**Semi-Automated Derivation of Personal Privacy Policies**
Yee, George; Korba, Larry

National Research
Council Canada

Institute for
Information Technology

Conseil national
de recherches Canada

Institut de technologie
de l'information

NRC · CNRC

# *Semi-Automated Derivation of Personal Privacy Policies **

Yee, G., and Korba, L.

May 2004

* published in Proceedings of the 2004 Information Resources Management Association International Conference (IRMA 2004). New Orleans, Louisiana, USA. May 23-26, 2004. NRC 46539.

Canada

# Semi-Automated Derivation of Personal Privacy Policies[1]

**George Yee and Larry Korba**

Institute for Information Technology
National Research Council Canada
1200 Montreal Road, Building M-50
Ottawa, Ontario, Canada K1A 0R6
{George.Yee, Larry.Korba}@nrc-cnrc.gc.ca

## ABSTRACT

Growth of the Internet has been accompanied by growth of Internet e-services (e.g. e-commerce, e-health). This proliferation of e-services has in turn fueled the need to protect the personal privacy of e-service users. We advocate a privacy policy negotiation approach to protecting personal privacy [1,2]. However, it is evident that the specification of a personal privacy policy must be as easy as possible for the consumer. In this paper, we define the content of personal privacy policies using privacy principles that have been enacted into legislation. We then present two semi-automated approaches for the derivation of personal privacy policies. The first approach makes use of common privacy rules obtained through community consensus. This consensus can be obtained from research and/or surveys. The second approach makes use of existing privacy policies in a peer-to-peer community.

## 1 INTRODUCTION

The rapid growth of the Internet has been accompanied by an avalanche of e-services targeting consumers. E-services are available for banking, shopping, learning, healthcare, and Government Online. However, each of these services requires a consumer's personal information in one form or another. This leads to concerns over privacy.

In order for e-services to be successful, privacy must be protected. An effective and flexible way of protecting privacy is to manage it using privacy policies. Where the privacy policy of an e-service consumer conflicts with the privacy policy of an e-service provider, we have advocated a negotiations approach to resolve the conflict [1,2]. Providers in general have sufficient resources to come up with their privacy policies. Consumers, on the other hand, need help in formulating privacy policies. In addition, the creation of such policies needs to be as easy as possible or consumers would simply avoid using them. Existing privacy specification languages such as P3P and APPEL [3,4] (XML-based) are far too complicated for the average Internet user to understand. Understanding or changing a privacy policy expressed in these languages effectively requires knowing how to program. What is needed is an easy, semi-automated way of deriving a personal privacy policy. In this paper, we present two semi-automated approaches for obtaining personal privacy policies for consumers.

We have not been able to find any other authors who have written on the derivation of personal privacy policies. However, there are many references to privacy policies for e-commerce web sites. These are not relevant for this work since they are based on the "take it or leave it" P3P view of privacy protection. We believe that privacy protection must meet the wishes of the consumer.

### 1.1 Privacy Legislation and Directives

In Canada, privacy legislation is enacted in the *Personal Information and Electronic Documents Act* [5] and is based on the Canadian Standards Association's *Model Code for the Protection of Personal Information* [6] recognized as a national standard in 1996. This Code consists of ten Privacy Principles [6] that for convenience, we label as CSAPP. Data privacy in the European Union is governed by a very comprehensive set of regulations called the Data Protection Directive [7]. In the United States, privacy protection is achieved through a patchwork of legislation at the federal and state levels. However, privacy has been recognized as a constitutional right and there exists a highly developed system of privacy protection under tort law for the past century [8].

