Visual analysis of privacy risks in web services

Yee, Georg

Publisher's version / Version de l'éditeur:

Proceedings of the IEEE International Conference on Web Services 2007 (ICWS 2007), 2007

NRC Publications Archive Record / Notice des Archives des publications du CNRC :
https://nrc-publications.canada.ca/eng/view/object/?id=51f55842-a93a-4cbe-a85f-6b4689baaad2
https://publications-cnrc.canada.ca/fra/voir/objet/?id=51f55842-a93a-4cbe-a85f-6b4689baaad2

National Research Council Canada
Conseil national de recherches Canada

Institute for Information Technology
Institut de technologie de l'information

NRC·CNRC

Visual Analysis of Privacy Risks in Web Services *

Yee, G.
July 9-13, 2007

* Proceedings Proceedings of the IEEE International Conference on Web Services 2007 (ICWS 2007). Salt Lake City, Utah, USA. July 9-13, 2007. NRC 49303.

Canada

Visual Analysis of Privacy Risks in Web Services[1]

George Yee
Institute for Information Technology
National Research Council Canada
george.yee@nrc.ca

Abstract

The growth of the Internet has been accompanied by the growth of web services (e.g. e-commerce, e-health) leading to the need to protect the privacy of web service users. However, before privacy can be protected, it is necessary to understand the risks to privacy that come with the service. Indeed, such understanding is key to protecting privacy throughout the service lifecycle. Unfortunately, there does not appear to be any existing method for privacy risk analysis specifically designed for web services. This paper presents a straightforward method for web services privacy risk analysis that uses visual techniques to improve effectiveness and illustrates the method with an example.

1. Introduction

This work considers web services to be: a) web-based services that employ XML (eXtensible Markup Language), WSDL (Web Service Definition Language), SOAP (Simple Object Access Protocol), and UDDI (Universal Description, Discovery, and Integration) in a service oriented architecture (SOA) [1], and b) existing and previous generations of web-based applications that involve web browsers interacting with web servers that do not employ XML, WSDL, SOAP or UDDI.

Numerous web services targeting consumers have accompanied the rapid growth of the Internet. Web services are available for banking, shopping, learning, healthcare, and Government Online. However, most of these services require a consumer's personal information in one form or another, leading to concerns over privacy. For web services to be successful, privacy must be protected.

Various approaches have been used to protect personal information, including data anonymization [2, 3] and pseudonym technology [4]. Approaches for privacy protection that are in the research stage include treating privacy protection as an access problem and then bringing the tools of access control to bear for privacy control [5], treating privacy protection as a privacy rights management problem using the techniques of digital rights management [6], and considering privacy protection as a privacy policy compliance problem, verifying compliance with secure logs [7].

The various approaches for protecting privacy described above all presume to know where and what protection is needed. They presume that some sort of analysis has been done that answers the question of "where" and "what" with respect to privacy risks. Without such answers, the effectiveness of the protection comes into question. For example, protection against house break-ins is totally ineffective if the owner only secures the front door without securing other vulnerable spots such as windows (the "where"). Of course, how the owner secures these spots is critical too ("what" protection). An effective break-in risk analysis would have identified the windows as being vulnerable to break-ins as well and, provided that the owner uses this information wisely, would have led to the owner additionally securing the windows. The result is a house that is better protected against break-ins. In the same way, privacy risk analysis considering "where" and "what" is essential to effective privacy protection - this work proposes a method for such analysis.

The objectives of this paper are to a) propose an effective method for privacy risk analysis that incorporates visual techniques to identify where and what protection (in terms of risk) is needed, and b) illustrate the method using a web service example. The privacy risk analysis is limited to the identification of privacy risks. It does not include estimating how likely it is that a risk will be realized. In addition, the web services to which this work applies make use of the service user's personal information in order to provide their services.

In the literature, there are significant works on security threat analysis but very little work on privacy

risk analysis. In fact, the only works that are directly related to privacy risk analysis appear to be the documents on "privacy impact assessment (PIA)" originating from government policy [8]. PIA is meant to evaluate the impact to privacy of new government programs, services, and initiatives. PIA can also be applied to existing government services undergoing transformation or re-design. However, PIA is a long manual process consisting mainly of self-administered questionnaires. It has not been tailored for use in a web service nor does it employ visual techniques as proposed in this work.

