Privacy Fox - A JavaScript-based P3P Agent for Mozilla Firefox

Fahd Arshad

17-801: Privacy Policy, Law, and Technology
School of Computer Science, Carnegie Mellon University
Pittsburgh, PA 15213, USA.
Email: fahd@cmu.edu

December 7, 2004

1 Abstract

Website operators and visitors are both interested in effective communication of the privacy policies of a website. In practice, however, privacy policies are either too long and difficult for visitors to read through, or do not provide the information that visitors are looking for. Automated translation of P3P policies into human readable form is a promising solution. P3P is a W3C-defined standard for publishing privacy policies in XML. Its adoption is growing, especially amongst the most visited websites on the Internet. This paper describes Privacy Fox, a plug-in for the Mozilla Firefox web browser that translates the P3P policy of the site being visited into two forms: a short privacy notice that highlights the privacy practices that are of most interest to the visitor, and a standard, detailed translation of the entire P3P policy. The translations warn the user about incomplete or unsatisfactory policies.

2 Introduction

The Fair Information Practice principles (FIP) are recognized as the cornerstone of the privacy debate [8]. Whereas there are a number of formulations of the FIP, often the first principle to be recognized is that of notice or awareness. A consumer must be made aware of an entity's privacy policies before the entity gathers any personal information on the consumer. Without this knowledge, the consumer is unable to make an informed decision on whether to participate in the data collection or not. In the digital marketplace, the number of actors and speed of transactions make it extremely hard for notice to be provided. Take the example of an average Web user: such a person may visit a single webpage published by one organization, but as is common today, the webpage may be hosted by one entity, the content provided by another, and the advertisements being served by still another entity altogether. With each click of her hypothetical mouse, our user may be engaging in one, two, or many more transactions with entities she may not have heard of. Each entity may well publish a privacy policy, but no sane user would be able to read the privacy policy of each party on the Web they may encounter. Yet the need for notice remains. This is the impetus behind the Platform for Privacy (P3P) project.

P3P is a W3C-sponsored effort that created a standard for publishing privacy policies in XML [19, 4]. When our user visits a website, her P3P-aware web browser can fetch the P3P policy and match it against the user's personal preferences. P3P-aware clients can inform the user of mismatches between the website's practices and the user's preferences, and optionally influence the flow of information (e.g., by refusing to store cookies set by the website). Existing P3P user agents are diverse both in their goals and their implementations. Some are host-based applications or browser plug-ins (AT&T Privacy Bird) and others are Web-based (the P3P Transmogrifier) [2, 10]. Some only perform translations from P3P to human-readable language, whereas others are capable of decision-making based on the policies they encounter and user-expressed privacy preferences. This paper describes Privacy Fox, a plug-in (extension) for the Mozilla

Firefox web browser that is built around a JavaScript-based P3P parser engine [11]. Privacy Fox's main goal is to translate P3P policies into short, easy-to-read privacy notices as well as into a standard detailed format that allows comparisons across websites.

Privacy Fox distinguishes itself amongst the community of P3P user agents in a number of ways. First, a parser engine written in JavaScript offers cross-platform compatibility that has been missing in client-side P3P user agents to date. The Privacy Fox extension works with the Firefox browser on the Windows, Linux, and Apple OSX platforms without any changes to the code-base whatsoever. Second, the JavaScript engine can easily be borrowed and wrapped into another user agent. This may be desirable because each computer that may need to translate P3P policies will most probably have a JavaScript interpreter already: a Web browser. Third, all P3P-enabled websites probably have human-readable versions as well, but these are often too comprehensive and time-consuming for users to browse through. A short, simplified privacy policy is much more user-friendly. Finally, any content, including privacy policies, becomes much more accessible to users if they control its display while the publisher decides the content. RSS feeds are a good example of this phenomenon. While the version of Privacy Fox described in this paper is able to display the translation in only two formats, it can be extended to allow the user to define what are the important aspects of a privacy policy to her. A user given such a choice will be more likely to be informed about privacy practices of various websites than a user who isn't given such a choice.

This paper provides an overview of previous work on P3P user agents and short privacy notices. It describes the architecture of Privacy Fox: how it is installed, how it retrieves policies, and translates and presents them to the user. Some evaluation results are presented, and finally future avenues of work on Privacy Fox are explored.