### 1.2 Privacy Management Model

As mentioned above, an effective and flexible way to protect privacy is to manage it using privacy policies. A provider has a privacy policy stating what personally identifiable information (PII) or private information (we use these terms interchangeably) it requires from a consumer and how the information will be used. A consumer has a privacy policy stating what PII the consumer is willing to share, with whom it may be shared, and under what circumstances it may be shared. An entity that is both a provider and a consumer has separate privacy policies for these two roles. A privacy policy is attached to a software agent that acts for a consumer or a provider. Prior to the activation of a particular service, the agent for the consumer and the agent for the provider undergo a privacy policy exchange, in which the policies are examined for compatibility. The service is only activated if the policies are compatible (i.e. there are no conflicts), in which case we say that there is a "match" between the two policies. In this paper, it is not necessary to consider the details of service operation. However, the provider is expected to comply with the consumer's privacy policy if service is initiated.

Section 2 examines the specification of privacy policies by identifying some attributes of private information collection. Section 3 shows how personal privacy policies can be semi-automatically generated. Section 4 gives conclusions and future research.

## 2 THE SPECIFICATION OF PRIVACY POLICY CONTENT

### 2.1 Requirements from Privacy Principles

In this section, we identify some attributes of private information collection using CSAPP as a guide. We use CSAPP because it is representative of privacy legislation in other countries and has withstood the test of time, originating from 1996. We will then apply these attributes to the specification of privacy policy contents. Tables 1 shows CSAPP.

We interpret "organization" as "provider" and "individual" as "consumer". In the following, we use CSAPP.n to denote Principle n of CSAPP. Principle CSAPP.2 implies that there could be different providers requesting the information, thus implying a *collector* attribute. Since on a provider policy, the collector is always the provider, *collector* is only used in consumer privacy policies. Principle CSAPP.4 implies that there is a *what* attribute, i.e. what private information is being collected. Principles CSAPP.2, CSAPP.4, and CSAPP.5 state that there are *purposes* for which the private information is being collected. Principles CSAPP.3, CSAPP.5 and CSAPP.9 imply that the private information can be disclosed to other parties, giving a *disclose-to* attribute. Principle CSAPP.5 implies a *retention time* attribute for the retention of private information. Thus, from the CSAPP we derive 5 attributes of private information collection: *collector*, *what*, *purposes*, *retention time*, and *disclose-to*.

**Table 1.** CSAPP - The Ten Privacy Principles from the Canadian Standards Association

| ***Principle*** | ***Description*** |
|---|---|
| **1. Accountability** | An organization is responsible for personal information under its control and shall designate an individual or individuals accountable for the organization's compliance with the privacy principles. |
| **2. Identifying Purposes** | The purposes for which personal information is collected shall be identified by the organization at or before the time the information is collected. |
| **3. Consent** | The knowledge and consent of the individual are required for the collection, use or disclosure of personal information, except when inappropriate. |
| **4. Limiting Collection** | The collection of personal information shall be limited to that which is necessary for the purposes identified by the organization. Information shall be collected by fair and lawful means. |
| **5. Limiting Use, Disclosure, and Retention** | Personal information shall not be used or disclosed for purposes other than those for which it was collected, except with the consent of the individual or as required by the law. In addition, personal information shall be retained only as long as necessary for fulfillment of those purposes. |
| **6. Accuracy** | Personal information shall be as accurate, complete, and up-to-date as is necessary for the purposes for which it is to be used. |
| **7. Safeguards** | Security safeguards appropriate to the sensitivity of the information shall be used to protect personal information. |
| **8. Openness** | An organization shall make readily available to individuals specific information about its policies and practices relating to the management of personal information. |
| **9. Individual Access** | Upon request, an individual shall be informed of the existence, use and disclosure of his or her personal information and shall be given access to that information. An individual shall be able to challenge the accuracy and completeness of the information and have it amended as appropriate. |
| **10. Challenging Compliance** | An individual shall be able to address a challenge concerning compliance with the above principles to the designated individual or individuals accountable for the organization's compliance. |

The Privacy Principles also prescribe certain operational requirements that must be satisfied between provider and consumer, such as identifying purpose and consent. Our service model and the exchange of privacy policies automatically satisfy some of these requirements, namely Principles CSAPP.2, CSAPP.3, and CSAPP.8. The satisfaction of the remaining operational requirements depends on compliance mechanisms (Principles CSAPP.1, CSAPP.4, CSAPP.5, CSAPP.6, CSAPP.9, and CSAPP.10) and security mechanisms (Principle CSAPP.7).