This paper is organized as follows. Section 2 defines privacy, privacy policies, privacy risks, and what they mean for web services. Section 3 presents the proposed method for web service privacy risk analysis, together with an application example. Section 4 discusses related work. Section 5 gives an evaluation of the proposed method. Section 6 presents conclusions and directions for future research.

2. Privacy and web services

As defined by Goldberg et al. in 1997 [9], privacy refers to the ability of individuals to *control* the collection, retention, and distribution of information about themselves. This leads to the following definitions for this work.

DEFINITION 1: *Privacy* refers to the ability of individuals to *control* the collection, use, retention, and distribution of information about themselves.

DEFINITION 2: A user's *privacy policy* is a statement that expresses the user's desired control over a web service's collection, use, retention, and distribution of information about the user.

DEFINITION 3: A *user privacy risk* of a web service is the potential occurrence of any action or circumstance that will result in a violation of a user's privacy policy.

Definition 1 is the same as given by Goldberg et al. except that it also includes "use". To see that "use" is needed, consider, for example, that one may agree to give out one's email address for use by friends to send email but not for use by spammers to send spam. This definition also suggests that "personal information", "private information" or "private data" is any information that can be linked to a person; otherwise, the information would not be "about" the person. Thus, another term for private information is "personally identifiable information (PII)". These terms are used interchangeably in this paper. The linking can be explicit, e.g. the person's name is attached to the information, or implicit, e.g. the information is part of a transaction that was initiated by a specific person.

Definition 2 refers to a user's privacy policy. In this work, the web service provider also has a privacy policy that details the control that the provider is willing to accept from the user's privacy policy. User information can only be disclosed to the provider if both the user's policy and the provider's policy are in agreement with each other. Figure 1 (adapted from [10]) gives an example of user/provider privacy policies for a web service that implements an online pharmacy. *Policy Use* indicates the type of web service for which the policy will be used. *Valid* holds the time period during which the policy is valid. The fields *collector*, *what*, *purposes*, *retention time*, and *disclose-to* are mandatory. They respectively indicate who is to receive the information, what is the information, for what purposes will the information be used, how long the provider can retain the information, and who outside the provider's organization can also receive the information. These fields derive from privacy principles that reflect privacy legislation shared by many countries, including Canada, the United States, the European Union, and Australia [11].

The policies in Figure 1 are minimum privacy policies in the sense that for any information item, the fields *collector*, *what*, *purposes*, *retention time*, and *disclose-to* form the minimum set of fields required to satisfy privacy legislation. Each set of such fields is termed a *privacy rule* describing a particular information item. For computer-based privacy management, privacy policies need to be machine-readable. This may be accomplished by expressing them in a XML-based language such as APPEL [12].

Policy Use: *Pharmacy* Owner: *Alice Buyer* Valid: *unlimited*	Privacy Use: *Pharmacy* Owner: *A-Z Drugs Inc.* Valid: *unlimited*
Collector: A-Z Drugs Inc. *What:* name, address, tel *Purposes:* identification *Retention Time:* unlimited *Disclose-To*: none *Collector:* A-Z Drugs Inc. *What:* drug name *Purposes:* purchase *Retention Time:* 2 years *Disclose-To*: none	*Collector:* Drugs Dept. *What:* name, address, tel *Purposes:* identification *Retention Time:* 1 year *Disclose-To*: none *Collector:* Drugs Dept. *What:* drug name *Purposes:* sale *Retention Time:* 1 year *Disclose-To*: none

Figure 1. Example user (left) and provider (right) privacy policies

In this work, a privacy risk analysis refers to an analysis of *user* privacy risk. This work considers only risks that involve potential violations of user privacy policies (Definition 3) where such policies are derived from privacy legislation. In other words, this work concerns only violations of the fields *collector*, *what*, *purposes*, *retention time*, and *disclose-to*, which have been enacted by privacy legislation as fully describing the privacy rights of individuals. Thus, this work can be seen to have a firm legislative basis, and at the same time can be extended to other concerns should that be necessary.

