User Agents for Matching Privacy Policies with User Preferences

Karin Bernsmed, Åsmund Ahlmann Nyre, and Martin Gilje Jaatun

***Abstract*—Privacy policies are commonly used by service providers to state how personal data obtained from users will be handled. However, the complexity and sheer length of such policies make them incomprehensible to the common web user. This paper surveys user agents that automatically fetch and compare privacy policies with privacy preferences, in order to help the end-user understand the implications of personal data disclosure. We discuss why previous efforts have experienced only moderate success, identify the main areas of improvement and point out directions for future work.**

***Index Terms*—P3P, PETs, privacy policies, privacy preferences, survey, user agents.**

I. Introduction

The future Internet will be more service oriented, have tighter social networks, and more ubiquitous communications. This entails increased communication and information sharing, requiring more stringent protection mechanisms to protect the privacy of its end-users. Privacy enhancing technologies (PETs) is a general term for technology that helps users protect and control their personal information and make informed decisions on when and what to share.

For PETs in general, attention has mainly been paid to hiding information or providing anonymity, whereas little has been done to mitigate the risks related to sharing personal data (by personal data, we mean any information that can be stored and associated with an identifiable person, such as name, e-mail address, digital identity, financial data and so on). The current lack of protection mechanisms is especially noticeable for online services, such as social networking and online shopping web sites, which are well-known for routinely collecting large amounts of personal data from their users.

While privacy policies are now commonplace for most websites, users are daunted by their complexity and length. Studies have shown that most Internet users claim to be concerned about their privacy [1], but when push comes to shove this has little effect on their actual behavior. This phenomenon is often referred to as “the privacy paradox” and is attributed to the difficulty of understanding privacy policies [2].

To enable the users to understand the implications of personal data disclosure and to make well-informed decisions, there is a need for tools informing them about the privacy policies of the online services providers [3][4]. Such tools should be able to match the user's privacy preferences against the service provider’s privacy policies, and visualize the conformance in an intuitive and easily understandable manner. This paper presents a survey of existing user agents for automatic privacy policy and privacy preferences matching. The purpose is to draw attention to the most promising research ventures, discuss their advantages and shortcomings, and to identify areas for further work.

The paper is organized as follows; Section II starts by explaining the concept of personal data management and control. This section also explains the basic model of a user agent for privacy preferences and policy matching. Section III discusses the support and limitations of today's browsers, regarding management of user privacy preferences and privacy policy interpretation. Sections IV and V give a short description and comparison of existing user agents. In Section VI we discuss our findings, before identifying areas for improvement in Section VII. Finally, in Section VIII we offer our concluding remarks.

II. Personal Data Management

Privacy is a fundamental human right, and a cornerstone of privacy is the individual's right to right to decide what to reveal about oneself [5]. In an ICT context, privacy is all about controlling personal information. While this may sometimes be accomplished through anonymity, many of the services we use on the Internet today are meaningless without some level of shared personal data. Imagine how Facebook would be with anonymous users, or how you would buy a book from an online bookstore without specifying a delivery address. Sharing personal data online is therefore often necessary, and in many cases also desirable.

A fundamental problem of sharing personal data when interacting with online services is the limited information available that the end-users can use to determine whether the service in question can be trusted to treat their personal data properly. Privacy policies have become the main instrument for service providers to explain how users' personal data are collected, used, disclosed, and managed. Unfortunately, due to their complexity, difficult language and sheer length, users tend to neither read nor understand the policies prior to acceptance [2]. This was one of the main motivations for launching the Platform for Privacy Preference Project (P3P) [6]; to make it easier for users to understand privacy policies and make well-informed decisions on how to interact with services that collect personal data.

Manuscript received February 25, 2012; revised March 31, 2012.

All the authors are with SINTEF Information and Communication Technology, N-7465 Trondheim, Norway (e-mail: {karin.bernsmed, asmund.a.nyre, martin.g.jaatun}@ sintef.no).

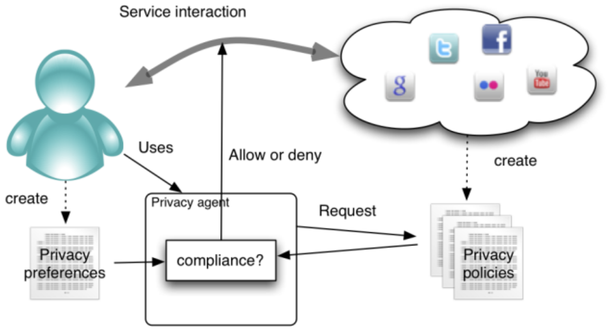

Fig. 1. The basic model for automatic matching of an end-user's privacy preferences with a service provider's privacy policy.