## 2.2 Privacy Policy Specification

Based on the above exploration, the contents of a privacy policy should, for each item of PII, identify a) *collector* - who wishes to collect the information (for consumer policies only), b) *what* - the nature of the information, c) *purposes* - the purposes for which the information is being collected, d) *retention time* – the amount of time for the provider to keep the information, and e) *disclose-to* – the parties to whom the information will be disclosed. Figure 1 gives 3 examples of consumer personal privacy policies for use with an e-learning provider, an online bookseller, and an online medical help clinic. The first item in a policy indicates the type of online service for which the policy will be used. Since a privacy policy may change over time, we have a *valid* field to hold the time period during which the policy is valid. For a consumer policy, the proxy field holds the name of the proxy if a proxy is employed to provide the information. Otherwise, this field has the default value of "no". For a provider policy, the proxy field has a default value of "yes" indicating that the consumer can use a proxy to provide the information. Otherwise, this field has the value "no".

| | | |
|---|---|---|
| Policy Use: *E-learning*<br>Owner: *Alice Consumer*<br>Proxy: *No*<br>Valid: *unlimited* | Policy Use: *Bookseller*<br>*Owner:* Alice Consumer<br>Proxy: *No*<br>Valid: *June 2003* | Policy Use: *Medical Help*<br>*Owner:* Alice Consumer<br>Proxy: *No*<br>Valid: *July 2003* |
| *Collector:* Any<br>*What:* name, address, tel<br>*Purposes:* identification<br>*Retention Time:* unlimited<br>*Disclose-To*: none<br><br>*Collector:* Any<br>*What:* Course Marks<br>*Purposes:* Records<br>*Retention Time:* 2 years<br>*Disclose-To*: none | *Collector:* Any<br>*What:* name, address, tel<br>*Purposes:* identification<br>*Retention Time:* unlimited<br>*Disclose-To*: none | *Collector:* Any<br>*What:* name, address, tel<br>*Purposes:* contact<br>*Retention Time:* unlimited<br>*Disclose-To*: pharmacy<br><br>*Collector:* Dr. A. Smith<br>*What:* medical condition<br>*Purposes:* treatment<br>*Retention Time:* unlimited<br>*Disclose-To*: pharmacy |

**Figure 1.** Example Consumer Personal Privacy Policies

A personal privacy policy thus consists of "header" information (policy use, owner, proxy, valid) together with 5-tuples or privacy rules

*<collector, what, purposes, retention time, disclose-to>*

where each 5-tuple or rule represents an item of private information and the conditions under which the information may be shared. A personal privacy policy therefore consists of a header plus one or more privacy rules.

# 3 SEMI-AUTOMATED DERIVATION OF PERSONAL PRIVACY POLICIES

A semi-automated derivation of a personal privacy policy is the use of mechanisms (described below) that may be semi-automated to obtain a set of privacy rules for a particular use (see Section 2.2). We present two approaches for such derivations. The first approach relies on third party surveys of user perceptions of data privacy. The second approach is based on retrieval from a community of peers.

### Derivation Through Third Party Surveys

(a) A policy provider makes use of third party surveys performed on a regular basis as well as those published in research literature to obtain user perceptions of the level of privacy for various sets of PII separated according to their uses. This gives a sensitivity or range of privacy levels for different PII in different situations.