3. Method for web service privacy risk analysis

3.1. Web service personal information model

The Web Service Personal Information Model (WSPIM) more formally describes the relationship between a web service and the personal information of a user of the web service. The proposed method for privacy risk analysis is based on this model. WSPIM comprises the following principles:

a) The web service requires the user's personal information in order to carry out its service to the user. For example, a book seller web service requires the user's address for shipping purposes.
b) The web service and the user exchange privacy policies prior to the start of the service. These policies must agree and be accepted by both the web service and the user before the service can begin. If there is disagreement, the web service provider and the user can try to negotiate to a mutually acceptable privacy policy [13, 14].
c) The web service obtains the user's personal information after agreeing with the user's privacy policy, either before the service begins, during the course of the service, or both.
d) The web service agrees that once it is in possession of the user's personal information, it will make every effort within its power to comply with the user's privacy policy.
e) Once the web service is in possession of the user's personal information, the web service may transmit the information (e.g. move it from one group to another within the web service's organization), store the information (e.g. store the information in a data base), and make use of the information to provide the service (e.g. print out shipping labels with the user's address).

In part e), there are many ways to "make use of the information". Some examples are: as input to a calculation, for payment (e.g. credit card number), as input to a search process, transformed as anonymized input to a survey, or combined with other data for display in a report. In part d), and for this work, the web service is assumed to make every reasonable effort to comply with the user's policy in *good faith*, i.e. the web service is not malicious. However, violations of the user's policy by malicious employees or other insiders of the web service are still possible and are not treated in this work.

3.2. Method for privacy risk analysis

The method for privacy risk analysis is based on the notion that potential violations of the user's privacy policy arise from where the personal information is *located.* This idea is well recognized and applied by traditional non-electronic services, where privacy may be protected by keeping paper documents containing sensitive personal information in a safe. For a web service, storing the user's personal information in an encrypted database with secure access controls is the equivalent of storing it in a safe, with corresponding minimal privacy risks. The method, then, consists of a) determining all the possible locations where the user's personal information could reside while in the possession of the web service, and b) evaluating at each of these locations the possible ways in which the user's privacy policy could be violated. More completely, the method is as follows:

Method for Web Service Privacy Risk Analysis

1. Draw a Personal Information Map (PIM) showing the paths of all personal information flows of the web service, based on part e) of the WSPIM, namely, that personal information can be transmitted, stored, and used. Use an arrow to represent the transmission of personal information items that are described by privacy rules in the user's policy (see Figure 1). Label the arrow with numbers, where each arrow number corresponds to a description of a single data item in a legend. Use a square to represent the storage of personal information. Use a circle to denote the use of the information. Use a dashed rectangle to enclose circles or squares into physically distinct units. For example, two circles representing two uses would be enclosed by a dashed square if both uses are run on the same computer. Physically separate units allow the identification of risks for any data transmission between them. Circles or squares not enclosed by a dashed rectangle are understood to be already physically separate units. Label the

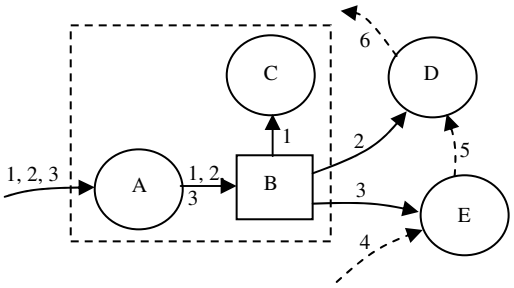
squares and circles with letters. Each such label corresponds to a description of the type of storage or the type of use as indicated in the legend.