P3P provides a policy markup language to enable automatic processing and assessments of policies. Combined with APPEL (A P3P Preference Exchange Language) [7], a P3P agent is capable of matching privacy policies to stated user preferences. This interaction model (Figure 1) is nowadays considered the standard model, as most of the web-based privacy preferences and policy matching agents proposed since P3P follow it. There are three main parts in this model; the user preferences, privacy policies and the user agent. The model lets users and service providers specify their privacy preferences and policies, respectively, in machine-understandable policy languages. Whenever a user makes a service request, a software agent can retrieve the privacy policy of the service and compare this to the user preferences. Depending on the capabilities of the agent and languages used for specification, the agent may enter a negotiating phase, initiate mitigating measures, block access to the service or simply issue a warning that the policy does not match the user's preferences. In Figure 1 this is illustrated by an allow/deny decision.

Policy languages are thus central to the above-mentioned model of web-based user agents, and many other security technologies for that matter. If privacy policies are written in natural language, as is common today, the user agent is incapable of determining whether a policy matches the user preferences. In addition to the already mentioned P3P and APPEL languages, several other machine-readable policy languages for privacy policies and preferences have been proposed. Examples are Ponder [8], which is a language for specifying management and security policies for distributed systems, and Rei [9], which is designed and distributed for dynamic environments like pervasive systems. Duma et al. [25] provide a comparison and review of a selection of the proposed policy languages with respect to privacy. In their paper they demonstrate how a user agent's capability is limited by the expressiveness of the policy language in use. They conclude that the more expressive the language is, the more features can be provided by the agent.

III. Browser Support

The web browser is by far the most common client interface for end-user interaction with service providers, and a browser may very well adopt the role of the user agent illustrated in the basic model in Figure 1. However, as far as we are aware, all browsers of today contain very limited support for matching user personal data privacy preferences against a service provider's privacy policy. One of the few automatic personal data collection mechanisms that the user can control in the browser settings is cookie management.

Since the launch of P3P in early 2000, Microsoft in particular has been an avid supporter of the technology, pushing the use of P3P in their software portfolio. Internet Explorer (IE) version 6 and its successors (IE7 and IE8) provide the ability to display P3P privacy policies and to compare the policies with the user's privacy preferences settings. However, the support is quite limited, in that only a subset of the specification has been implemented in the conformance checking; namely the compact P3P policy, which covers the use of cookies. IE will not alert the user if the web site violates the privacy preferences regarding any other personal information, such as user-provided data. IE uses the P3P compact policy, which is transferred in the HTTP headers, to make cookie blocking decisions. If the cookie policy of the service provider does not match the user's preferences, IE will display an eye covered by a do-not-enter sign in the browser frame. P3P user agents based on compact policies were also implemented as a part of Mozilla Firefox and Netscape web browsers in early 2000, however the functionality has since then been removed.

Even though most browsers do not support automatic user preferences and privacy policy matching as core functionality, this feature can in many cases be implemented as an add-on. There are several examples of how a user agent can be implemented with this technique, some of which will be presented in the next section.

IV. User Agents Overview

In this section we present the most promising research efforts on user agents that are capable of matching privacy policies and end-user privacy preferences. The approaches are then compared, and their advantages and shortcomings are discussed.

A. *AT and T Privacy Bird*

The ATandT Privacy Bird [11] was one of the first P3P user agents. It is still the most complete P3P tool currently available [12]. Its design was formed through the experience with four early prototype user agents that were developed in parallel with the creation of the P3P specification [2]. The ATandT Privacy Bird is implemented as a browser helper object. A graphical window allows the user to set up his/her privacy preferences, based on a subset of the P3P specification vocabulary. It is also possible to import privacy preferences written in APPEL. The ATandT Privacy Bird will then automatically retrieve privacy policies from service providers and compare these with the user's specified privacy preferences. When surfing the web, a bird icon in the browser title bar changes color and shape to indicate whether or not the web site is P3P enabled, and whether or not the P3P policy matches the user's privacy preferences. The user can click on the bird to view a summary of the P3P policy. If there is a mismatch between the user's privacy preferences and the privacy policy, the summary will explain where and why the mismatch occurred.

B. PIPWatch Toolbar

The PIPWatch Toolbar [4] is a web browser add-on that helps users to interpret whether the privacy policies of the websites they visit comply with the Canadian private-sector privacy legislation - the Personal Information Protection and Electronic Documents Act (PIPEDA). Its appearance is similar to the ATandT Privacy Bird; when surfing the web the browser toolbar provides the user with privacy-related information of the website and a “'privacy beaver” changes color according to the current privacy risk.