3 Background and related work

3.1 User agents

A number of existing P3P user agents translate P3P policies into human-readable format for the user. Perhaps the best known is AT&T Privacy Bird [2]. Amongst its many P3P-related features, Privacy Bird is able to translate a site's P3P policy into plain English. It displays individual policy segments as a collapsible bulleted list. Privacy Bird's biggest drawback is that it is limited to IE 5.x and 6 on Windows. This is unfortunate, because a survey of Privacy Bird users found that its users paid more attention to privacy policies and took steps to safeguard their information; Privacy Bird leaves out the growing community of non-IE browser users [5].

Middle-ware or remote P3P clients, such as JRC's P3P Proxy Service or the P3P Transmogrifier are not as handicapped by platform dependencies [13, 10]. The Proxy Service acts as a go-between for the client and a website, matching the client's stated privacy preferences against a site's P3P policy. It also has the rudimentary ability to translate the P3P policy into plain English. However, this proxy is a proof-of-concept tool and not meant for day-to-day use. Both the proxy and the Web application models such as the P3P Transmogrifier shift the burden of computation to a remote entity, and hence create bottlenecks.

Implementation issues such as retrieval and parsing are affected by the choice of implementation method. However, there are other issues all P3P user agents must tackle. As Ackerman and Cranor point out, P3P transactions present a huge and complicated multivariate decision space between user preferences and site policies [1]. An n-dimensional decision space does not lend itself to easy representation in tabular form. But if the user agent simplifies the decision space by hiding some of the information, it risks misleading the user. Finally, there is the issue of the accuracy, or more accurately, the inaccuracy of P3P translators: if a P3P user agent's translation of the P3P policy differs materially from the site's own plain language privacy statement, serious legal issues arise in addition to ambiguities and consumer confusion. Cranor and Reidenberg suggest clear documentation, certification, and standard user agent guidelines to avoid such a dilemma [7].

3.2 Short privacy notices

Short privacy notices have engendered a lot of interest in various parties concerned about privacy issues. This is because a well-designed short notice would allow the reader to glean relevant information without spending

time sifting through useless details. The content of proposed short notice formats usually corresponds closely with the FIPs. One recommendation is that concise prose or phrases be used to specify the content of the short notice [3]. Gellman, on the other hand, suggests the use of checklists to describe privacy practices [9]. Privacy Bird also tries to reduce the length of privacy policies by using summarized bulleted lists and a layered approach to interfaces.

4 System design

4.1 Anatomy of a Firefox extension

The Mozilla Firefox browser provides the extension mechanism as a way of easily adding functionality to the browser. An extension is modular in structure. The interface elements are specified in an XML-based language called XUL (XML User-interface Language). The actual logic is written in JavaScript. The Mozilla engine's API, XPCOM (Cross-Platform Component Access Model), is exposed to JavaScript via XPConnect. XPConnect is the glue that can be used to access XPCOM from JavaScript, and can thus be used to perform various tasks that are otherwise beyond the capability of JavaScript, such as fetching remote XML [15, 14]. Webpages as well as XML documents can be manipulated through DOM (Document Object Model), a standard specified by the W3C [21]. If an application needs to store any data, it can use RDF (Resource Description Framework), another W3C-defined, XML-based standard [20]. The presentation of any output, either in XUL or in HTML, can be handled by CSS (Cascading Style Sheets) [17].

An extension is a collection of XUL, JavaScript, RDF, CSS, and other resource files (such as icon images) packed into an archive. Such an extension can be installed on Firefox browsers running under Windows, Linux, or Mac OS X to extend their functionality. Extensions available online perform simple but useful tasks such as switching between proxies, synchronizing bookmark files via an FTP server, etc. Other extensions provide much more complex additions to the browser, such as ad blocking or the ability to perform navigation tasks through mouse gestures [12].

4.2 Locating P3P policies

According to the P3P specification, websites that are P3P-enabled may advertise so by three methods: they may place a policy reference file at a well-known location (`http://server_name/w3c/p3p.xml`), publish it via sending a specific HTTP header, or use a <LINK> tag in the HTML pages themselves to point to the reference file. Of these three methods, the first is used by an overwhelming majority of P3P-enabled websites and is the only method Privacy Fox uses.