2. Use dashed arrows, numbered in the same way as the arrows in Step 1, to add to the map all non-personal information flows, if any, that are involved with the transmission, storage and use of the personal information. Non-personal information is information that is not personal or not private, i.e. information that cannot identify any particular individual, e.g. the price of a book. Figure 2 illustrates steps 1 and 2 for a book seller web service that requires the user's name, address, book selection, and credit card number. These are considered as three personal information items where name and address together are considered as one item. Figure 2 also shows three non-personal information flows (4, 5, 6). The dashed rectangle enclosing A, B, and C indicates that A, B, and C all run on the same physical computer.
3. Inspect the map resulting from step 2, and for each location (transmission path arrow, storage square, and use circle) and each personal information item, enumerate the possible ways in which a privacy rule may be violated in terms of violations of each of *collector*, *what*, *purposes*, *retention time*, and *disclose-to* (see Section 2) in turn. This may be achieved by asking risk questions for each field, as suggested in Table 1, and drawing conclusions based on knowledge and experience with information security and systems. The risk questions are "how" questions, based on the idea that a risk arises where there is some way (i.e. how) for a violation to occur. Record the results in a Privacy Risks Table containing two columns: the left column for records of the form "(PII_1, PII_2, …/ locations)" and the right column containing the corresponding privacy risks. The Privacy Risks Table is the goal of the method. Table 2 illustrates this step for the book seller of Figure 2.

Table 1. Risk questions

Field	Risk Questions
collector	How can the PII be received by an unintended collector either in addition to or in place of the intended collector?
what	How can the user be asked for other PII, either intentionally or inadvertently?
purposes	How can the PII be used for other purposes?
retention time	How can the PII retention time be violated?
disclose-to	How can the PII be disclosed either intentionally or inadvertently to an unintended recipient?

It is important to remember that the PIM resulting from Step 2 is not a program logic flow diagram and one should not try to interpret it as such. It shows *where* personal information goes, *where* it is stored, and *where* it is used, corresponding to the notion that the location of personal information is key to understanding potential violations of privacy policy, mentioned at the start of this section.



Legend:

A: receive and store data	2: book selection
B: database	3: credit card number
C: print shipping label	4: company account number
D: pack book for shipping	5: payment status
E: charge credit card	6: shipping status
1: name and address	

Figure 2. PIM for a book seller web service

Table 2. Partial Privacy Risks Table corresponding to Figure 2

(PIIs / locations)	Privacy Risks
(1, 2, 3 / path into A); (2 / path into D); (3 / path into E)	Man-in-the-middle attack violates *collector*, *purposes*, and *disclose-to*; for path into A, user could be asked for personal information that violates *what*
(1, 2, 3 / A, B); (1 / C); (2 / D); (3 / E)	Trojan horse, hacker, or SQL attack (for B) violates *collector*, *purposes*, and *disclose-to*; for B, information could be kept past *retention time*

Adding other non-personal information flows in Step 2 is important to identify potential unintended leakages of personal information. For example, personal information may be "anonymized" (any obvious links to the information owner removed) and placed in a report together with non-personal information for public distribution. The presence of the non-personal information flows together with the personal information flows, both directed to a "produce report" use circle could lead to identifying a personal information leakage risk.

This method targets the identification of all risks and hence does not take into account any existing

security technology that might lessen or eliminate a risk.

A web service may make use of other web services in offering its service. For example, a book seller web service may make use of a payment processing web service and a shipping web service to sell a book. For the sake of exposition, the *primary* web service is the service that the user chooses to use. *Secondary* web services are services that the primary service makes use of to operate its service. A primary web service that makes use of secondary web services in this way is called a *multi-provider* web service. In order to analyze the privacy risks of a multi-provider web service, it is also necessary to analyze the privacy risks of all its secondary web services that receive the user's personal information. Secondary web services are identified in the user's privacy policy under *disclose-to*. Of course, all secondary services have to also agree to comply with the user's privacy policy.

The above method is best implemented by a privacy risks analysis team, consisting of no more than three or four people, selected for their technical knowledge of the web service and their knowledge of the web service organization's work procedures and processes. Good candidates for the team include the web service's design manager, test manager, and other line managers with the required knowledge. The team should be led by the privacy analyst who must also be knowledgeable about security threats and who should have the support of upper management to carry out the privacy risks analysis. A definite advantage of the team approach would accrue to step 3, where the enumeration would be more thorough by virtue of more people being involved in its brainstorming.

