Web Services Privacy Measurement Based on Privacy Policy and Sensitivity Level of Personal Information

Chaiwongsa P.

Computer Science Program, Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand
Email address: punyaphat.c@student.chula.ac.th

Senivongse T.

Computer Science Program, Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand
Email address: twittie.s@chula.ac.th

Abstract

Web services technology has been in the mainstream of today's software development. Software designers can select Web services with certain functionality and use or compose them in their applications with ease and flexibility. To distinguish between different services with similar functionality, the designers consider quality of service. Privacy is one aspect of quality that is largely addressed since services may require service users to reveal personal information. A service should respect the privacy of the users by requiring only the information that is necessary for its processing as well as handling personal information in a correct manner. This paper presents a privacy measurement model for service users to determine privacy quality of a Web service. The model combines two aspects of privacy. That is, it considers the degree of privacy principles compliance of the service as well as the sensitivity level of user information which the service requires. The service which complies with the privacy principles and requires less sensitive information would be of high quality with regard to privacy. In addition, the service WSDL can be augmented with semantic annotation using SAWSDL. The annotation specifies the semantics of the user information required by the service, and this can help automate privacy measurement. We also present a measurement tool and an example of its application.

Keywords: *privacy; privacy policy; personal information; measurement; Web services; ontology*

1 Introduction

Web services technology has been in the mainstream of software development since it allows software designers to use Web services with certain functionality in their applications with ease and flexibility. Software designers study service information that is published on service providers' Web sites or through service directories and select the services that have the functionality as required by the application requirements. For those with similar functionality, different aspects of quality of service (QoS) are usually considered to distinguish them.

Privacy is one aspect of quality that is largely addressed since Web services may require service users to reveal personal information. An online shopping Web service may ask a user to give personal information such as name, address, phone number, and credit card number when buying products, and a student registration Web service of a university would also ask for students' personal information to maintain student records. A Web service should respect the privacy of service users by requiring only the information that is necessary for its processing as well as handling personal information in a correct manner. From a view of a service user, proper handling of the disclosed personal information is highly expected. From a view of a software designer who is developing a service-based application, it is desirable to select a Web service with privacy quality into the application since the privacy quality of the service contributes to that of the application. The application itself should also respect the privacy of the application users.

In this paper, we present a privacy measurement model for service users to determine privacy quality of a Web service. The model combines two aspects of privacy. That is, it considers the degree of privacy principles compliance of the service as well as the sensitivity level of user information which the service requires. The model follows the approach by Yu et al. [1] which assesses if the privacy policy of a Web service complies with a set of privacy principles. We enhance it by also considering sensitivity level of users' personal information. The approach by Jang and Yoo [2] is adapted to determine sensitivity level of personal information that is exchanged with the service. According to our privacy measurement model, a service which complies with the privacy principles and requires less sensitive information would be of high quality with regard to privacy. In addition, we develop a supporting tool for the model. The tool relies on augmenting WSDL data elements of the service with semantic annotation using the SAWSDL mechanism [3]. The annotation specifies the meaning of WSDL data elements based on personal information ontology, i.e., a semantic term associated with a data element indicates which personal information the data element represents. Semantic annotation is useful for disambiguating user information that may be named differently by different Web services. As a result, it helps automate privacy measurement and facilitates the comparison of privacy quality of different Web services. Combining these two aspects of privacy, the model is considered practical for service users since the assessment is based on the privacy policy and service WSDL which can be easily accessed.

Section II of this paper discusses related work. Section III describes an assessment of privacy policy of a Web service based on privacy principles and Section IV presents measurement of sensitivity level of personal information. The privacy measurement model combining these two aspects of privacy is proposed in Section V. The supporting tool is described in Section VI and the paper concludes in Section VII.

2 Related Work

W3C has stated in the Web Services Architecture Requirements [4] that Web services architecture must enable privacy protection for service consumers. Web services must express privacy policy statements which comply with the Platform for Privacy Preferences (P3P), and the policy statements must be accessible to service consumers. Service providers generally publish privacy policy statements which follow privacy protection guidelines proposed by governmental or international organizations, and these statements are the basis for privacy protection measurement.