A policy reference file identifies where different policies for the website reside, and which sections of the website each policy applies to. Each policy itself contains various mandatory elements, such as the website's contact information, what types of data are collected and why, etc. as well as certain optional elements, such as <CONSEQUENCE>.

4.3 Privacy Fox in action

Privacy Fox, once installed, adds an option called "Translate Privacy Policy" under the "View" menu of the Firefox browser (Figure 1). When a user visits a website and chooses this option, Privacy Fox checks the well-known location of the website being visited for a policy reference file. If one is found, Privacy Fox parses it and identifies the applicable policy file. Next, the policy file is retrieved and parsed into an internal data structure (a hash table) that is designed for efficient retrieval during translation and validation.

4.4 Validation

Privacy Fox performs a loose validation of the P3P policy according to the P3P 1.1 specification [23]. It aims to extract as much useful information as it can from even a malformed P3P policy. It keeps track of the mandatory elements in its hash table. During translation, if it finds that a mandatory element is missing, Privacy Fox includes a color-coded warning in its output to warn the user of this fact. An alternative would

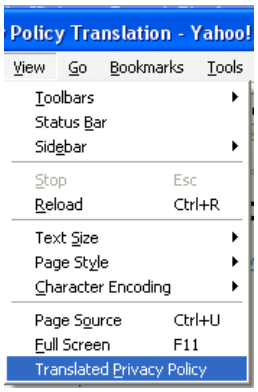

Figure 1: Privacy Fox: Launching from the View Toolbar

be to use the W3C P3P Validation service, but this is implemented using CGI scripts and not as a remote API or Web services, which would have allowed Privacy Fox to intelligently perform validation [22]. The W3C validator outputs only HTML, and parsing it is too non-granular and resource-intensive to be a viable solution for Privacy Fox's JavaScript engine. Also, validation mostly has binary output. It can only tell the user whether the policy follows the specification or not. It is more useful for the user to receive specific warnings about the missing elements, which Privacy Fox provides.

4.5 Translation

The next stage is the translation. Firefox provides "tabs" – nested windows within the main window that allow the user to browse multiple pages within the same browser window. If Privacy Fox finds a privacy policy, it loads two new tabs. One contains a short privacy notice version of the website's P3P policy. The other contains a more detailed translation of the P3P policy. The translation happens on-the-fly. The internal representation of the policy is quizzed and the information channeled into the output tabs. Currently, debugging information and progress reports are sent to a third tab, but this tab can be turned off in non-development versions by toggling a single flag in the code. If the engine encounters an unrecoverable error, such as being unable to find the P3P reference file, or being unable to load the policy XML file, it will use the JavaScript alert mechanism to warn the user. As long as it can find a properly formatted P3P policy file, Privacy Fox aims to handle missing elements etc. as warnings in its output. If Privacy Fox is unresponsive or hangs during a translation, the error can usually be discovered by looking at the output in the debugging tab as well as the warnings provided by the JavaScript Console that is part of the Firefox browser.

The latest P3P specification has guidelines for user agents that translate P3P policies into human readable form [24]. It provides phrases that can be used when different P3P policy tags are encountered. Due to shortage of space, Privacy Fox's short notice format does not follow these recommendations very closely. However, Privacy Fox uses this language as much as it can for the detailed translation; thus this view is called the standard translation.

Figure 2 shows the short privacy notice Privacy Fox generates for Yahoo!'s website. The purpose of this notice is to give the users a quick overview of the privacy practices of the site. It provides links to the website's privacy policy, as well as highlights of the website's privacy practices as declared in the P3P policy: what data is being collected, for what purposes is it being collected, can the user be personally identified from the data, and who is the data shared with?