(b) Corresponding to a provider's privacy policy (which specifies what PII is required), the policy provider or a software application constructs and ranks the privacy rules for each use using the PII in (a), according to their sensitivity levels, such that the rules are selectable by a single value privacy level from a "privacy slider". The outcome of this process is a set of consumer privacy rules, ranked by PII sensitivity, for different providers. The policy provider would express the resulting privacy rules in a policy language such as APPEL. There are different ways to do this ranking. One way is to assign a privacy rule the median of its sensitivity range as its privacy level (illustrated below).

(c) Consumers obtain online from the policy provider the privacy rules that make up whole policies. They do this by specifying the use for the rules, the provider for which a consumer privacy policy is required, and the level of privacy required using the privacy slider. The consumer then completes each privacy policy by adding the rest of the header information (i.e. *owner*, *proxy*, *valid – use* is already there). This can be done through a human computer interface that shelters the user from the complexity of the policy language. In this way, large populations of consumers may quickly obtain privacy policies for many service providers that reflect the privacy sensitivities of the communities surveyed.

(d) Consumers may interactively adapt their privacy policies for different service providers based on their current policies, the sensitivities of the privacy rules, and the policy of the service provider. This assumes the availability of an easy to understand interface for the interaction as well as software to reflect the changes back into the policy language.

This approach requires trust in the policy provider. Effectively the policy provider becomes a trusted third party. A certification process for the policy provider is probably required. For instance, in Canada, the offices for the provincial and federal privacy commissioners could be this certification body. They could also provide this policy creation service.

A notification process should be used during the policy exchange phase between a consumer and a service provider to let the consumer know when "sensitive data" is exchanged. The degree of consumer sensitivity to different PII for different situations would also be available from the policy provider. This information could be updated regularly by the policy provider, or updated through a short online survey. The sensitivities would either modulate the base policy or set a trigger level for user warnings during policy exchange. During the warning, the user is presented with options that may allow the "degradation" or shoring up of the privacy policy. Figure 2 illustrates this approach.

**Example:**
Suppose the item of PII for which we wish to derive a privacy rule is "course marks retention time" from the e-learning privacy policy in Figure 1.

Then the above steps are implemented as follows:

a) The third party survey generates the following results for course marks retention time (the higher the privacy level, the higher the privacy; the highest level is 5, the lowest level is 1).

| PII | Privacy Level |
|---|---|
| course marks retention time 6 months | 3 |
| course marks retention time 6 months | 4 |
| course marks retention time 6 months | 4 |
| course marks retention time 6 months | 5 |
| course marks retention time 12 months | 1 |
| course marks retention time 12 months | 1 |
| course marks retention time 12 months | 2 |
| course marks retention time 12 months | 3 |

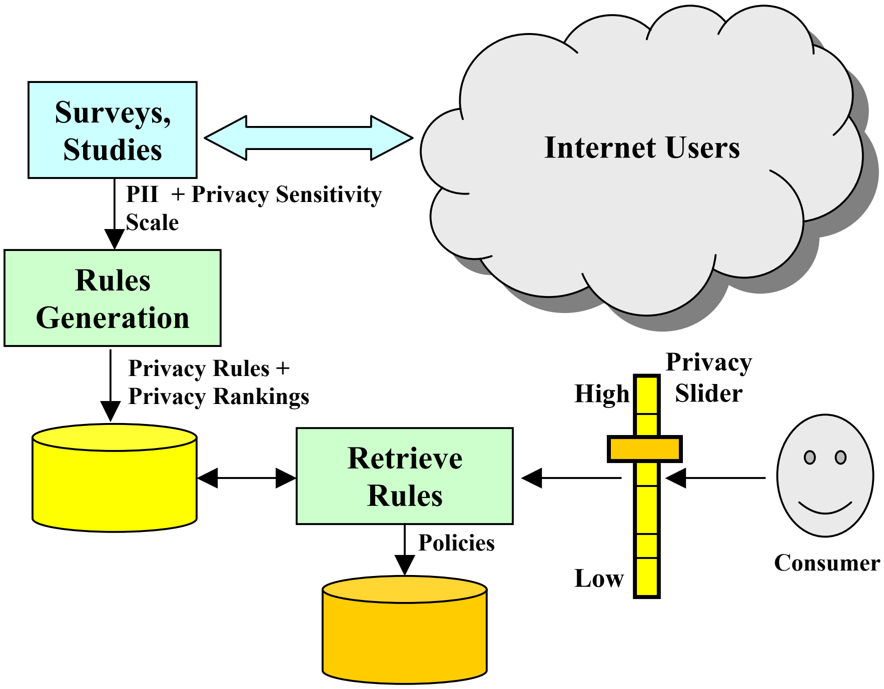
Note that the other parameters in a privacy rule may change too, not just retention time. We change retention time only to keep the example simple. Actually, each different combination of parameters represents a different privacy level. Also, the higher the number of months that the marks are retained, the lower the privacy level. The different privacy levels obtained for the same PII constitute one part of the privacy sensitivity scale.