In contrast to the ATandT Privacy Bird, the PIPWatch Toolbar is not based on any machine-readable privacy policy or preference language. Instead users are expected to contribute with privacy related information on the service provider by using functionality embedded in the toolbar. This includes filling out basic information on the website and to send email to privacy officers, asking them to fill out a questionnaire. The purpose of the questions is to find out to what degree the web sites complies with the PIPEDA. The responses from the privacy officers will then be stored on a central server and used by the toolbar to evaluate to what degree the site fulfils the users’ privacy expectations with respect to the Canadian privacy legislation.

C. Collaborative Privacy Management (CPM)

Kolter et al. [13] present a privacy management tool based on user collaboration. Its appearance is very similar to the general model presented in Figure 1; it consists of a privacy preference generator and a privacy agent. As an additional feature it also includes a data disclosure log. The privacy preference generator assists the user in specifying his own privacy preferences using a vocabulary derived from P3P (see [14] for a detailed description), while the privacy agent provides recommendations to the user whenever personal data are to be disclosed to a service. The data disclosure log is assumed to record all events where personal data are disclosed, in order to track the flow of personal data and react to repetitive data disclosure. This will prevent service providers from violating user privacy by requesting small pieces of personal data that individually adheres to the preferences of the user, but when aggregated constitute a privacy violation. Before providing recommendations to the user, the privacy agent consults all sources of information (preferences, policies and disclosure log). The degree of conformance between the user’s privacy preferences and the policy is indicated with a traffic-light symbol.

Kolter et al. note that a weakness of the P3P concept is that it requires support from services providers in creating compliant privacy policies. To overcome this dependency, the authors propose an online web community that can collaboratively create and share annotated privacy policies of sites and services through a Wiki-like community portal. Whenever users request a service, the tool will request the privacy policy from the community portal rather than the website in question. With a setup similar to that of Wikipedia, revision control would allow for changes to be tracked as new revisions of privacy policies are published. The community portal will also allow experts to explain privacy policies to novice users, share preferences and rate service providers' adherence to their policies.

D. Privacy and Identity Management for Europe (PRIME)

The purpose of the EU project PRIME [15] was to develop privacy enhancing identity management systems in order to protect and improve end-user privacy. PRIME is based on the principle that design must start from maximum privacy [16]. The PRIME architecture lets users interact with service providers through their web browsers running on top of the PRIME software. Service providers use PRIME middleware for all privacy-sensitive interactions. The architecture handles both privacy preferences and policies for end-users and service providers, respectively, and is designed to enforce these in an automated way as far as possible.

PRIME points out that “informative and intuitive user interfaces are crucial for effective privacy protection” [17]. The guiding principle is to let the users be in control of their personal data. The end-user interface is therefore an important part of the PRIME architecture. It manages the user data and credentials, controls interactions with other parties, and tracks the user data once it has been disclosed. In addition, it interacts with the user when browsing the web by displaying privacy policies and allowing the user to decide on what personal data to release. Several prototype user interfaces have been implemented and tested.

The PRIME user interface aims to help novice users minimize their personal data disclosure, which is achieved through the use of “PrivPrefs”. PrivPrefs are predefined privacy preferences that can be used on the fly, to avoid forcing the user to specify privacy preferences in advance. For example, if an online web service requests information from a user, the user can choose to use an anonymous PrivPref to avoid releasing any personal data. The PRIME user interface also contains a tracking function, which makes it possible for end-users to check what personal data that has been disclosed, when, to whom and how the data has been processed.

A special feature within the PRIME architecture is the bundling of personal data management with pseudonyms by using a “TownMap” metaphor. This means that the user can specify different disclosure preferences for different data types by using areas on a street map that visualizes the privacy preferences settings. The map is also used to show users how personal data transfers take place. In the TownMap-based user interface, the PRIME project implements a novel approach to let users express consent to personal data disclosure, which is called Drag-And-Drop Agreements (DADAs). The purpose is to avoid the automated behavior that often follows when users are forced to click through a number of confirmation boxes [17].

When releasing personal data, PRIME will match the user's privacy preferences with the service provider's privacy policy. The user can choose to either disclose data automatically if all conditions are met (e.g. by using PrivPrefs), or to be notified by a confirmation box. In PRIME the user is able to negotiate the amount of personal data disclosure to service providers [18]. PRIME has defined its own policy language for access control, data release and data handling.

E. Integrated Privacy View (IPV)

The Integrated Privacy View (IPV) [19] is a tool designed to match user privacy preferences and service provider privacy policies on a fine-grained level, something that is not possible with P3P in its original form (because it can only operate on a page level). The tool is based on an extension of P3P and APPEL. More specifically, IPV links a particular P3P policy statement to a particular input field on a web page. When the user surfs the web, and arrives at a page that contains input fields or forms, IPV will insert icons beside each field to indicate whether or not the personal data collected in that particular field will be used in a way that conforms to the user's stated privacy preferences.