For the data being collected, Privacy Fox relies heavily on the use of the <CATEGORY> element, since it is much easier to describe in a condensed human-readable form. The P3P specification-recommended language is used here for the translation. Data can also be described in more granular terms by using data elements such as *name.gender*. For most of these, a one-to-one mapping exists between the data elements and categories. For example, *user.id* can be definitively placed under the *unique* category. However, there are variable category elements for which such mapping does not exist (e.g., *date.ymd.year*). Privacy Fox currently only interprets explicit categories; other data is displayed as-is. In future iterations, this will be

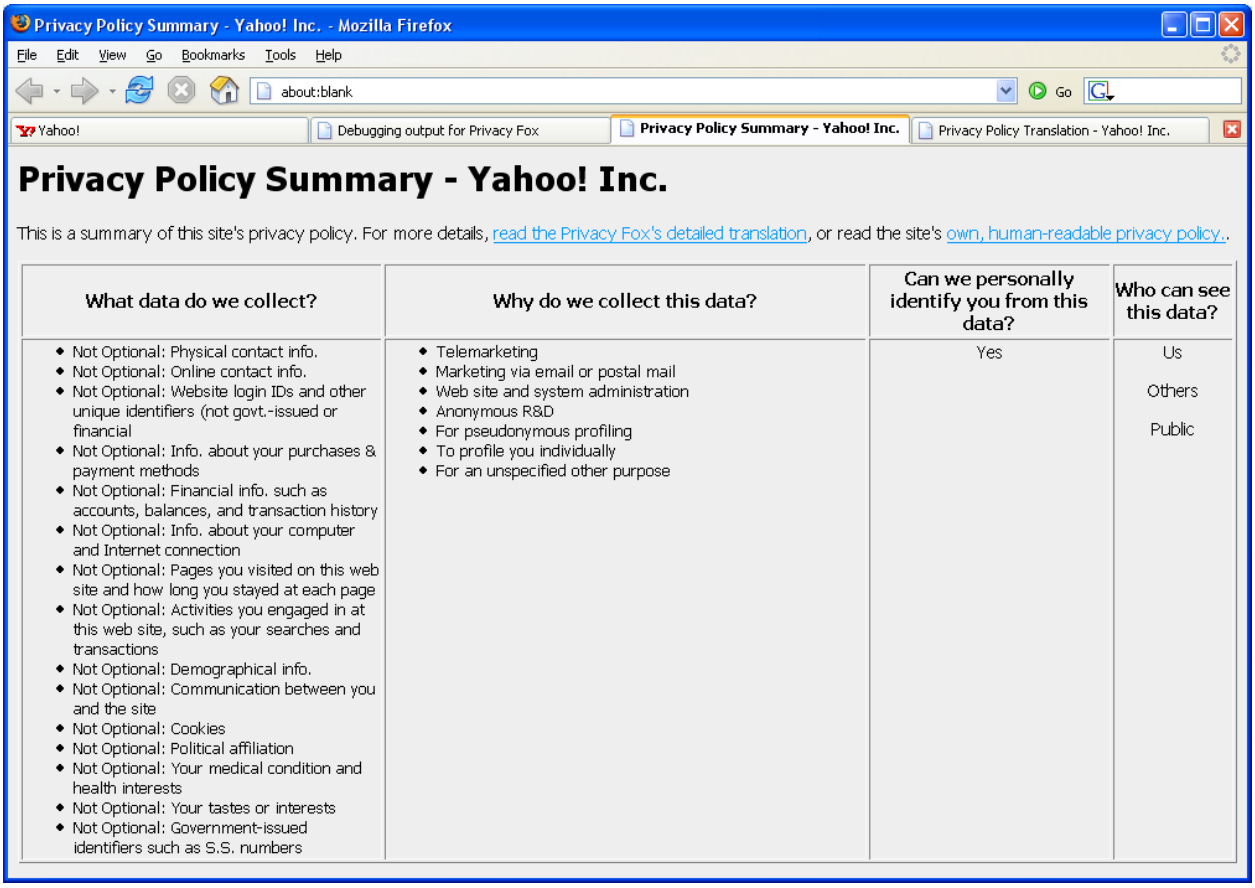

Figure 2: Privacy Fox: Short Privacy Notice for www.yahoo.com's P3P privacy policy

handled in two different ways: the granular elements will be "rounded up" to their categories for the sake of the short privacy notice. A more descriptive translation can be made available in the standard translation. Though the P3P specification does not provide language guidance for these data elements, it does provide the category mappings as part of its description of the data schema.

Figure 3 shows an example of the standard translation format. Most of the details expressed in the P3P policy are displayed here. The translations of the <ACCESS> and <DISPUTE> elements are presented here, in addition to the contact information of the policy issuer. Again, crucial information that is missing from the policy file is highlighted. For each statement made in the P3P policy, four categories of information are extracted: what data is collected, for what purpose, who is it shared with, and how long is it retained? Furthermore, the standard translation also displays the human-readable explanations included in the policy, for example, under the <CONSEQUENCE> element.

