Searching for Privacy: Design and Implementation of a P3P-Enabled Search Engine

Simon Byers[1], Lorrie Faith Cranor[2], Dave Kormann[1], and Patrick McDaniel[1]

[1] AT&T Research, Florham Park, NJ
{byers, davek, pdmcdan}@research.att.com
[2] Carnegie Mellon University, School of Computer Science, Pittsburgh, PA
lorrie@cs.cmu.edu

Abstract. Although the number of online privacy policies is increasing, it remains difficult for Internet users to understand them, let alone to compare policies across sites or identify sites with the best privacy practices. The World Wide Web Consortium (W3C) developed the Platform for Privacy Preferences (P3P 1.0) specification to provide a standard computer-readable format for privacy policies. This standard enables web browsers and other user agents to interpret privacy policies on behalf of their users. This paper introduces our prototype P3P-enabled Privacy Bird Search engine. Users of this search service are given visual indicators of the privacy policies at sites included in query results. Our system acts as a front end to a general search engine by evaluating the P3P policies associated with search results against a user's privacy preference settings. To improve system performance we cache unexpired P3P policy information (including information about the absence of P3P policies) for thousands of the most popular sites as well as for sites that have been returned in previous search results. We discuss the system architecture and its implementation, and consider the work necessary to evolve our prototype into a fully functional and efficient service.

1 Introduction

As people increasingly use the Internet for shopping and other activities, the level of online privacy concern is rising [14]. Many web sites have attempted to address privacy concerns by posting privacy policies and participating in self-regulatory privacy programs. However, it remains difficult for Internet users to understand privacy policies [15], let alone to compare policies across sites or identify sites with the best privacy practices. The World Wide Web Consortium (W3C) developed the Platform for Privacy Preferences (P3P 1.0) Specification to provide a standard computer-readable format for privacy policies, thus enabling web browsers and other user agents to read privacy policies on behalf of their users [7]. However, the P3P user agents available to date have focused on blocking cookies and on providing information about the privacy policy associated with a web page that a user is requesting [8]. Even with these tools, it remains difficult for users to ferret out the web sites that have the best policies. We have developed a prototype P3P-enabled search engine called Privacy Bird Search that offers users the ability to perform Web searches that return privacy policy information along side search results.

D. Martin and A. Serjantov (Eds.): PET 2004, LNCS 3424, pp. 314–328, 2005.

1.1 P3P and APPEL

The P3P 1.0 Specification defines a standard XML format for a computer-readable privacy policy called a *P3P policy*. Although P3P policies contain some human-readable elements, they consist mostly of multiple-choice elements, which facilitate automated evaluation. A P3P policy includes elements that describe the kinds of a data a web site collects, the purposes for which data is used, potential data recipients, data retention policies, information on resolving privacy-related disputes, an indication as to whether a site allows individuals to gain access to their own data, and other information.

P3P became an official W3C Recommendation in April 2002 and has since been adopted by nearly a third of the most popular (top 100) web sites [4]. P3P user agent software is built into the Microsoft Internet Explorer 6 (IE6) and Netscape Navigator 7 web browsers. In addition, a P3P user agent called AT&T Privacy Bird can be downloaded for free and used as an add-on to the IE5 and IE6 web browsers. Other experimental P3P user agents are also available. In addition, a variety of tools have been developed to help web site operators generate P3P policies.

W3C also produced a specification for a language called A P3P Preference Exchange Language (APPEL) that can be used to encode user privacy preferences. APPEL is not an official W3C Recommendation; however, it has been implemented in Privacy Bird and other P3P user agents. APPEL is an XML-based language in which privacy preferences are encoded as rules that can be used to evaluate a P3P policy and control user agent behavior [6]. For example, an APPEL ruleset might specify that access to a web site should be blocked if the site collects data for telemarketing purposes without providing opportunities to opt-out.

1.2 Privacy Bird

AT&T Privacy Bird is implemented as an Internet Explorer browser helper object. The software adds a bird icon to the top right corner of the IE title bar. Users can configure Privacy Bird with their personal privacy preferences using a graphical user interface or by importing APPEL files. The preference interface allows users to select from pre-set high, medium, and low settings, or to configure their own custom setting. The user's preference settings are encoded as as APPEL rule set. At each web site a user visits, Privacy Bird checks for P3P policies. When Privacy Bird finds a policy, it uses an APPEL evaluation engine to compare the policy to the user's preferences. The Privacy Bird icon appears as a green "happy" bird at sites with policies that match a user's preferences. At sites with policies that do not match a user's preferences the icon appears as a red "angry" bird. The icon appears as a yellow "uncertain" bird at sites that have no P3P policy. A user can click on the bird to get a summary of the site's privacy policy, including the specific points where the site's policy differs from the user's preferences [9].

1.3 Related Work

A wide variety of web privacy tools are available that perform functions such as identifying web bugs, blocking cookies, reducing the amount of information transmitted by web browsers to web sites, and facilitating anonymous or pseudonymous browsing [8]. Several now-defunct dot coms offered privacy-related services including an electronic