To specify a P3P policy statement, IPV uses a concept called fine-grained policy anchors, which is defined as an extension to P3P. This means that a policy statement can be linked directly to an HTML element. IPV thereby defines a new attribute for the HTML element that specifies an input field; the *p3pdataelement*, which indicates that a specific policy statement is associated with this input field. The IPV privacy agent will then look for HTTP responses that contain the p3pdataelement attribute, perform privacy policy conformance evaluation, and modify the original page in order to inform the user of the result.

The IPV prototype tool is currently implemented as an HTTP proxy. As presented in [19], the tool does not include any user interface to specify privacy preferences, but assumes the existence of an APPEL file on the user's local machine.

F. Privacy Finder

Privacy Finder [20] is a privacy-enhanced online search engine, based on the technology developed in the ATandT Privacy Bird project [11]. Privacy Finder orders search results according to their P3P privacy policies. A "privacy meter" next to the search result indicates whether a P3P policy exists, and to what degree the policy corresponds with a list of preset user privacy preferences.

Clicking on the privacy meter will open a more detailed privacy policy report of the site. The report is based on the Privacy Nutrition Label concept presented in [21], which is an approach aimed at improving the visual presentation of privacy policies to end users. The report is constructed by combining symbols and color codes to illustrate how the user's personal data will be treated. The data is organized according to what type of information it represents, how it will be used by the service provider and whether it will be shared with 3rd party service providers. The terminology used in the policy report is derived from the P3P specification, but simplified in order to fit into a one-page summary.

V. Comparison of User Agents

Table 1 provides an overview over the six user agents presented in the previous section. The table summarizes questions and answers related to how privacy policies are implemented, whether the user can specify his/her own privacy preferences, and how policy-preferences matching is performed. More details are provided in the following.

A. How is the Privacy Policy Declared?

As can be seen from the table, there are different approaches to declare privacy policies. Three of the tools are based on P3P (ATandT Privacy Bird, IPV and Privacy Finder). PRIME has developed its own language, while the community approach of CPM uses a combination of privacy policies (textual as well as P3P) together with other privacy related information. PIPWatch does not use any formal language, but focuses on issues regarding the Canadian privacy legislation. A clear benefit of using P3P is the standardized way of expressing machine-readable policies, which later on facilitates automatic policy-preferences matching. However, as pointed out in [2], the usefulness of any P3P user agent is limited by P3P adoption.

B. Who Provides the Privacy Policy?

In order for the end-users to benefit from a tool that compares privacy policies with user preferences, there must be a large selection of applications that support the concept. If privacy policies exist only for a small fraction of all websites, using such a tool will tend to be meaningless from the users' perspective. Someone must therefore take on the responsibility to create and maintain privacy policies and practices. As can be seen in Table 1, the tools presented in this paper take on different approaches. Four of the tools (ATandT Privacy Bird, PRIME, IPV and PrivacyFinder) rely on the service providers to implement privacy policies according to a specified standard. However, the history of P3P has shown that one cannot expect service providers to voluntarily contribute to the widespread availability of accurate machine-readable privacy policies. As pointed out by Kolter et al. [13], a privacy architecture that accepts today's service landscape on the web may be more practical. An alternative approach is therefore to rely on assistance from the user community. Finally, PIPWatch requires support from both the user community and the service providers, without any automation, which in our opinion does not bode well for the scalability of the technology.

C. Can Users Specify Individual Privacy Preferences?

This is possible for all tools except IPV and Privacy Finder. These latter tools use preset preferences (declared in APPEL), however, adding such functionality should be straight-forward for both tools.

D. Is There a Policy-Preferences Conformance Matching?

All the tools share the ability of automatic policy-preferences conformance matching. Icons are apparently a popular way of notifying the user of possible mismatches and privacy risks. As can be seen in the table, the icons can appear in either the browser frame (ATandT Privacy Bird, PIPWatch, CPM) or directly in the body of the particular web page (IPV, Privacy Finder).

E. How is the Tool Implemented?

Even though the implementation details differ, the end-user will interact with the tool via a web browser in all the proposals. Regarding the implementation, it should be noted that all tools are research efforts. PIPWatch Toolbar, CPM and IPV are in an early stage. Neither of them, nor the PRIME user interface is accessible for public use. The ATandT Privacy Bird is available for download, but the software works only with IE 5.01/5.5/6.0 on Windows. Privacy Finder is up and running, and is at the time of writing operated by the CyLab Usable Privacy and Security Laboratory (CUPS) at Carnegie Mellon University.