4.6 Update mechanism

The Firefox extensions framework allows extensions to be updated automatically. This is achieved by the extension carrying a "phone home" URI referencing a special RDF file. This file contains information on the most up-to-date version of the extension for the version of Firefox being used. The extension can be manually updated by going to the Extension Manager and choosing update from the context menu of Privacy Fox (figure 4.6). Firefox also checks for updates for itself as well as extensions periodically. However, this functionality can be disabled by users.

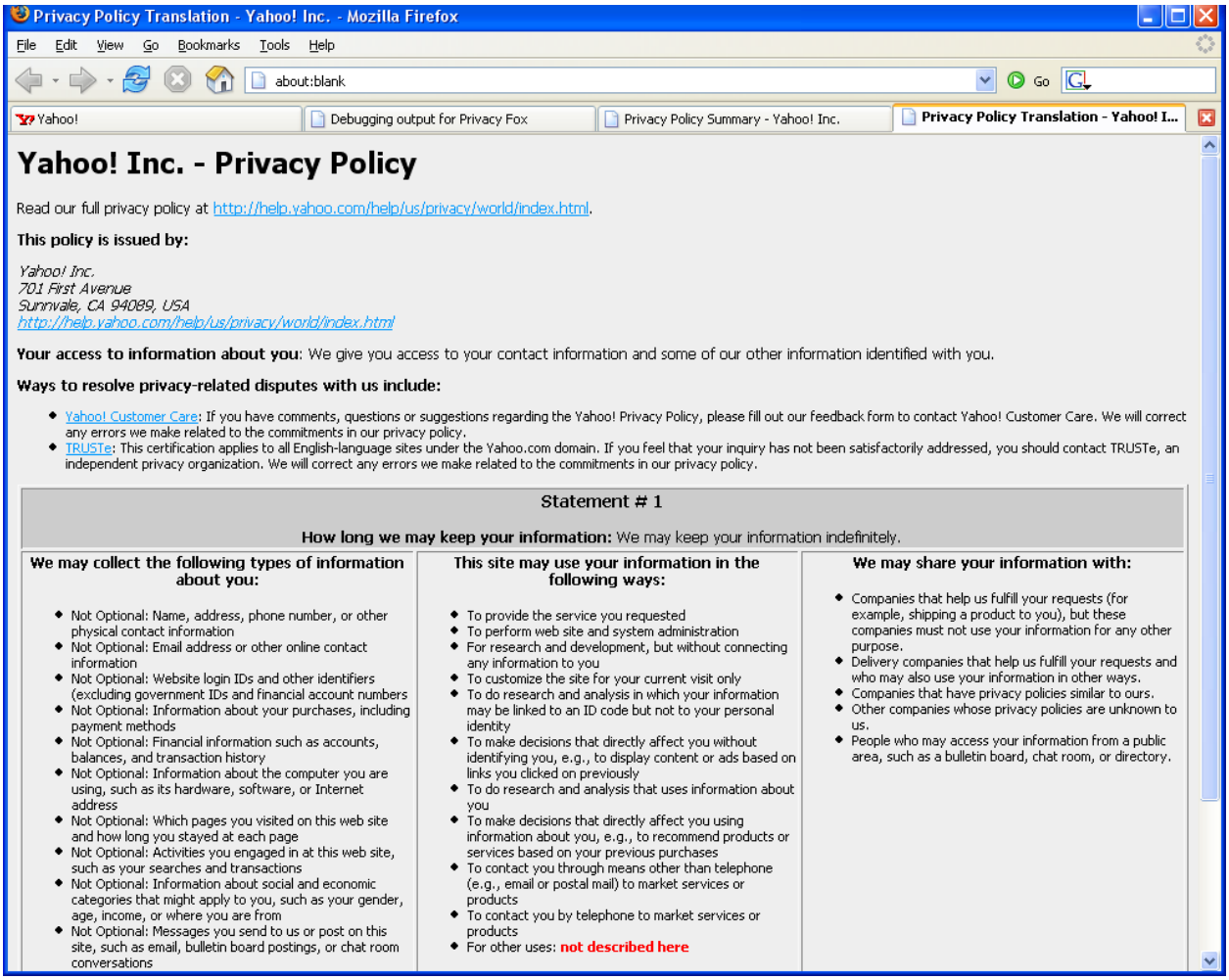

Figure 3: Privacy Fox: Standard Translation for www.yahoo.com's P3P privacy policy