A. Related Work in Privacy Measurement Based on Privacy Policy

Following Canadian Standards Association Privacy Principles, Yee [5] specifies how to define privacy policy, and a method to measure how well a service protects user privacy based on measurement of violations of the user's privacy policy. The work is extended to consider compliances between E-service provider privacy policy and user privacy policy using a privacy policy agreement checker [6]. Similarly, Xu et al. [7] provide for a composite service and its user a policy compliance checker which considers sensitivity levels of personal data that flow in the service together with trust levels and data flow permission given to the services in the composition. Tavakolan et al. [8] propose a model for privacy policy and a method to match and rank privacy policies of different services with user's privacy requirements. We are particularly interested in the work by Yu et al. [1] which follows 10 privacy principles defined in the Australia National Privacy Principles (Privacy Amendment Act 2000). The work proposes a checklist to rate privacy protection of a Web service with regard to each privacy principle. A privacy policy checker which can be plugged into the Web service application is also developed to check for privacy principles compliance.

B. Related Work in Privacy Measurement Based on Sensitivity Level of Personal Information

Yu et al. [9] present a QoS model to derive privacy risk in service composition. The privacy risk is computed using the percentage of private data the users have to release to the services. The users can define weights that quantify a potential damage if the private data leak. Hewett and Kijsanayothin [10] propose privacy-aware service composition which finds an executable service chain that satisfies a given composite service I/O requirements with minimum number of services and minimum information leakage. To quantify information leakage, sensitivity levels are assigned to different types of personal information that flows in the composition. The composition also complies with users' privacy preferences and providers' trust. We are particularly interested in the comprehensive view of privacy sensitivity level of Jang and Yoo [2]. They address four factors of sensitivity, i.e. degree of conjunction, principle of identity, principle of privacy, and value of analogism. They also give a guideline to evaluate these sensitivity factors which we can adapt for the work.

3 Assessment of Web service Privacy Policy

For the privacy policy aspect, we simply adopt a privacy principles compliance assessment by Yu et al. [1]. According to the Australia National Privacy Principles (Privacy Amendment Act 2000), there are 10 privacy principles for proper management of personal information. For each principle, Yu et al. list a number of criteria to rate privacy compliance of a service. For full detail of the compliance checklist, see [1]. Here we show a small part of the checklist through our supporting tool in Figure 1. For instance, there are 3 criteria that a service has to follow to comply with the collection principle, i.e., the privacy policy statements must state (1) the kind of data being collected, (2) the method of data collection, and (3) the purpose of data collection. The service user can check with the published privacy policy how many of these criteria the service satisfies, and then give the compliance rating score. Thus for the collection principle, the maximum rating is 3; the rating ranges between 0-3. The service user can also define a weighted score for each privacy principle denoting the relative importance of each principle. The total privacy principle compliance (P_{com}) score of a service is computed by (1) [1]:

$$P_{com} = \sum_{i=1}^{10} \frac{r_i}{r_{imax}} * p_i \qquad (1)$$

where

r_i = rating for principle i assessed by service user

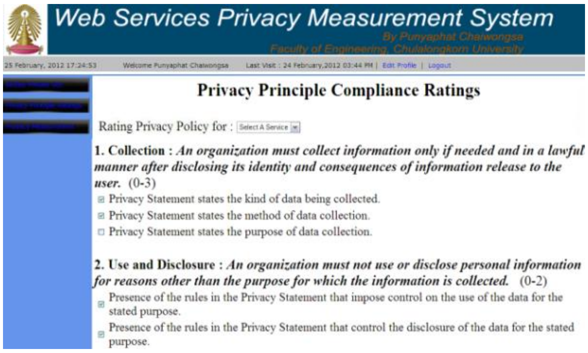
r_{imax} = maximum rating for principle i

p_i = weighted score for principle i assigned by service user, and $\sum_{i=1}^{10} p_i = 100 \cdot$

P_{com} ranges between 0-100. Instead we will later use a normalized NP_{com}, as in (2), which ranges between 0-1 in our privacy measurement model in Section V:

$$NP_{com} = \sum_{i=1}^{10} \left(\frac{r_i}{r_{imax}} * p_i \right) / 100 = \frac{P_{com}}{100} \cdot \qquad (2)$$

As an example, a user of a Register service of a university, which registers student information, rates and gives a weight for each privacy principle as in Table 1 P_{com} of this service then is 87.08 and NP_{com} is 0.87.