3.3. Application example

Consider a drug store web service Easy123Drugs.com (e.g. Walgreens.com). Easy123Drugs is a multi-provider service that makes use of two business web services: an accounting service AccountingAsNeeded.com (e.g. cbiz.com), and an online payment service PayAsYouLikeIt.com (e.g. Paypal.com). Suppose Easy123Drugs, AccountingAsNeeded, and PayAsYouLikeIt (all fictitious names with no hits on Google) are all web services that are based on the Service Oriented Architecture [1], employing XML-based protocols. Due to space limitations in this paper, the details regarding UDDI lookup and service binding via SOAP and WSDL [1] will not be described here. It is assumed that these initialization steps occur as required.

Table 3 shows the user's personal information required by each service. The user provides her private information to Easy123Drugs once her privacy policy has been accepted and agreed to by all the services. Easy123Drugs then discloses this information to AccountingAsNeeded and PayAsYouLikeIt according to the user's privacy policy.

Table 3. Personal information required

Web Service	User Personal Information Required
Easy123Drugs	name and address, prescription (patient name, drug name, doctor's name, authorization), credit card details (name, credit card number, expiry date)
PayAsYouLikeIt	credit card details (as above)
AccountingAs Needed	name and address, prescription (as above)

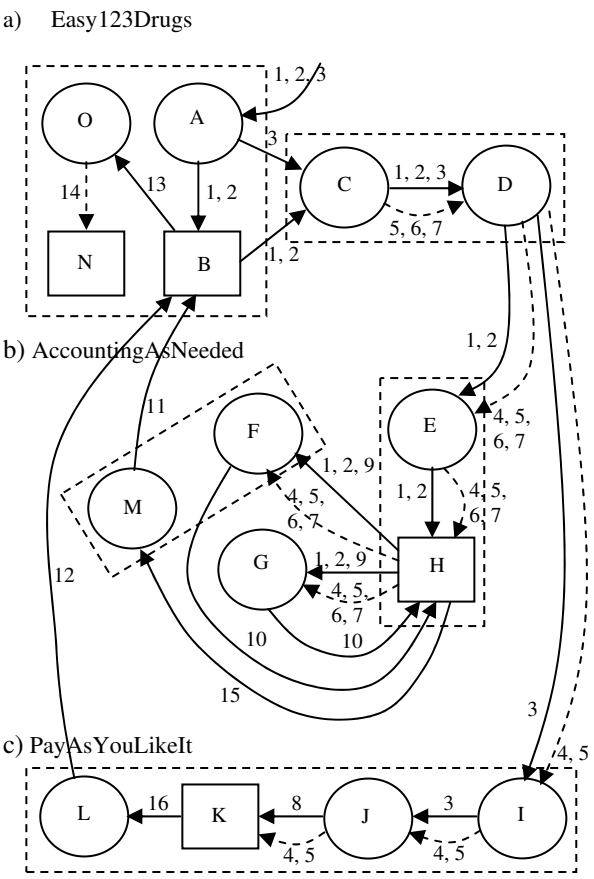
The proposed method for privacy risks analysis is carried out as follows:

Steps 1 and 2: Draw the PIM for each web service (see Figure 3, next page). As shown in Figure 3, there are some uses of personal information that have not been mentioned. First, both AccountingAsNeeded and PayAsYouLikeIt send activity reports back to Easy123Drugs that contain personal information (L and M). These reports contain selections and re-arrangements of the original personal data (15, 16). Second, Easy123Drugs produces a share holders' report that is in the public domain, and to do so, it selects, re-arranges, and anonymizes original personal data (13, 14). Third, AccountingAsNeeded allows its employees to partially work from home (G). Finally, the three web services do not store the user's credit card details in their databases.