(b) In this step, the policy provider constructs privacy rules from the PII in (a) and ranks them using the median value from the corresponding sensitivity range. Thus for the 4 course mark retention times of 6 months, the lowest value is 3, the highest value is 5, and the median is 4. Therefore the rule < any, course marks, records, 6 months, none > is ranked with privacy level 4. Similarly, the rule < any, course marks, records, 12 months, none > is ranked with privacy level 2.

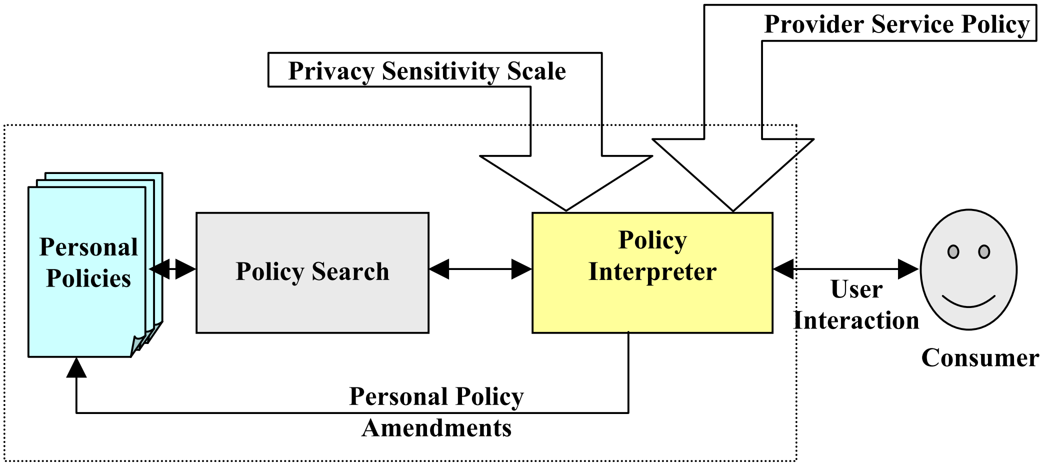
(c) To obtain his/her privacy rules, the consumer specifies the use as e-learning and a privacy slider value of 4 (for example). He/she then obtains the rule

< any, course marks, records, 6 months, none >

and proceeds to complete the policy by adding header values for *owner*, *proxy*, and *valid*.



a) Derivation of personal privacy policies from surveys



b) Adapting personal privacy policies to the service provider

**Figure 2.** Derivation of personal privacy policies through surveys

## Retrieval from a Community of Peers

This approach assumes an existing community of peers already possessing specific use privacy policies with rules according to desired levels of privacy. A new consumer joining the community searches for personal privacy rules or whole personal privacy policies (sets of rules). The existing personal privacy policies may have been derived using the third party surveys as above. The privacy policy rules are each stored along with its privacy level so that it may be selected according to this level. Where a rule has been adapted or modified by the owner, it is the owner's responsibility to ensure that the slider privacy value of the modified rule is consistent with the privacy sensitivity scale from surveys.