To summarize, the ATandT Privacy Bird was one of the first P3P user agents that appeared and it was considered a very promising approach. Its main advantage was its potential for gaining a wide user acceptance, which most likely is due to its easily comprehensible user interface. Its potential has been shown in several user studies [2]. However, in practice its usefulness has turned out to be very limited, due to the low number of web sites with P3P privacy policies. The PIPWatch toolbar is adapted to the Canadian market, and is therefore in its current shape not very useful for the international World Wide Web. CPM and IPV are both promising attempts to improve some of the shortages of P3P; CPM by relaying on user community support and IPV by extending P3P. Privacy Finder is useful for anyone interested in interacting with web services with personal data privacy in mind. The PRIME architecture is a very ambitious approach. It represents a complete solution for privacy policy and user preference management (amongst other tings). However, in order to use PRIME, both users and service providers need to use the PRIME middleware. PRIME will therefore only become successful if a majority of the service providers accept the technology, and integrate the middleware into its components. As far as we are aware, the results from PRIME have not yet resulted in any concrete product.

VI. Discussion

In this section we discuss the implications of the different strategies that can be taken for automatic privacy policy and user preferences matching in more depth. We have noted that the design of most of the user agents reviewed in this paper spring from the P3P specification, either in model (as explained in Section II), or in language or protocol. Therefore, when discussing the strengths and weaknesses of the different approaches we will keep the P3P specification in mind.

A. *Privacy Policy Languages*

The functionality and flexibility of the user agents we have reviewed in this paper are to a great extent dependent on the underlying policy language for both privacy preferences and policies. Since the majority of the proposed user agents rely on P3P and APPEL for policy and preference declaration, respectively, they inherit many of the shortcomings identified for P3P. Hochheiser [22] identified the limited scope (only web sites), lack of limitations on personal data collection and the restricted vocabulary as the main areas of technical critique against the P3P specification. Particularly the precision of terms is said to be sub-optimal, which for example makes it difficult for service providers to express conformance to privacy legislation. While extensions may be developed to answer several of these shortcomings, this would conflict with the P3P intention of being simple and easy to use.

As indicated by Duma et al. [25], there exist several more expressive policy languages that do not suffer from the same limitations. However, this comes at the cost of adding complexity to the policy and preferences specification processes, as well as to user agent development. But, as with other seemingly complex systems, a solution may be to provide an abstraction layer on top that is tailored for specific contexts and user skills. This is to some extent what CPM and PRIME proposed, since they do specify extensions to P3P and a new policy language, respectively, and also use a privacy preference generator, which may be viewed as an abstraction of the preference language. However, they do not extend this to providing similar abstractions offering more functionality to advanced users.

B. *Dependence on Critical Mass of Adoption*

At its birth, P3P was praised by the Internet community and widely believed to be the keystone to resolving large privacy issues on the web. In practice, the specification has never been widely adopted. Even though there have been occasional reports on increasing rates of P3P adoption, especially among e-commerce sites [23], it seems like adoption of the specification is doomed to fail. Today, major sites like Google.com, Apple.com and CNN.com do not use P3P to summarize their privacy policies.

P3P's dependence on service providers to declare their privacy policy using the P3P policy language is one of the factors explaining why it never reached a critical mass of adoption. The problem is that with few P3P compliant web sites, the user demand for P3P user agents is low, which in turn reduces the usage of P3P-declared policies. And if no one is using the P3P policies, why should service providers bother providing them? Reaching the critical mass is what turns this around to a positive reinforcement, rather than a negative one.

To have third parties involved in translating the plain text privacy policy into a proper machine readable policy language is one way of reducing the dependency. Kolter et al. [13] and Clement et al. [4] suggest collaborative communities where individuals perform the translation. Another option may be to have professional translators selling annotated privacy policies as a service on the web, akin to how anti-virus software vendors currently operate. However, this introduces several new challenges. The goal is of course to ensure correct and complete translations, but how should statements be treated that cannot be translated? How can one ensure integrity and authenticity of the transformed policies? And, perhaps more importantly; in the event of an error, who is responsible and liable for it?

TABLE I: AN OVERVIEW OF THE USER AGENTS PRESENTED IN SECTION IV.

	Privacy policy		User preferences		Policy-preferences interaction		Implementation
	How is the privacy policy declared?	Who provides the privacy policy?	Can the user specify individual privacy preferences?		Is there any policy-preferences conformance matching?		How is the tool implemented?
AT&T Privacy Bird	P3P	Service providers	Yes	User controlled privacy preferences generator (APPEL).	Yes	"Bird icon" in browser title bar changes color and shape to indicate mismatch or lack of policy.	Browser add-on
PIPWatch Toolbar	Canadian privacy legislation: PIPEDA issues	User community & service providers	Yes	Users can weight privacy concerns related to PIPEDA issues.	Yes	"Beaver icon" in browser frame changes color to indicate privacy risk.	Browser add-on
Kolter et.al.	User provided privacy-related info (policies, ratings, etc)	User community	Yes	User controlled privacy preferences generator (XML-based)	Yes	?	Browser add-on
PRIME	PRIME policy languages	Service Providers	Yes	Using "PrivPrefs" and "TownMap"	Yes	Automatic disclosure or "Send personal data" confirmation box	PRIME user interface and middleware
Integrated Privacy View	P3P + extension	Service Providers	No	Only preset preferences (APPEL)	Yes	"Smiley face icons" on web page indicates conformance or mismatches	HTTP proxy
Privacy Finder	P3P	Service Providers	No	Only preset preferences (APPEL)	Yes	"Privacy meter" on web page indicates degree of conformance.	Search engine (web site)