Web Services Privacy Measurement System
By Punyaphat Chaiwongsa
Faculty of Engineering, Chulalongkorn University
25 February, 2012 17:24:53 Welcome Punyaphat Chaiwongsa Last Visit : 24 February,2012 03:44 PM | Edit Profile | Logout

Privacy Principle Compliance Ratings

Rating Privacy Policy for : Select A Service

1. Collection : ***An organization must collect information only if needed and in a lawful manner after disclosing its identity and consequences of information release to the user.*** (0-3)

☑ Privacy Statement states the kind of data being collected.
☑ Privacy Statement states the method of data collection.
☐ Privacy Statement states the purpose of data collection.

2. Use and Disclosure : ***An organization must not use or disclose personal information for reasons other than the purpose for which the information is collected.*** (0-2)

☑ Presence of the rules in the Privacy Statement that impose control on the use of the data for the stated purpose.
☑ Presence of the rules in the Privacy Statement that control the disclosure of the data for the stated purpose.

Figure 1: Assessing privacy principles compliance using our tool.

Table 1: Example of Privacy Principles Compliance Rating

No.	**Privacy Principles**	**Rating** r_i	**Max Rating** r_{imax}	**Weight** p_i	**Score** $r_i/r_{imax}*p_i$
1	Collection	2	3	20	13.33
2	Use and Disclosure	2	2	10	10
3	Data Quality	2	2	5	5
4	Data Security	2	2	10	10
5	Openness	2	2	5	5
6	Access and Correction	3	4	5	3.75
7	Identifiers	2	2	2	2
8	Anonymity	0	1	5	0
9	Transborder Data Flows	2	2	8	8
10	Sensitive Information	1	1	30	30
			Total	100	$\mathbf{P_{com}}$ = 87.08 $\mathbf{NP_{com}}$ = 0.87

4 Assessment of Sensitivity Level of Personal Information

The motivation for assessing sensitivity level of personal information is that, for different Web services with similar functionality, a service user would prefer one to which disclosure of personal information is limited. It is therefore desirable that less number of personal data items is required by the service and the data items that are required are also less sensitive. We adapt from the approach by Jang and Yoo [2] which analyzes sensitivity level of personal information based on personal information classification.

A. Formal Concept Analysis and Ontology of Personal Information

Jang and Yoo represent personal information classification using a formal concept analysis (FCA) [11]. The formal definition of a data group, i.e., personal information in this case, is given as

$D_G = (G, N, R)$

where G is a finite set of concepts and can be described as $G = \{g_1, g_2, ..., g_n\}$,

N is a finite set of attributes which describe the concepts and can be described as $N = \{n_1, n_2, ..., n_m\}$, and

R is a binary relation between G and N, i.e., $R \subseteq G \times N$. For example, g_1 R n_1, or $(g_1, n_1) \in R$, represents that the concept g_1 has an attribute n_1.

The formal concepts can also be described using a cross table. We extend the cross table of [2] to create one as shown in Table 2. Here personal information is classified into 7 concepts, i.e., G = {*Basic*, *Career*,…, *Finance*}, and there are 37 personal information attributes, i.e., N = {*BirthPlace*, *BirthDay*, …, *CreditcardNumber*}. The cross table shows the relation, marked by an x, between each concept and attributes of the concept. For example, *BirthPlace* belongs in the *Basic* and *Private* concepts while the *Basic* concept has 15 attributes, i.e., *BirthPlace*, *BirthDay*, …, *DrivingLicenseNumber*.

For a Web service, its WSDL interface document defines what users' personal information is required for the processing of the service. However, different services with similar functionality may name the exchanged data elements differently. A service, for example, may require a data element called *Address* whereas another requires *Addr*. In order to infer that the two services require the same personal data, both *Address* and *Addr* elements in the two WSDLs can be annotated with the same semantic information. To disambiguate user information that may be named differently by different services, we augment WSDL data elements of a service with semantic annotation using the SAWSDL mechanism [3]. The annotation specifies the meaning of WSDL data elements based on personal information ontology. We represent the personal information concepts and attributes in the cross table (Table 2) as an OWL-based personal information ontology as in Figure 2. The attribute sawsdl:modelReference is associated with a data element in the WSDL document to reference to a semantic term in the ontology. In the WSDL of the Register service in Figure 3, the meaning of the data element called *Name* is the term *PersonName* in the ontology in Figure 2, etc. Semantic annotation is useful for automating privacy measurement and facilitates comparison of privacy quality of different services.