(a) All online users are peers and everyone has a privacy slider. The new consumer broadcasts a request for privacy rules to the community, specifying use and slider value. This is essentially a peer-to-peer search over all peers.
(b) The community responds by forwarding matching (in terms of use and slider value) rules to the consumer. This match may be a fuzzy match as well.
(c) The consumer compares the rules and selects them according to use, popularity (those that are from the greater number of peers), and best fit in terms of privacy. After obtaining the rules, the consumer completes the privacy policies by completing the headers as in the above derivation from surveys approach.
(d) Consumers may adapt their privacy policies for different service providers as in the derivation by surveys approach.

There is a challenge here regarding how to carry out this approach in a timely fashion. Efficient peer-to-peer search techniques will collect the policies in a timely manner, but the amount of information collected by the requester may be quite large. As well, since the various policies collected will probably differ from each other, the requestor will have to compare them to determine which one to select. Quick comparison so as to reduce the amount of data collected would be through a peer-to-peer policy search that employs a policy hash array, containing hashed values for different portions of the policy for more rapid comparison.

## 4 CONCLUSIONS AND FUTURE RESEARCH

The protection of personal privacy is paramount if e-services are to be successful. A privacy policy approach to privacy protection seems best. However, for this approach to work, consumers must be able to derive their personal privacy policies easily. In order to describe semi-automated approaches to derive personal privacy policies, we first defined the content of a personal privacy policy using the Canadian Privacy Principles. We then presented two semi-automated approaches for obtaining the policies: one based on third party surveys of consumer perceptions of privacy, the other based on retrieval from a peer community. Both approaches reflect the privacy sensitivities of the community, giving the consumer confidence that his/her privacy preferences are interpreted with the best information available.

Clearly, the notion of a trusted third party as a personal policy provider may be controversial to some. Any error made by the policy provider could affect PII for many hundreds or thousands of people. Having privacy commissioners' offices take responsibility for this process seems to be a natural fit, given their mandate as privacy watchdog for the consumer. However, the process would have a cost. Costs might be recovered via micro-charges to the consumer, or the service provider for the policies provided. Aggregated information from the PII surveys might be sold to service providers.

An interesting aspect to this approach not discussed above, is the prospect of continuously updating PII related information and privacy policies, based upon policy updates fed back to the policy provider from consumers. Users will be changing their policies to suite their desired or perceived needs over time, when interacting with different service providers and for different services. The policy provider could gather updates made to policies dynamically, and analyze them to adjust the typical policies it distributes to better reflect the experiences of the population with different service providers.

For future research, we plan to investigate other ways of deriving privacy policies easily. As well, we plan to construct simulations of the approaches presented in this paper to look for and resolve any scalability/performance issues.

# 5 REFERENCES

[1] 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.

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

[3] W3C, “The Platform for Privacy Preferences”, http://www.w3.org/P3P/

[4] W3C, “A P3P Preference Exchange Language 1.0 (APPEL1.0)”, W3C Working Draft 15 April 2002, http://www.w3.org/TR/P3P-preferences/

[5] Department of Justice, Privacy Provisions Highlights,
http://canada.justice.gc.ca/en/news/nr/1998/attback2.html

[6] Canadian Standards Association, “Model Code for the Protection of Personal Information”, retrieved Sept. 5, 2003 from:
http://www.csa.ca/standards/privacy/code/Default.asp?articleID=5286&language=English

[7] European Union, “Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data”, unofficial text retrieved Sept. 5, 2003 from: http://aspe.hhs.gov/datacncl/eudirect.htm

[8] Industry Canada, “Privacy and the Information Highway, Regulatory Options for Canada”, chapter 6, retrieved Sept. 5, 2003 from: http://strategis.ic.gc.ca/SSG/ca00257e.html#6

[1] NRC Paper Number: NRC 46539