Semi-Automated Seeding of Personal Privacy Policies in E-Services

George Yee
National Research Council Canada, Canada

Larry Korba
National Research Council Canada, Canada

INTRODUCTION

The rapid growth of the Internet has been accompanied by a proliferation of e-services targeting consumers. E-services are available for banking, shopping, learning, government online, and healthcare. However, each of these services requires a consumer's personally identifiable information (PII) in one form or another. This leads to concerns over privacy.

In order for e-services to be successful, privacy must be protected (Ackerman, Cranor, & Reagle, 1999). An effective and flexible way of handling privacy is management via privacy policies. In this approach, a consumer of an e-service has a personal privacy policy that describes what private information the consumer is willing to give up to the e-service, with which parties the provider of the e-service may share the private information, and how long the private information may be kept by the provider. The provider likewise has a provider privacy policy describing similar privacy constraints as in the consumer's policy, but from the viewpoint of the provider, (i.e., the nature of the private information and the disclosure/retention requirements that are needed by the e-service). Before the consumer engages the e-service, the provider's privacy policy must match with the consumer's privacy policy. In this way, the consumer's privacy is protected, assuming that the provider complies with the consumer's privacy policy. Note that policy compliance is outside the scope of this work but see Yee and Korba (July, 2004).

Initial attempts at conserving consumer privacy for e-services over the last few years have focused on the use of Web site privacy policies that state the privacy rules or preferences of the Web site or service provider. Some of these policies are merely statements in plain English and it is up to the consumer to read it. This has the drawback that very few consumers take the trouble to read it. Even when they do take the time to look at it, online privacy policies have been far too complicated for consumers to understand and suffer from other deficiencies (Lichtenstein, Swatman, & Babu, 2003; Jensen & Potts, 2004). Still other privacy policies are specified using P3P (W3C) that allows a consumer's browser to automatically check the privacy policy via a browser plug-in. This, of course, is better than plain English policies but a major drawback is that it is a "take-it-or-leave-it" approach. There is no recourse for the consumer who has a conflict with the Web site's P3P policy, except to try another Web site. In this case, we have advocated a negotiations approach to resolve the conflict (Yee & Korba, Jan., May, 2003). However, this requires a machine-processable personal privacy policy for the consumer.

We assume that providers in general have sufficient resources to generate their privacy policies. Certainly, the literature is full of works relating to enterprise privacy policies and models (e.g., Barth & Mitchell, 2005; Karjoth & Schunter 2002). Consumers, on the other hand, need help in formulating machine-processable 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, APPEL (W3C; W3C, 2002), and EPAL (IBM) 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. Moreover, most of these languages suffer from inadequate expressiveness (Stufflebeam, Anton, He, & Jain, 2004). What is needed is an easy, semi-automated way of seeding a personal privacy policy with a consumer's privacy preferences. In this work, we present two semi-automated approaches for obtaining consumer personal privacy policies for e-services through seeding. This article is based on our work in Yee and Korba (2004).

The section "Background" examines related work and the content of personal privacy policies. The section "Semi-Automated Seeding of Personal Privacy Policies" shows how personal privacy policies can be semi-auto-

matically seeded or generated. The section "Future Trends" identifies some of the developments we see in this area over the next few years. We end with "Conclusion".

BACKGROUND

We have been able to find only two other authors who have written on the derivation of personal privacy policies. Dreyer and Olivier (1998) describe a tool called the "Privacy Workbench" for creating and analyzing privacy policies. However, it is not clear from their article how one comes up with the privacy policy in the first place, as it seems to just appear followed by a description of how the tool can perform conflict analysis. It is a model-based rules inference approach for validating an existing privacy policy. More importantly, Privacy Workbench is a tool for a programmer, as it is far too complex for the average consumer to understand and use. Snekkenes (Snekkenes, 2001) wrote about the derivation of personal location privacy policies for use with a location-based service, (e.g., E911 emergency location service in the United States). Snekkenes' view is that "individuals should be equipped with tools to become in the position to formulate their own personal location privacy policies". This author provided concepts as well as fragments of a language for formulating personal location privacy policies. Unfortunately, the language presented can only be understood by programmers and not the average consumer. Our approaches for generating personal privacy policies are not model-driven or service specific and have been designed for ease-of-use by the average consumer.

Privacy Legislation and Directives

Before we can consider how to seed a personal privacy policy, we need to know what such a policy should contain in terms of privacy provisions. We use privacy legislation to obtain what must be specified in a personal privacy policy. Therefore, this gives a minimum policy in the sense that all elements required by law have been specified, but additional provisions can be included at the discretion of the consumer.

In Canada, privacy legislation is enacted in the *Personal Information Protection and Electronic Documents Act* (Department of Justice, 2005; Government of Canada) and is based on the Canadian Standards Association's Model Code for the Protection of Personal Information *(Canadian Standards Association)* recognized as a national standard in 1996. This Code consists of ten Privacy Principles *(Canadian Standards Association)* that for convenience, we label as CSAPP. Data privacy in the European Union is governed by a comprehensive set of regulations called the Data Protection Directive (Euro-

Figure 1. Example consumer personal privacy policies

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