of *exact match*, if the *similarity* for each pair of element in two policies is greater thanα, policy *pol* and pol_i are called *exact match*. Otherwise, we calculate the *fuzzy match degree*. We introduce a function *match()* to judge if a seller policy pol_i satisfies the buyer policy *pol*.

$$match(pol_i)=\begin{cases}\text{true} & \text{if } similarity(pol_i.element, pol.element) > \alpha, \forall pol_i.element \\ \sum_{t_j \in pol_i.PIn, t_k \in pol.PIn} similarity(t_j.a, t_k.a) & \text{otherwise}\end{cases}$$

If $match(pol_i)$=true, it means the policy satisfies the buyer policy *pol*, then put the seller policy pol_i in a set *p'*. After getting the values of *match()* of all policies in the set p_{set}, we use the algorithm *Policy Match* to acquire the *good* seller policy. If the set *p'* is not empty, we find the *good* policy with the minimal information requirement and the minimal sum of sensitivities. If *p'* is null, we choose the *good* policy with the maximum *fuzzy match degree*. The algorithm *Policy Match* is given below:

Algorithm *PolicyMatch*(*pol*, p_{set}, *match*)

```
For each pol_i in p_set
  If (match(pol_i)==true)
    Put pol_i in p';
If (p'!=Φ)
  Sort p' by size of PIn in descending order;
  Choose the minimal ones, and put in p_0;
  For each pol_i in p_0
pol_i.s=∑_{t_j∈pol_i.PIn} t_j.s;
  Sort p_0 by pol_i.s in descending order;
  Choose the minimal one, and put in p_1;
  Return p_1;
Else
  Sort p_set by match(pol_i);
  Choose the maximum one, and put in p_2;
  Return p_2;
```

We would illustrate our method with a practical example in next section.

5 System Implementation and an Illustrative Example

We implement our model in an environment with Intel 2 cores, 2.4GHz CPU, 2GB of memory and 320GB disk. We choose the ontology editor Protégé to convert privacy policies into ontology based expressions, and choose COMA++ as the schema matching tool for policy match. The selected data set PRI_ITM= {register name, gender, age, name, e-mail, company phone, home address, ID number}, the set PURPOSE= {inform, delivery, verify, refund}, and the set PROCESS= {disclosure, retain, delete} to present all available data in the system.

In our model, we use OWL to define each *element* in a policy as a class and define *sensitivities* of attributes as data property [10]. Then according to operating on data properties, we get the sum of sensitivities to realize the algorithm *Policy Match*. COMA ++ [2] is a schema and ontology matching tool. Figure 2 describes how COMA++ works. The left figure shows the *similarity* of the corresponding nodes. Experiences show that if the *similarity* is larger than 0.8, this pair of nodes can be considered as the same, i.e. the two nodes express the same object. The right part of the figure 2 shows the whole policy match in COMA++.

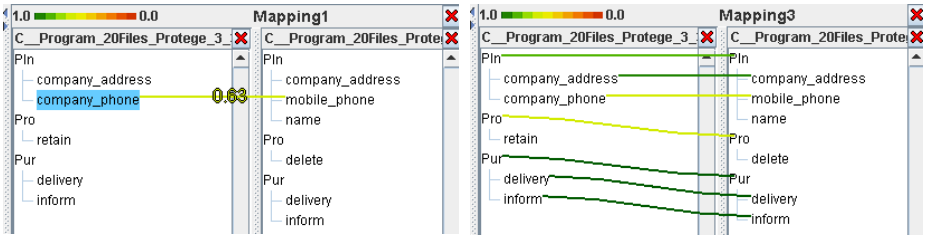

Fig. 2. the Policy Match in COMA++

Example. Suppose that a buyer privacy policy is pol = ({<companyphone,0.3>, <companyaddress,0.3>,<mobilephone,0.4>}, {inform, delivery}, {delete}), and the policies of candidate sellers are pset={pol_x, pol_y, pol_z, pol_w},where pol_x=({<company adress,1>,<companyphone,1>},{delivery,inform},{retain}),pol_y=({<officeaddress,1>, <mobilephone,1>}, {delivery,inform},{delete}), pol_z=({<companyaddress,1>, <officephone,1>},{delivery,inform},{delete}), pol_w=({<officeaddress,1>, <companyphone,1>,<mobliephone,1>},{delivery, inform},{delete}).