Table 2: Cross Table of Personal Information, Adapted from [2]

Concepts/Attributes	BirthPlace	BirthDay	BirthYear	Age	BloodType	BloodPressure	CellphoneNumber	PersonalEmailAddress	Family	BusinessEmailAddress	Gender	HomeAddress	HomephoneNumber	SocialSecurityNumber	PassportNumber	PersonName	DrivingLicenseNumber	CompanyName	CompanyAddress	CompanyphoneNumber	JobTitle	StaffID	Income	DiseaseName	DoctorName	HospitalName	Medicine	Weight	Height	PatientID	BankAccount	StudentID	SchoolName	SchoolAddress	Department	BankName	CreditcardNumber
Basic	X	X	X	X	X		X	X	X		X	X	X	X	X	X	X																				
Career							X			X				X		X		X	X	X	X	X	X														
Health			X	X	X	X	X				X			X		X								X	X	X	X	X	X	X							
Identity														X	X		X					X								X	X	X					
Private	X			X			X																X	X			X										
School			X				X	X			X			X		X																X	X	X	X		
Finance							X	X								X							X								X					X	X

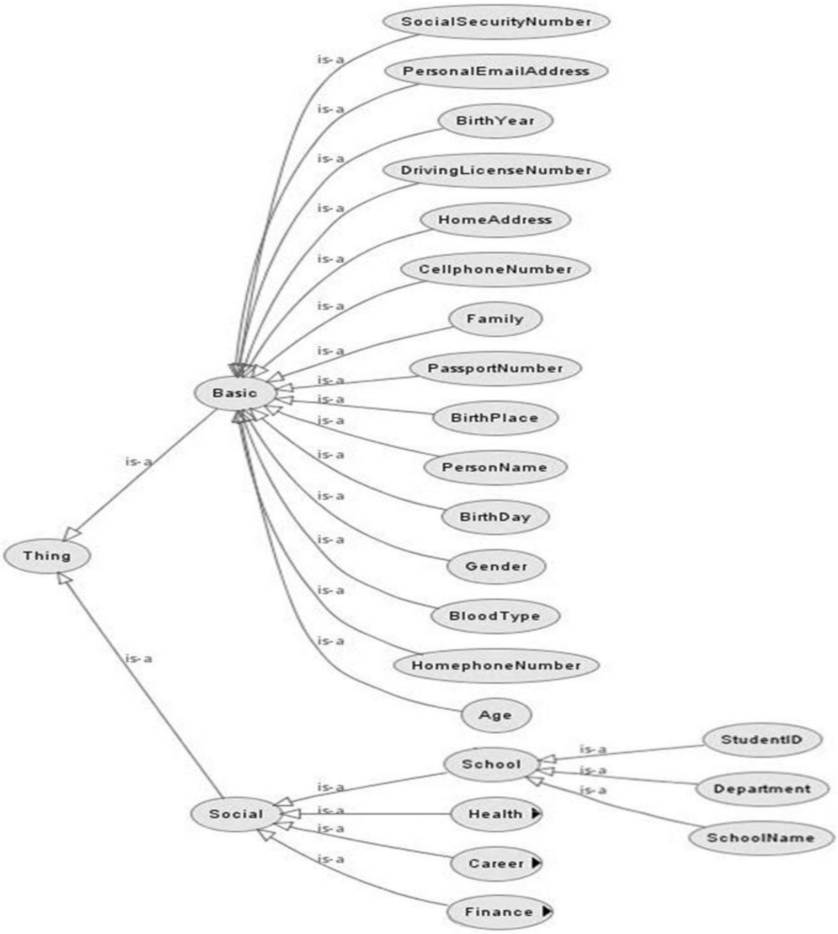
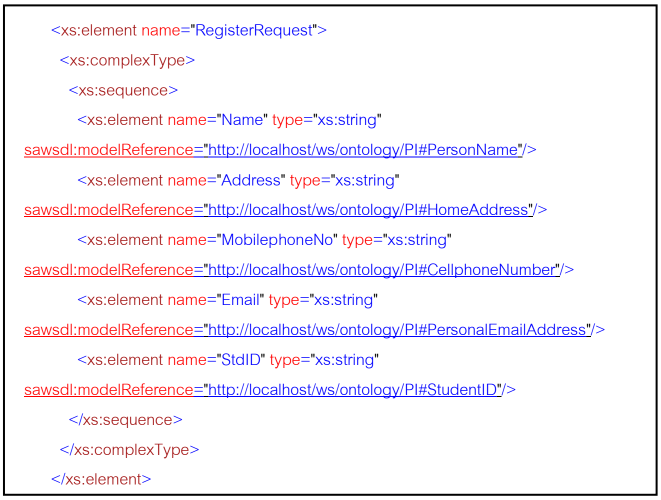