Step 3: Enumerate privacy risks at private information locations. Table 4 gives a partial Privacy Risk Table for locations in Figure 3 that have interesting or serious privacy risks. The theft of personal information means that the information is under the control of an unintended party. Clearly, this can violate the corresponding privacy rule or rules in terms of violating *collector*, *purposes*, *retention time*, and *disclose-to*. The risk of personal information theft arises so often that it is convenient to call it *CPRD-risk*, from the first letters of collector, purposes, retention time, and disclose-to.

To illustrate this step, the risks in the first 3 rows of Table 4 were obtained as follows. For the first row, it was noticed that the personal information flowed through transmission paths connecting physically distinct units. The risk questions of Table 1 were then considered, leading to possible man-in-the-middle

attacks that give rise to CPRD-risk. In addition, violations of *what* are always possible unless strict controls are in place against it. For the second row, it was observed that the associated personal data are input to information use processes (e.g. A, C, D). The risk questions of Table 1 were again considered, leading to possible Trojan horse or hacker attacks that again give rise to CPRD-risk. For the third row, it was noticed that personal data are stored in databases. Once again the risk questions were considered, leading to possible SQL attacks against the databases, giving rise to CPRD-risk. In each of these three cases, knowledge of the system (personal data locations) and knowledge of information security (possible attacks) were needed to identify the risks. The remaining risks in Table 4 were derived in a similar fashion.



Legend:

A: receive and store data
B: database
C: process order
D: disclose data
1: name and address
2: prescription
3: credit card details
4: business id
5: order id
6: quantity of drug sold
7: price paid by user
8: user account update
9: current ledger record
10: updated ledger record
11: accounting report
12: payment report
13: performance data
E: receive and store data
F: update ledgers at work
G: update ledgers at home
H: database
I: receive and store data
J: charge credit card and update business account
K: database
L: compose payment report
M: compose accounting report
N: share holders' report
O: compose share holder's report
14: anonymized performance data
15: accounting data
16: payment data

Figure 3. PIMs for a) Easy123Drugs, b) AccountingAsNeeded, and c) PayAsYouLikeIt

Table 4. Partial Privacy Risks Table corresponding to Figure 3

(PIIs / locations)	Privacy Risks
(1, 2, 3 / path into A); (1, 2 / path between D and E); (3 / path between D and I); (12 / path between L and B); (11 / path between M and B)	Man-in-the-middle attacks lead to CPRD-risk; corresponding to 1, 2, 3, the user could be asked for personal information that violates *what.*
(1, 2, 3 / A, C, D); (13 / O); (1, 2 / E); (1, 2, 9 / F, G); (15 / M); (3 / J); (16 / L)	Trojan horse, or hacker attacks on the personal information use processes lead to CPRD-risk.
(1, 2, 11, 12 / B); (1, 2, 10 / H); (8 / K)	Potential SQL attacks on B, H, and K lead to CPRD-risk.
(13 / O)	A bad anonymization algorithm can expose personal information, leading to CPRD-risk.
(1, 2, 9 / G)	An insecure home environment, e.g. people looking over the shoulder or printed personal information lying on a desk in the clear, can also lead to CPRD-risk.
(1, 2, 9 / G)	If an employee works from home on a laptop and carries the laptop back and forth between home and work, possible theft or loss of the laptop can also lead to CPRD-risk for any of 1, 2, or 9 that might be temporarily stored in the laptop.
(1, 2, 9 / G)	If an employee works from home on a home PC and stores 1, 2, 9 on a flash memory stick, carrying the memory stick back and forth between home and work, possible theft or loss of the memory stick can also lead to CPRD-risk.

4. Related work

The literature on works dealing directly with privacy risk analysis for web services appears to be non-existent. However, the following authors have written on privacy topics that relate well with privacy risk analysis. Hong et al. [15] propose the use of privacy risk models to help designers design ubiquitous computing applications that have a reasonable level of privacy protection. Their privacy risk model consists of two parts: a privacy risk analysis part and a privacy risk management part. The risk analysis identifies the privacy risks while the risk management part is a cost-benefit analysis to prioritize the risks and design artifacts to manage the risks. Hong et al.'s privacy risk analysis is similar to a privacy impact analysis, consisting of a series of questions for the designer to answer that help to identify the privacy risks. Visualization is not used. Karger [16] presents a privacy and security threat analysis of the American Federal Employee Personal Identity Verification Program based on the standard FIPS PUB 201 [16]. However, the privacy threat analysis does not appear to be based on any published method but is done in an ad hoc fashion based on personal knowledge and thinking through scenarios.