Particularly the community approach, which is based on volunteers, is dependent on reaching a critical mass of members and translated privacy policies relatively soon in order to create the positive reinforce mentioned above. Unless users see the usefulness of the translation service, the willingness to participate is assumed to be reduced. For professional translators, this may be less important since they may follow common business development phases and build their database of translated policies in advance, only introducing users to the service once it is fairly complete.

C. *Retroactive Effect of Policy Changes*

Both privacy policies and preferences are assumed to be dynamic and may therefore change over time. However, none of the user agents surveyed in this paper handle directly the retroactive effect of such changes. Instead, they verify adherence to preferences before information is shared and are therefore unable to detect changes that occur at a later point in time. Service providers will typically let changes to their privacy policy affect all collected data, including that prior to the change. Users' privacy may therefore have been violated even though the privacy policy was in line with their preferences at the time information was shared.

To overcome this, a user agent needs to continuously or periodically assess whether the privacy policies are in line with the user's preferences. One may argue that CPM and PRIME do keep track of what data has previously been shared with a service, but that is assumed to only be used whenever more data is about to be shared with the service. So, if a user should stop interacting with a service, changes to the policy and possible privacy violations will not be picked up by the user agent.

D. *Privacy Policy Enforcement*

A crucial point in all of the user agents we have presented is the service providers' adherence to their own privacy policies. How can the end-user be sure that the service provider is not collecting more data than the policy states? A common solution is to have public authorities issue a certification ("privacy seal") to web sites and services that demonstrate compliance with their policies, however, a common problem with such an approach is that it requires the user to trust a (possibly unknown) third party certifier.

CPM utilizes the community portal to let users rate the service providers' adherence to their policies. However, it may be difficult for users to know what data is being collected, how it is stored and whether it is used for other purposes. That is, for users to rate the providers adherence to their policy, they must be able to reveal any policy violations. If not, the rating system is meaningless. Another option may be to include trust management systems in general to assess the likelihood of service providers adhering to their policies.

VII. THE WAY FORWARD

Even though there are many promising research efforts in this area, none of the proposed technologies presented in this paper have thus far left the research stage and made the leap into industry adoption. There are several reasons for this and the solution is not straight-forward. However, we believe that there are some particular aspects that need special attention in order to find a solution.

A. *User Interface Design*

In the introduction of this paper we pointed to the privacy paradox; even though users claim to be concerned about privacy, this is not reflected in their behavior. The Technology Acceptance Model (TAM) [24] identifies *perceived* usefulness and *perceived ease of use* as the main variables for determining user acceptance of new technology. Since many of the existing user agents have had a strong focus on building privacy technology that is easy to use (in particular ATandT Privacy Bird, Privacy Finder and PRIME have put a strong effort into the design of the user interface), a reason for the paradox may be that users do not perceive the technology as being *useful*. We believe that greater effort must be placed in conveying to the users the benefit of using privacy enhancing technology on the web. Hence, more efforts must be invested into determining what user interface design would make privacy technology seem both usable *and* useful.

B. Flexibility

As pointed out in the previous section, the flexibility of the user agent is greatly dependent on the flexibility of the underlying policy and preference specification language. Since most of the user agents presented in this paper either rely on fixed or generated privacy preferences there is no need to keep the policy language simple. Instead we view the preference (or policy) generator as an abstraction layer between the user and the actual preferences. Using a more expressive and fine-grained language will therefore not have a negative impact on the usability of the tool, seen from the end-user's perspective. The benefit of this approach is that the user interface of the generator may be tailored for specific purposes, domains or knowledge level of the users. It should therefore be investigated whether other, more powerful, policy languages than P3P (e.g. XACML [25]) can be used for expression and matching of privacy policies and preferences.

C. Continuity of Decision

One problem with privacy policies and preferences is that they evolve over time. To resolve this problem, the matching and evaluation of policies with respect to a set of preferences must continue throughout the personal data life-cycle. In an ideal world, the user agent will conduct periodic evaluation of the policies of services that the user is currently sharing information with. Should a discrepancy be detected, the user will be informed and mitigating measures may be taken. This idea has been inspired by the continuity of decisions envisioned by Park and Sandhu [26] for the UCON access control model.