Figure 2: Part of personal information ontology.



```
<xs:element name="RegisterRequest">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="Name" type="xs:string"
sawsdl:modelReference="http://localhost/ws/ontology/PI#PersonName"/>
      <xs:element name="Address" type="xs:string"
sawsdl:modelReference="http://localhost/ws/ontology/PI#HomeAddress"/>
      <xs:element name="MobilephoneNo" type="xs:string"
sawsdl:modelReference="http://localhost/ws/ontology/PI#CellphoneNumber"/>
      <xs:element name="Email" type="xs:string"
sawsdl:modelReference="http://localhost/ws/ontology/PI#PersonalEmailAddress"/>
      <xs:element name="StdID" type="xs:string"
sawsdl:modelReference="http://localhost/ws/ontology/PI#StudentID"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
```

Figure 3: Part of semantics-annotated WSDL document.

B. Sensitivity Level of Personal Information

Jang and Yoo [2] address four factors of privacy sensitivity for personal information, i.e. degree of conjunction, principle of identity, principle of privacy, and value of analogism. They also give a guideline to evaluate these sensitivity factors which we can adapt for the work. We define the formula to compute the scores of these factors based on the cross table (Table 2) as follows.

1) *Degree of conjunction* of an attribute (personal data item) *n* is derived from the number of concepts which the attribute *n* describes. This means *n* is associated with these concepts and the disclosure of *n* may lead to other information belonging in these concepts. The degree of conjunction of *n* or $D_C(n)$ is determined by (3):

$$D_C(n) = \frac{number\ of\ concepts\ in\ which\ n\ belongs}{total\ number\ of\ concepts}. \quad (3)$$

For example, from Table 2, *PersonName* is associated with 5 out of 7 concepts, i.e., *Basic*, *Career*, *Health*, *School*, and *Finance*. Therefore $D_C(PersonName)$ = 5/7.

2) *Principle of identity* of an attribute *n* indicates that *n* is an identity attribute of the concept with which it is associated, i.e., *n* is used as a key information to access other attributes in that concept. Disclosure of *n* may then lead to more problems than disclosure of other attributes. The principle of identity of *n* or $I_A(n)$ is determined by (4):

$$I_A(n) = \begin{cases} 0 & if\ n\ is\ not\ identity\ attribute \\ \dfrac{number\ of\ attributes\ in\ the\ concepts}{total\ number\ of\ attributes} & if\ n\ is\ identity\ attribute\ for\ the\ concepts. \end{cases} \quad (4)$$

For example, from Table 2, *StudentID* is an identity attribute (i.e., it belongs in the concept *Identity*) for the concept *School*. There are 10 attributes associated with *School* and there are 37 attributes in total. Therefore $I_A(StudentID)$ = 10/37. For *HomeAddress*, it is not an identity attribute and $I_A(HomeAddress)$ = 0.

3) *Principle of privacy* of an attribute *n* indicates that *n* is private information. Note that this is subjective to the service users, e.g., some users may consider *Age* as private information whereas others may not. We let the service users customize the cross table by specifying which attributes are considered private, i.e., belong in the concept *Private*. The principle of privacy of *n* or $P_A(n)$ is determined by (5):

$$P_A(n) = \begin{cases} 0 & if\ n\ does\ not\ belong\ in\ the\ concept\ Private \\ 1 & if\ n\ belongs\ in\ the\ concept\ Private. \end{cases} \quad (5)$$

For example, from Table 2, *CellphoneNumber* is private and $P_A(CellphoneNumber)$ = 1, whereas *PersonalEmailAddress* is not and $P_A(PersonalEmailAddress)$ = 0.