Another class of related work is of course the work on privacy impact analysis (PIA) [8] already mentioned in Section 1. To reiterate, PIA is a manual process, consisting of a series of questionnaires that are answered by the privacy analyst or a team of privacy analysts in order to identify "impacts" to privacy of a new service or a change to an existing service. It is not specifically designed for web services nor does it use the visualization techniques proposed here.

A third class of related work is the work on security threat analysis, e.g. [17]. Security threats are related to privacy risks because such threats can increase privacy risks. For example, a Trojan horse attack (security threat) can lead directly to the lost of privacy when private data is unwittingly disclosed to the attacker. Security threat analysis for a computer system involves a) understanding the system, b) identifying the parts of the system that are vulnerable to attack, c) identifying the possible attacks and how those attacks can be carried out, and d) identifying the likelihood of those attacks occurring. Appropriate countermeasures are then installed for the vulnerable parts of the system that are associated with high likelihood attacks.

Finally, the notation proposed here for the PIM is similar in style to the Data Flow Diagram (DFD) that was popular in the 1970's and 1980's in the context of structured programming [18]. However, DFDs are used to represent all data flow, for the purpose of understanding system functionality. The contributions of this work lie in proposing new notation, as well as specializing and adapting existing notation such as DFDs, for use in visual privacy risk analysis, together with a structured method for such analysis.

5. Evaluation

This section evaluates the proposed method primarily based on its strengths and weaknesses, since there appears to be no other similar method (as mentioned in the section above on related works) with which to do a comparison evaluation.

Some of the strengths of the method include: a) provides a structured way to evaluate privacy risks, b) easy-to-use graphical notation, c) focuses the attention of the privacy analyst on risks that arise based on the locations that hold the personal information, and d) it appears (but remains to be shown) that the method is scalable, i.e. it seems that larger systems simply require more (in a linear fashion) time and paper with which to draw the PIM and construct the Privacy Risks Table.

Some weaknesses of the method are: a) drawing the PIM and filling out the Privacy Risks Table require expertise in how personal information is used in the service as well as expertise in security and privacy, b) the method is a manual process that is prone to error, and c) the method can never identify all the risks. Weakness a) is unavoidable as even expert systems must get their expertise from people. Also, this "weakness" is common to many analytical methods, e.g. designing good software. Weakness b) can be attenuated by building tools with which to semi-automatically draw the PIM. Similar tools already exist for rendering a software architecture diagram from the reverse engineering of code. A rules engine could also be used to partially automate the enumeration of privacy risks based on machine understanding of the above graphical notation. Weakness c) may have to stand as it is very difficult if not impossible to overcome. It is due to the nature of security, that no system can be completely secure. However, the above mentioned rules engine and automated tools could improve risk coverage.

The proposed method can be applied to any web service provider that offers its service to end users (e.g. e-banking, e-learning, e-health, B2C e-commerce), since user personal privacy is more likely to be required in such applications. The method can also be easily adapted to B2B e-commerce provided that privacy policies are used to manage privacy at both ends of the transaction (adaptations may be needed to accommodate new types of privacy rules).

6. Conclusions and future research

The rapid growth of web services has led to various methods to protect privacy. However, before privacy can be protected, it is necessary to understand the risks to privacy that come with the service. Indeed, such understanding is key to protecting privacy throughout the service lifecycle. This work has proposed a straightforward method for visual analysis of privacy risks in web services, focusing the analyst's attention at locations that hold personal information at one time or another. The method only identifies possible privacy risks and does not evaluate the likelihood of a risk being realized. However, identifying the risks is a necessary first step.