D. Context Dependency

The need for privacy will of course depend on the context. In the normal case, medical history is not something a user would distribute to anyone asking for it. However, should a life-threatening situation occur, the same user would probably give up his/her privacy in a heart beat. Thus, there is a cost-benefit trade-off where the perceived benefit of a service is compared to the perceived cost of releasing personal data. This idea has been partly elaborated by Hong et al. [27], who proposed a middleware for context-aware application to automatically generate and match user privacy preferences and privacy policies

E. Service Provider Buy-in

Service providers currently have little incentive to implement or facilitate the use of machine-readable privacy policies (the low adoption of P3P is an illuminating example of this). From the service providers' perspective there is a trade-off between obtaining as much information about the users a possible, while still preserving the users' trust. If service providers are to implement privacy enhancing measures, there will have to be a very concrete and measurable upside. We believe that for a user agent to become useful in practice, convincing the service providers to adopt the concept is the key to success.

In cases where collecting personal information is not central to the service to be provided, but still considered by the provider as a nice option, a specific PET feature may be turned into a competitive advantage, as explained in the following. Assuming that a user employs a user agent that has access to the user's privacy preferences, these can be communicated to the service provider during session initiation. The service provider can then refrain from asking for information which would violate the user's privacy preferences. This would improve the user experience, saving the user from eschewing a site which is too noisy, and the service provider would not lose a potential customer because of information that the provider does not really need. For other users with more relaxed preferences (or even without privacy concerns), the service provider can collect more information, possibly in return for more services.

We are in the process of refining these aspects into more detailed requirements and subsequently a design specification. In our work we will evaluate several privacy policy languages and determine how privacy policies from different web service providers can be interpreted by a user agent, and similarly how the end-users' privacy preferences should be registered and rendered into a corresponding set of privacy control rules. A key element will be to continuously assess whether the privacy policies meet end-user expectations and requirements. This will require delving into an important area not covered by any of the user agents surveyed in this paper, namely context-sensitive privacy policies.

Our next step will then be to develop a mock-up prototype of a user agent, with an interface that allows the end-user to specify his privacy preferences in an intuitive and easily understandable way. The user agent will perform policy compliance assessments and display the result through an intuitive graphical user interface. The user agent will then be implemented and tested by a representative selection of the intended end-user community.

VIII. Conclusion

In this paper we have described and compared six different user agents for automatic privacy policy and user preferences matching. We have compared their underlying privacy policy languages, preferences generation methods, matching functionality and implementation levels. We found that four out of six user agents are based on the P3P policy language, while the other either do not use a formal policy language or have developed their own. Most of these user agents rely on service providers to provide policies, with CPM as an exception relying on a community-based service for annotating privacy policies.

The limited adoption of user agents such as the ones presented in this paper is arguably due to a wide range of factors. We have pointed at the limited expressiveness in the selected privacy policy languages, the dependence on a critical mass of adoption and the lack of perceived usefulness as the main reasons for failure.

This paper also stakes out directions for future work. We have identified user interface design, flexibility, continuity of privacy decisions, context dependency and service provider buy-in as the main areas for improvement when developing a future user agent for automatic privacy policy and user preferences matching.

References

[1] S. Spiekermann, J. Grossklags, and B. Berendt, "E-privacy in 2nd generation E-commerce: privacy preferences versus actual behavior," in *Proceedings of the 3rd ACM conference on Electronic Commerce (EC'01)*, 2001.

[2] L. F. Cranor, P. Guduru, and M. Arjula. "User interfaces for privacy agents," *ACM Trans. Comput. - Hum. Interact*, vol. 13, no. 2, pp. 135–178, 2006.

[3] G. Iachello and J. Hong, "End-user privacy in human-computer interaction," *Found. Trends Hum.-Comput. Interact*, vol. 1, no. 1, pp. 1–137, 2007.

[4] A. Clement, D. Ley, T. Costantino, D. Kurtz, and M. Tissenbaum. "The PIPWatch toolbar: Combining PIPEDA, PETs and market forces through social navigation to enhance privacy protection and compliance," in *Proceedings of the IEEE International Symposium on Technology and Society*, 2008.

[5] A. Westin, *Privacy and Freedom*, New York Atheneum, New York, 1967.

[6] WC3. Platform for Privacy Preferences (P3P) Project [Online]. Available: http://www.w3.org/P3P/

[7] L. Cranor, M. Langheinrich, and M. Marchiori, "A P3P Preference Exchange Language 1.0 (APPEL1.0)," World Wide Web Consortium, 2002.