4) *Value of analogism* of an attribute *n* indicates that *n* can be used to derive other attributes. This means the knowledge of *n* can also reveal other personal information. The value of analogism of *n* or $A_A(n)$ is determined by (6):

$$A_A(n) = \begin{cases} 0 & if\ n\ cannot\ derive\ other\ attributes \\ 1 & if\ n\ can\ derive\ other\ attributes. \end{cases} \quad (6)$$

The analogy between attributes has to be defined and associated with the cross table and the personal information ontology. For example, *SocialSecurityNumber* can derive other attribute such as *BirthPlace*, and $A_A(SocialSecurityNumber)$ = 1, whereas *Age* cannot and $A_A(Age)$ = 0.

All four sensitivity factor scores range between 0-1. Based on these scores, Jang and Yoo suggest that the sensitivity level of an attribute *n* or $S_L(n)$ be determined by (7) [2]:

$$S_L(n) = D_C(n) + I_A(n) + P_A(n) + A_A(n). \quad (7)$$

We propose to compute the sensitivity level of all personal information exchanged with a Web service using (8):

$$S_{Lws} = \sum_{i=1}^{k} S_{Li} \quad (8)$$

where k = number of exchanged personal data elements

S_{Li} = sensitivity level of personal data element i computed by (7).

We will later use a normalized NS_{Lws}, as in (9), which ranges between 0-1 in our privacy measurement model in Section V:

$$NS_{Lws} = \sum_{i=1}^{k} \frac{S_{Li}}{4k} = \frac{S_{Lws}}{4k}. \qquad (9)$$

As an example, suppose a Register service of a university requires the following personal information: Name, Address, MobilephoneNo, Email, and StdID. In the WSDL in Figure 3, these data elements are annotated with semantic terms described in the personal information ontology in Figure 2. We can determine the sensitivity level of each data element by calculating the sensitivity level of the associated semantic term using (7), and the total sensitivity level of all personal data required by the service using (8) and (9) as in Table 3.

5 Web Services Privacy Measurement Model

We combine the two privacy aspects in Sections III and IV into a privacy measurement model. The normalized privacy principles compliance NP_{com} of a service is a positive aspect. A service user would prefer a service with high compliance rating. The service provider is encouraged to follow privacy principles, provide proper management of users' personal information, and publish a clear privacy policy that can facilitate compliance rating by the service users. On the contrary, the normalized sensitivity level NS_{Lws} for the service is a negative aspect. Using a service which exchanges highly sensitive personal data could mean high risk of privacy violation if these data are disclosed or not protected properly.

Table 3: Example of Sensitivity Level Measurement

Data Element	**Semantic Annotation *n***	$D_C(n)$ **(3)**	$I_A(n)$ **(4)**	$P_A(n)$ **(5)**	$A_A(n)$ **(6)**	$S_L(n)$ **(7)**
Name	PersonName	5/7	0	0	0	0.71
Address	HomeAddress	1/7	0	0	0	0.14
Mobilephone Number	Cellphone Number	6/7	0	1	0	1.86
Email	PersonalEmail Address	3/7	0	0	0	0.43
StdID	StudentID	2/7	10/37	0	0	0.56
					Total	S_{Lws} =3.7 NS_{Lws} =3.7/ 4*5 =0.19

The privacy quality *P* of a service is computed by (10). The service user can also define weighted scores α and β to denote relative importance of the two privacy aspects; α and β are in [0, 1] and α + β = 1. The service which complies with the privacy principles and requires less sensitive information would be of high quality with regard to privacy.

$$P = \alpha NP_{com} + \beta(1 - NS_{Lws}). \qquad (10)$$

As an example, given equal weights to the two privacy aspects and the assessment in Tables 2 and 3, the privacy quality of the Register service is

P = (0.5)(0.87) + (0.5)(1 - 0.19)

= 0.435+0.405 = 0.84.

The Register service has high privacy principles compliance level and requires personal data that are relatively not so sensitive. It is therefore desirable in terms of privacy.

6 Development of Supporting Tool

Besides the proposed model, we have developed a Web-based tool called a privacy measurement system to support the model. To be able to automate privacy measurement, the tool relies on the service WSDL being annotated with semantic terms described in the personal information ontology. The usage scenario of the privacy measurement system is depicted in Figure 4 and can be described as follows.

1) The privacy measurement system obtains the cross table and personal information ontology from a privacy domain expert. In the prototype of the tool, the cross table in Table 2 and a personal information ontology that corresponds to the cross table are used.

2) A service user specifies the Web service to be measured the privacy. Together with the service WSDL URL, the user uses the tool to specify the following:

a) Privacy principles compliance rating r_i and weight p_i for each privacy principle; the user will have to check with the privacy policy of the service in order to rate.