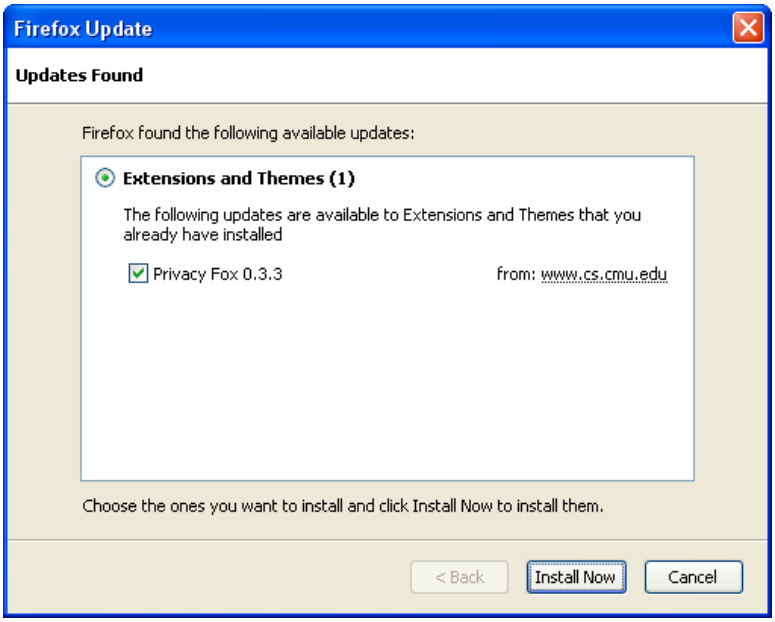

Figure 4: Privacy Fox: Updating the extension through Firefox's Extension Manager

5 Evaluation

5.1 P3P engine & translation

Privacy Fox's engine has not yet undergone significant evaluation for robustness. At present it has been tested successfully on a handful of websites. This list is available at (`http://privacyfox.cs.cmu.edu/`). The most significant source of its failure is that the XPConnect call currently being used to fetch remote XML can only process XML that the remote webserver publishes as the content-type text/xml. If the policy file has a .xml extension, some webservers will serve it correctly. However, a significant number of servers do not, and most who use extensions other than .xml, such as .p3p (the default extension used by IBM's P3P Policy Editor), are not configured to serve the extensions as text/xml. Luckily, this problem can be solved by upgrading Privacy Fox to use a more powerful XPCOM object called XMLHttpRequest [16]. Another problem may be that it can only fetch policies advertised via the well-known location, and not the other two methods, HTTP headers and the <LINK> tag embedded in HTML documents. Please refer to the Discussion & Future Work section for more on solving the limitations of Privacy Fox.

Evaluation of a more robust version of Privacy Fox may be carried out by creating a list of about 20 or so P3P-enabled websites, asking Firefox to translate their privacy policies, and manually inspecting the output. This is as much a debugging process as it is an evaluation, and thus can be performed over and over until a satisfactory performance level is achieved. Privacy Fox's current version has been tested on 6 different websites to date, and it works on 5 of them, the sixth suffering from the content-type mismatch described above. Of the five sites, at least two had invalid P3P policies – many mandatory elements were missing, and it was clear from both the nature of the websites and the content of the human-readable sections of the policy that P3P elements were misused. The most glaring example is the use of the <NON-IDENTIFIABLE> element in a statement together with a <CONSEQUENCE> tag claiming that financial information was collected for billing purposes. If not credit card information does not fall under personally identifiable information, it is hard to imagine what else will.

Another approach to evaluation would be to run Privacy Fox's translation consecutively with an external validation. Privacy Fox may be asked to check a number of websites known to be P3P-enabled, and report whether it manages to finish its translation process without errors or not. This can then be compared to the

validation output from the W3C P3P Validation service. Although this approach can not detect whether Privacy Fox accurately reflects each P3P policy, it can precede a manual examination of the translations and quickly point to the cases where Privacy Fox is expected to parse successfully (i.e., W3C-validated policies) but does not.