After policy parsing and matching, the algorithm returns the policy p_1=pol_z and its related service information to the buyer. The analysis of the results is as follows. In the phase of exact match, due to similarity (<pol.delete, pol_x.retain>) =0.65<α, the policy polx doesn't satisfy the buyer's preference, so it is removed. After exact match, we can get the policy set p'= {pol_y, pol_z, pol_w}. Next, based on algorithm Policy Match, due to the size of pol_w. PIn=3>2(the number of the least information need is 2) and pol_y.s=0.7>0.6(the lowest sum of sensitivities is 0.6), so the policies pol_y, pol_w are removed. At last, the algorithm returns the seller policy p_1=pol_z to the buyer. And this result is in accordance with our intuitive understanding.

6 Conclusions

In this paper, we present an ontology-based policy definition and matching model for privacy protection on third-party trading platform. Our model allows buyers and sellers to define privacy policy according to their preferences and then converts policies into ontology based expression forms. We introduce the concept good property of sellers who require the minimal information from buyers while satisfying other trading requirements. We propose a privacy policy matching algorithm to find such seller for a user request. Some experiments are performed to verify our model and an illustrative example is given. In the future, we would improve the policy matching algorithm and conduct more experiments to get an exact value of the threshold in the algorithm.

Acknowledgment. Part of this work is supported by the National Natural Science Foundation of China (61173140), the Science Foundation of Shandong Province(Y2008G28), and the Independent Innovation Foundation of Shandong University (2010JC010).

References

1. Denker, G., Kagal, L., Finin, T.: Security in Semantic Web using OWL. Information Security Technical Report 10, 51–58 (2005)
2. Aumueller, D., Do, H.H., Massmann, S., Rahm, E.: Schema and Ontology Matching with COMA++. In: Proceedings of the 5th ACM SIGMOD International Conference on Management of Data, New York, pp. 906–908 (2005)
3. Bonatti, P.A., Faella, M., Galdi, C., Sauro, L.: Towards a Mechanism for Incentivating Privacy. In: Atluri, V., Diaz, C. (eds.) ESORICS 2011. LNCS, vol. 6879, pp. 472–488. Springer, Heidelberg (2011)
4. Fenz, S.: An Ontology-and Bayesian-based Approach for Determining Threat Probabilities. In: Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security, New York, pp. 344–354 (2005)
5. Hecker, M., Dillon, T.S., Chang, E.: Privacy Ontology Support for E-Commerce. In: Proceeding of IEEE Computer Society 2008, Internet Computing, pp. 54–61 (2008)
6. Garcia, D.Z., Toledo, M.B.: A Web Service Privacy Framework Based on a Policy Approach Enhanced with Ontologies. In: Proceedings of the 11th IEEE International Conference on Computational Science and Engineering, San Paulo, pp. 209–214 (2008)
7. Tumer, A., Dogac, A., Toroslu, I.H.: A Semantic-Based User Privacy Protection Framework for Web Services. In: Mobasher, B., Anand, S.S. (eds.) ITWP 2003. LNCS (LNAI), vol. 3169, pp. 289–305. Springer, Heidelberg (2005)
8. Gao, F., He, J., Peng, S.: An Approach for Privacy Protection Based-on Ontology. In: Proceedings of 2010 Second International Conference on Networks Security, Wireless Communications and Trusted Computing, Wuhan, Hubei, pp. 397–400 (2010)
9. Carminati, B., Ferrari, E., Heatherly, R.: A Semantic Web Based Framework for Social Network Access Control. In: Proceedings of the 14th ACM Symposium on Access Control Models and Technologies, New York, pp. 177–186 (2009)
10. Masoumzadeh, A., Joshi, J.: OSNAC: An Ontology-Based Access Control Model for Social Networking Systems. In: Proceedings of IEEE International Conference on Social Computing, Minneapolis, MN, pp. 751–759 (2010)
11. Qian, J.A., Jiang, X.H., Sun, T.F.: Privacy Ontology-based Personalized Access Control Model. Information Security and Communications Privacy 2, 67–73 (2011)
12. Lan, L.H., Ju, S.G., Liu, S.C.: Survey of study on privacy preserving data publishing. Application Research of Computers 27, 2822–2827 (2010)
13. Hu, Y.J., Wu, W.N., Cheng, D.R.: Towards law-aware semantic cloud policies with exceptions for data integration and protection. In: Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics (2012)