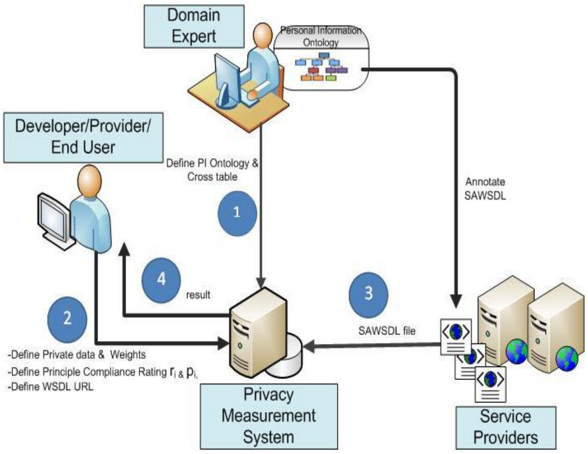

Domain Expert
Personal Information Ontology
Developer/Provider/ End User
Define PI Ontology & Cross table
1
Annotate SAWSDL
4
result
3
2
SAWSDL file
-Define Private data & Weights
-Define Principle Compliance Rating r_i & p_i,
-Define WSDL URL
Privacy Measurement System
Service Providers

Figure 4: Usage scenario of privacy measurement system.

b) Personal data attributes that are considered private; these attributes will be associated with the concept *Private* of the cross table.

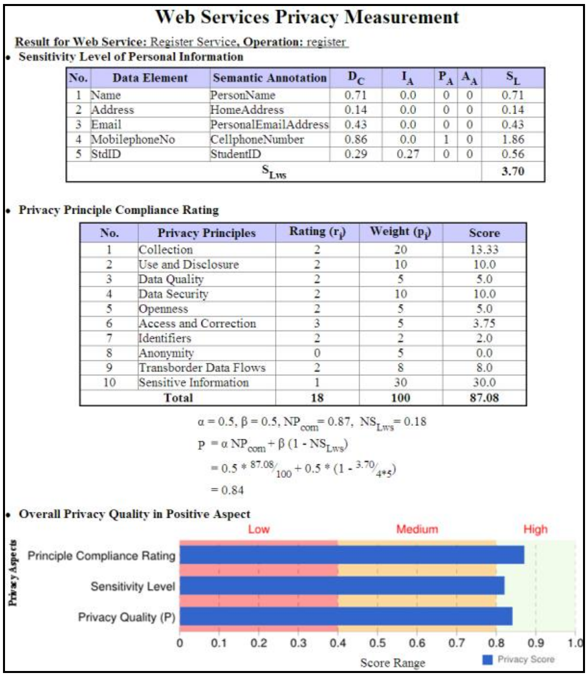
c) Weights *α* and *β* for the privacy measurement model.

The users of the tool could be end users of the services or software designers who are assessing privacy quality of the services to be aggregated in service-based applications. Additionally, service providers may use the tool for self-assessment; the measurement can be used for comparison with competing services and as a guideline for improving privacy protection.

3) The tool imports the WSLD document of the service. It is assumed that the service provider annotates the WSDL based on the personal information ontology.

4) The tool calculates the privacy score of the service and informs the user.

As an example, a screenshot reporting privacy measurements of the Register service is shown in Figure 5.



Web Services Privacy Measurement

Result for Web Service: Register Service, **Operation:** register

- **Sensitivity Level of Personal Information**

No.	Data Element	Semantic Annotation	D_C	I_A	P_A	A_A	S_L
1	Name	PersonName	0.71	0.0	0	0	0.71
2	Address	HomeAddress	0.14	0.0	0	0	0.14
3	Email	PersonalEmailAddress	0.43	0.0	0	0	0.43
4	MobilephoneNo	CellphoneNumber	0.86	0.0	1	0	1.86
5	StdID	StudentID	0.29	0.27	0	0	0.56
S_{Lws}							**3.70**

- **Privacy Principle Compliance Rating**

No.	Privacy Principles	Rating (r_i)	Weight (p_i)	Score
1	Collection	2	20	13.33
2	Use and Disclosure	2	10	10.0
3	Data Quality	2	5	5.0
4	Data Security	2	10	10.0
5	Openness	2	5	5.0
6	Access and Correction	3	5	3.75
7	Identifiers	2	2	2.0
8	Anonymity	0	5	0.0
9	Transborder Data Flows	2	8	8.0
10	Sensitive Information	1	30	30.0
Total		**18**	**100**	**87.08**