5.2 User interpretation and usefulness

Evaluation of whether users are able to interpret the translated privacy policies correctly or not, and whether they find Privacy Fox useful, requires running controlled user studies, or at least user surveys. This is beyond the scope of the current work, but may be considered in the future. However, some of the design principles that form the basis of this work have been verified from lessons learned from previous user agent designs. Privacy Bird designers found that users who had access to privacy policies via Privacy Bird's icon were more likely to be informed of privacy practices than those who were not [6]. The relevance of short privacy notices is discussed in the earlier section on related work.

6 Discussion & Future Work

The future work that needs to be done on Privacy Fox can be divided into three main categories. First is the work that needs to be done for the current translation goals of Privacy Fox to be completed. This would allow Privacy Fox to be used by the wider community as a complete P3P translation tool. Second is the potential work that would allow users to express their privacy preferences to Privacy Fox. This could be stored internally and used by Privacy Fox to provide alerts and modify the translation to highlight or present items of interest to the user. Finally, Privacy Fox could be extended to a forensic tool.

6.1 Improving existing translations

As mentioned earlier, one of the major improvements in Privacy Fox could result from using the XMLHttpRequest object to fetch remote XML instead of the *load* call currently being used. Fetching of policies can also be improved by Privacy Fox being able to process HTTP headers and <LINK> tags to retrieve P3P reference file locations. XMLHttpRequest allows access to full headers also, which should facilitate the former, whereas the embedded tags can be retrieved through DOM.

Interface work can be improved by turning the two tabs into a single, layered interface. This can be accomplished by hiding the standard translation until the user requests more information than is presented in the short privacy notice. There is already a placeholder link in the short privacy notice where this will be implemented. The interface can also be improved further by using template HTML files and using DOM to fill them in instead of creating the entire document on-the-fly. This will not only make Privacy Fox more efficient and maintainable but separate the style information from the translation. Also, this will facilitate the customization of the presentation of the translations by the user, e.g., changing the font color or size of missing elements of the policy.

Further work is needed to group discrete data elements into categories and to provide better translations for them. The detection of personally identified information can also be improved. Currently, Privacy Fox assumes that a person can not be personally identified if the <NON-IDENTIFIABLE> element is used. If the purpose of data collection includes non-anonymous profiling, Privacy Fox reports that the user can be identified. In other cases, it currently withholds judgment. Privacy Fox should be able to further analyze the data collected and the purpose of the collection to inform the user whether or not identified data is being collected.

Some other P3P-related improvements are also needed. The opt-in/opt-out attributes under the PURPOSE and RECIPIENT sections of the policy are not yet captured by Privacy Fox's translations. This is important information that needs to be included. Privacy Fox currently does not match the URL of the website being visited with the scoping rules in the policy reference file; this is crucial to functioning on sites which have more than one privacy policies. Finally, Privacy Fox does not handle extensions at all. While gaining comprehension of various extensions may not be a crucial goal, Privacy Fox should be able to terminate gracefully if it encounters a mandatory extension. The P3P specification asks that extensions

be declared mandatory only when non-awareness of the extension changes the fundamental nature of the interpretation of the policy being extended.

6.2 User preferences & forensic analysis

A P3P agent should be able to obtain the user's privacy preferences, store them, match them against a visited website's privacy practices declaration, and warn the user of any mismatches. The warnings, as well as general availability of a P3P policy, can be conveyed to the user by using either a persistent widget, such as a widget on the browser's toolbar, or transient notifications, such as a sliding panel that appears for a short time and then disappears unless acted upon. The user's preferences could be stored in APPEL (A P3P Preference Exchange Language) [18]. Or, an internal representation more closely tied to privacy issues of concern to users may be utilized.

As discussed in the Evaluation section, a number of websites have self-contradictory P3P policies. Some of these can be detected within the P3P framework using validation. However, many of the contradictions are evident only when the human-readable fields are compared with other elements of the policy. Some heuristics may be discovered and used to extend Privacy Fox to detect such logical mismatches. For example, a website that declares no data is being collected and yet uses cookies may warrant a warning to this effect being added to the translations. Collection of sensitive data may also be inferred from the existence of HTML forms or SSL connections being accepted by the website.