[8] N. Damianou, N. Dulay, E. Lupu, and M. Sloman. "The Ponder Policy Specification Language," *Policies for Distributed Systems and Networks*, pp. 18–38, 2001.

[9] L. Kagal, T. Finin, and A. Joshi, "A policy language for a pervasive computing environment," *Policies for Distributed Systems and Networks*, IEEE International Workshop on, vol. 0, no. 63, 2003.

[10] C. Duma, A. Herzog, and N. Shahmehri. "Privacy in the semantic web: What policy languages have to offer," *Policies for Distributed Systems and Networks (POLICY'07)*, 2007, Eighth IEEE International Workshop on, pp. 109–118, 2007.

[11] L. F. Cranor, M. Arjula, and P. Guduru. "Use of a P3P user agent by early adopters," *Proceedings of the 2002 ACM workshop on Privacy in the Electronic Society* (WPES '02), 2002.

[12] Privacy Bird [Online]. Available: http://www.privacybird.org/

[13] J. Kolter, T. Kernchen, and G. Pernul, "Collaborative privacy management," *Computers and Security*, vol. 29, no. 5, pp. 580–591, 2010.

[14] K. Jan and G. Pernul, "Generating user-understandable privacy preferences," in *Proceedings of the 2009 International Conference on Availability, Reliability and Security (AReS)*, 2009.

[15] PRIME – Privacy and Identity Management for Europe [Online]. Available: http://www.prime-project.eu/

[16] C. Andersson, J. Camenisch, S. Crane, S. Fischer-Hübner, R. Leenes, S. Pearson, J. S. Pettersson, and D. Sommer, "Trust in PRIME," *Proceedings of the 5th IEEE Int. Symposium on Signal Processing and IT*, 2005.

[17] J. S. Pettersson (Ed). HCI Guidelines. PRIME Deliverable vol.1, 2008.

[18] C. Bournez and G. Neven (Eds). Final requirements and state-of-the-art for next generation policies. PrimeLife Deliverable: D5.1.1, 2009.

[19] S. E. Levy and C. Gutwin, "Improving understanding of website privacy policies with fine-grained policy anchors," in *Proceedings of the 14th international conference on World Wide Web (WWW '05)*, 2005.

[20] Privacy Finder [Online]. Available: http://www.privacyfinder.org/

[21] P. G. Kelley, J. Bresee, L. F. Cranor, and R. W. Reeder, "A "nutrition label" for privacy," in *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS'09)*, 2009.

[22] H. Hochheiser, "The platform for privacy preference as a social protocol: An examination within the u.s. policy context," *ACM Trans. Internet Technol.*, vol. 2, no. 4, pp. 276–306, 2002.

[23] S. Egelman, L. F. Cranor, and A. Chowdhury, "An Analysis of P3P-Enabled Web Sites among Top-20 Search Results," in *Proceedings of the 8th International Conference on Electronic Commerce*, 2006.

[24] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly*, vol. 13, no. 3, pp. 319–340, 1989.

[25] OASIS. eXtensible Access Control Markup Language (XACML) Version 2.0. OASIS Open, 2005.

[26] J. Park and R. Sandhu, "The UCON_{ABC} usage control model," *ACM Transactions on Information Systems Security*, vol. 7, no. 1, pp. 128–174, 2004.

[27] D. Hong, M. Yuan, and V. Y. Shen, "Dynamic privacy management: a plug-in service for the middleware in pervasive computing," in *Proceedings of the 7th international conference on Human computer interaction with mobile devices and services*, 2005.

Dr. Karin Bernsmed received her MSc degree from Linköping University in 2003 and her PhD in Telematics from the Norwegian University of Science and Technology (NTNU) in 2007. She worked as a research scientist at Telenor Research and Innovation until 2010, after which she joined SINTEF ICT where she is currently heading the information security group. Her research interests include network security and privacy, security in cloud computing, and stochastic modeling and analysis.

Åsmund Ahlmann Nyre received his MSc degree in Communication Technology from the Norwegian University of Science and Technology (NTNU) in 2005. He worked until 2007 as Security Engineer at the Norwegian road-tolling company Q-Free, after which he joined SINTEF ICT. As of July 2009, Mr. Nyre additionally holds a position as PhD candidate at NTNU. His research interests include information control, privacy and network security.

Martin Gilje Jaatun is a Senior Scientist at SINTEF ICT. He received his MSc degree in Telematics from the Norwegian Institute of Technology in 1992. Previous positions include scientist at the Norwegian Defence Research Establishment (FFI), and Senior Lecturer in information security at the Bodø Graduate School of Business. His research interests include security in cloud computing and security of critical information infrastructures. He is vice chairman of the Cloud Computing Association (cloudcom.org), and a Senior Member of the IEEE.