$\alpha = 0.5$, $\beta = 0.5$, $NP_{com} = 0.87$, $NS_{Lws} = 0.18$

$$P = \alpha\, NP_{com} + \beta\,(1 - NS_{Lws})$$
$$= 0.5 * {}^{87.08}/_{100} + 0.5 * (1 - {}^{3.70}/_{4*5})$$
$$= 0.84$$

- **Overall Privacy Quality in Positive Aspect**

Figure 5: Example of measurements screen.

7 Conclusions

This paper presents a privacy measurement model which combines and enhances existing privacy measurement approaches. The model considers both privacy principles compliance and sensitivity level of personal information. The basis of the measurement is the privacy policy published by the service provider and user's personal information that is exchanged with the service. The model can be applied even in the absence of any of such information. We present also a supporting tool which can automate privacy measurement based on semantic annotation added to WSDL data elements.

Generally a service user can consider the privacy score as one of the QoS scores to distinguish services with similar functionality. As discussed earlier, the privacy score is subjective to the users who assess the service. The score may vary depending on how the service provider provides a proof of privacy principles compliance, the expectation of the user when rating the compliance, and the user's personal view on private data. Also, the cross table presented in Table 2 is an example but not intended to be exhaustive. A privacy measurement system can adjust the concepts, attributes, and their relations within the cross table as well as the corresponding personal information ontology.

Since the measurement tool makes use of semantics-enhanced WSDLs, a limitation would be that we require the service providers to specify semantics. However, semantic information only helps automate the calculation and the measurement model itself does not rely on semantic annotation. The approach can still be followed and the measurement model can still be used even though WSDL documents are not semantics-annotated.

At present, we target privacy of single Web services. The approach can be extended to composite services. We are planning for an empirical evaluation of the model by service users and an experiment with real-world Web services as well as cloud services.

References

[1] W. D. Yu, S. Doddapaneni, and S. Murthy, "A privacy assessment approach for service oriented architecture applications," in Procs. of 2nd IEEE Int. Symp. on Service-Oriented System Engineering (SOSE 2006), 2006, pp. 67-75.

[2] I. Jang and H. S. Yoo, "Personal information classification for privacy negotiation," in Procs. of 4th Int. Conf. on Computer Sciences and Convergence Information Technology (ICCIT 2009), 2009, pp. 1117-1122.

[3] W3C, Semantic Annotations for WSDL and XML Schema, http://www.w3.org/TR/2007/REC-sawsdl-20070828/, 28 August 2007.

[4] W3C, Web Services Architecture Requirements, http://www.w3.org/TR/wsa-reqs/, 11 February 2004.

[5] G. Yee, "Measuring privacy protection in Web services," in Procs. of IEEE Int. Conf. on Web Services, 2006, pp.647-654.

[6] G. O. M. Yee, "An automatic privacy policy agreement checker for E-services," in Procs. of Int. Conf. on Availability, Reliability and Security, 2009, pp. 307-315.

[7] W. Xu, V. N. Venkatakrishnan, R. Sekar, and I. V. Ramakrishnan, "A framework for building privacy-conscious composite Web services," in Procs. of IEEE Int. Conf. on Web Services, 2006, pp. 655-662.

[8] M. Tavakolan, M. Zarreh, and M. A. Azgomi, "An extensible model for improving the privacy of Web services," in Procs. of Int. Conf. on Security Technology, 2008, pp. 175-179.

[9] T. Yu, Y. Zhang, Y., K. J. Lin, "Modeling and measuring privacy risks in QoS Web services," in Procs. of 8th IEEE Int. Conf. on E-Commerce Technology and 3rd IEEE Int. Conf. on Enterprise Computing, E-Commerce, and E-Services, 2006.

[10] R. Hewett and P. Kijsanayothin, "On securing privacy in composite web service transactions," in Procs. of 5th Int. Conf. for Internet Technology and Secured Transactions (ICITST'09), 2009, pp. 1-6.

[11] Uta Priss, "Formal Concept Analysis," http://www.upriss.org.uk /fca/ fca.html/, Last accessed: 24 February 2012.