Plans for future research include: a) experimenting with the method to determine its effectiveness by applying it to real world web services, b) building tools for use in drawing the PIM, c) programming the graphical notation to be machine-readable, d) experimenting with a rules engine to semi-automatically enumerate the privacy risks based on reading the PIM and a set of rules, and e) extending the method to evaluate the likelihoods of risk realization.

7. References

[1] M. O'Neill et al, *Web Services Security*, McGraw-Hill/Osborne, 2003.

[2] V.S. Iyengar, "Transforming Data to Satisfy Privacy Constraints", *Proceedings, SIGKDD'02*, Edmonton, Alberta, 2002.

[3] A. Kobsa, J. Schreck, "Privacy through Pseudonymity in User-Adaptive Systems", *ACM Transactions in Internet Technology*, Vol. 3, No. 2, May 2003, pp. 149-183.

[4] R. Song, L. Korba, and G. Yee, "Pseudonym Technology for E-Services", chapter in *Privacy Protection for E-Services*, edited by G. Yee, Idea Group, Inc., 2006.

[5] C. Adams and K. Barbieri, "Privacy Enforcement in E-Services Environments", chapter in *Privacy Protection for E-Services*, edited by G. Yee, Idea Group, Inc., 2006.

[6] S. Kenny and L. Korba, "Adapting Digital Rights Management to Privacy Rights Management", *Computers & Security*, Vol. 21, No. 7, November 2002, pp. 648-664.

[7] G. Yee, L. Korba, "Privacy Policy Compliance for Web Services", *Proceedings, 2004 IEEE International Conference on Web Services (ICWS 2004)*, San Diego, California, July 6-9, 2004.

[8] Treasury Board of Canada, "The Privacy Impact Assessment Guidelines: A Framework to Manage Privacy Risk", available as of May 6, 2006, from: http://www.tbs-sct.gc.ca/pgol-pged/piatp-pfefvp/course1/mod2/mod2-5_e.asp

[9] I. Goldberg, D. Wagner, and E. Brewer, "Privacy-Enhancing Technologies for the Internet", *IEEE COMPCON'97,* 1997*,* pp. 103-109.

[10] G. Yee, L. Korba, "Semi-Automatic Derivation and Use of Personal Privacy Policies in E-Business", *International Journal of E-Business Research*, Vol. 1, No. 1, Idea Group Publishing, 2005, pp. 54-69.

[11] G. Yee, L. Korba, R. Song, L., "Legislative Bases for Personal Privacy Policy Specification", chapter in *Privacy Protection for E-Services*, edited by G. Yee, Idea Group, Inc., 2006.

[12] W3C; "A P3P Preference Exchange Language 1.0 (APPEL1.0)"; W3C Working Draft 15 April 2002, available as of February 28, 2007 at: http://www.w3.org/TR/P3P-preferences/

[13] G. Yee, L. Korba, "Bilateral E-services Negotiation Under Uncertainty", *Proceedings, The 2003 International Symposium on Applications and the Internet (SAINT2003)*, Orlando, Florida, Jan. 27-31, 2003.

[14] G. Yee, L. Korba, "The Negotiation of Privacy Policies in Distance Education", *Proceedings, 14th IRMA International Conference*, Philadelphia, Pennsylvania, May 18-21, 2003.

[15] J.I. Hong, J.D. Ng, S. Lederer, J.A. Landay, "Privacy Risk Models for Designing Privacy-Sensitive Ubiquitous Computing Systems", Proceedings, 2004 Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, Cambridge, MA, USA, 2004, pp. 91-100.

[16] P.A. Karger, "Privacy and Security Threat Analysis of the Federal Employee Personal Identity Verification (PIV) Program", *Proceedings of the Second Symposium on Usable Privacy and Security*, Pittsburgh, Pennsylvania, USA, 2006, pp. 114-121.

[17] C. Salter, O. Sami Saydjari, B. Schneier, J. Wallner, "Towards a Secure System Engineering Methodology", *Proceedings of New Security Paradigms Workshop*, Sept. 1998.

[18] Wikipedia, "Data Flow Diagram", available at: http://en.wikipedia.org/wiki/Data_flow_diagram (visited July 1, 2006).

[1] NRC Paper No.: NRC 49303