Privacy Fox is very much a work in progress. I have initiated what I hope will be a publicly collaborative project at `http://privacyfox.mozdev.org/`. My goal is to at least accomplish the first half of the future work discussed here.

7 Conclusion

Privacy Fox fills the void of a missing P3P user agent for Mozilla Firefox, and in turn for non-Windows platforms. Its layered approach strikes a balance between a user's need for a quick overview of a website's privacy policy as well as a complete view of the same. By following the P3P specification's guidance on plain language translation, it avoids introducing bias and ambiguity in its output. Due to the simplicity of the engine and the modular design of Firefox extensions in general, it can be rapidly extended. Privacy Fox's biggest contribution may be to make users more aware of the privacy practices of the websites they patronize. Hopefully continued development of Privacy Fox will make it more effective at accomplishing this goal.

References

[1] Mark S. Ackerman and Lorrie Cranor. Privacy critics: UI components to safeguard users' privacy. In *CHI '99 extended abstracts on Human factors in computing systems*, pages 258–259. ACM Press, 1999.

[2] AT&T. AT&T Privacy Bird. `http://www.privacybird.com`.

[3] The Center for Information Policy Leadership. The Short Notices Project. `http://www.hunton.com/Resources/Sites/general.aspx?id=132`.

[4] Lorrie Faith Cranor. *Web Privacy with P3P*. O'Reilly and Associates, 2002.

[5] Lorrie Faith Cranor, Manjula Arjula, and Praveen Guduru. Use of a P3P user agent by early adopters. In *Proceeding of the ACM workshop on Privacy in the Electronic Society*, pages 1–10. ACM Press, 2002.

[6] Lorrie Faith Cranor, Manjula Arjula, and Praveen Guduru. User interfaces for privacy agents. Under review. `http://lorrie.cranor.org/pubs/privacy-bird-20040802.pdf`, August 2004.

[7] Lorrie Faith Cranor and Joel Reidenberg. Can user agents accurately represent privacy notices. In *Proceedings of the Telecommunications Policy research Conference*, 2002.

[8] Finance and Trade Commission (FTC). The Fair Information Practice Principles. http://www.ftc.gov/reports/privacy3/fairinfo.htm.

[9] Robert Gellman. The Short Notice Checkbox Model. http://www.privacyconference2003.org/pdf/gellman_snsample.doc.

[10] Phillip Karlsson. The P3P Transmogrifier. http://goats.com/features/privacy/.

[11] The Mozilla Foundation. Mozilla Firefox 1.0. http://www.mozilla.org/products/firefox/.

[12] The Mozilla Foundation. Mozilla update :: Extensions. https://update.mozilla.org/extensions/?application=firefox.

[13] Joint Research Centre JRC. JRC P3P Proxy Service. http://p3p.jrc.it/.

[14] The Mozilla Foundation. Scriptable Components (XPConnect). http://www.mozilla.org/scriptable/.

[15] The Mozilla Foundation. XPCOM. http://www.mozilla.org/projects/xpcom/.

[16] The XUL Planet. XMLHttpRequest. http://www.xulplanet.com/references/objref/XMLHttpRequest.html.

[17] Worldwide Web Consortium. Cascading Style Sheets. http://www.w3.org/Style/CSS/.

[18] Worldwide Web Consortium. A P3P Preference Exchange Language 1.0. http://www.w3.org/TR/P3P-preferences/.

[19] Worldwide Web Consortium. The Platform for Privacy Preferences (P3P) Project. http://www.w3.org/P3P/.

[20] Worldwide Web Consortium. The RDF Primer. http://www.w3.org/TR/rdf-primer/.

[21] Worldwide Web Consortium. The W3C Document Object Model. http://www.w3.org/DOM/.

[22] Worldwide Web Consortium. The W3C P3P Validator. http://www.w3.org/P3P/validator.html.

[23] Worldwide Web Consortium. The Platform for Privacy Preferences 1.1 (P3P1.1) specification (working draft). http://www.w3.org/TR/2004/WD-P3P11-20040720/, July 2004.

[24] Worldwide Web Consortium. User agent guidelines. the Platform for Privacy Preferences 1.1 (P3P1.1) specification (working draft). http://www.w3.org/TR/2004/WD-P3P11-20040720/#ua, July 2004.