Electronic Commerce Research and Applications 7 (2008) 274–293

Contents lists available at ScienceDirect

Electronic Commerce Research and Applications

journal homepage: www.elsevier.com/locate/ecra

P3P deployment on websites

Lorrie Faith Cranor [a], Serge Egelman [a,*], Steve Sheng [a], Aleecia M. McDonald [a], Abdur Chowdhury [b]

[a] Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15217, United States
[b] Illinois Institute of Technology, 10W 31St. Chicago, IL 60616, United States

ARTICLE INFO

Article history:
Received 13 April 2008
Received in revised form 15 April 2008
Accepted 15 April 2008
Available online 22 April 2008

Keywords:
P3P
Privacy policies
Search engines
E-commerce

ABSTRACT

We studied the deployment of computer-readable privacy policies encoded using the standard W3C platform for privacy preferences (P3P) format to inform questions about P3P's usefulness to end users and researchers. We found that P3P adoption is increasing overall and that P3P adoption rates greatly vary across industries. We found that P3P had been deployed on 10% of the sites returned in the top-20 results of typical searches, and on 21% of the sites returned in the top-20 results of e-commerce searches. We examined a set of over 5000 websites in both 2003 and 2006 and found that P3P deployment among these sites increased over that time period, although we observed decreases in some sectors. In the Fall of 2007 we observed 470 new P3P policies created over a 2-month period. We found high rates of syntax errors among P3P policies, but much lower rates of critical errors that prevent a P3P user agent from interpreting them. We also found that most P3P policies have discrepancies with their natural language counterparts. Some of these discrepancies can be attributed to ambiguities, while others cause the two policies to have completely different meanings. Finally, we show that the privacy policies of P3P-enabled popular websites are similar to the privacy policies of popular websites that do not use P3P.

1. Introduction

According to a 2005 poll conducted by CBS News and the New York Times, 82% of Americans believe that the right to privacy in the US is either under serious threat or is already lost. This same poll also found that 83% of Americans are concerned about companies collecting their personal information because of the risk that companies might share their personal information inappropriately [8]. These responses are similar to a 2000 survey conducted by The Pew Internet & American Life Project, in which 86% of respondents said that they wanted companies to require permission before using personal information for purposes other than those for which it was provided [24]. To address concerns about their handling of personal data, many websites are posting their privacy policies. However, most users do not read these policies [36]. Furthermore, a majority of individuals surveyed held the mistaken belief that the mere presence of a privacy policy means that a corporation will not share their data [38]. Even those who do bother to read privacy policies often cannot understand what the policies mean [18]. Additionally, websites with poor privacy practices have little incentive to disclose these practices, while websites with good practices may view the posting of their policies as a burden [39]. Thus, privacy policies do not seem to be serving website visitors well.

The platform for privacy preferences (P3P) was created by the World Wide Web Consortium (W3C) to make it easier for website visitors to obtain information about sites' privacy policies [11]. P3P specifies a standard XML format for machine-readable privacy policies that can be parsed by a user agent program. This allows users to specify their privacy preferences to their web browser or other application. When a user encounters a website that does not conform to the user's preferences, the agent can alert the user or take other actions such as blocking cookies.

Both end users and researchers may benefit from increasing P3P adoption. P3P best serves end users when a large number of websites with which users share data make their privacy policies available in the P3P format. Even if only a fraction of websites are P3P-enabled, user agents can help users identify the websites that do use P3P, as well as those that have privacy policies that users deem acceptable. Automated tools can also be used to collect and analyze P3P policies for research purposes. This makes it easy for researchers to collect large numbers of policies and compare them across legal jurisdictions or industry sectors, and to track policy changes over time.

This study aims to assess the state of P3P adoption to inform questions about P3P's usefulness to end users and researchers. In Section 2, we provide background on P3P and existing P3P user

* Corresponding author.
E-mail addresses: lorrie@cs.cmu.edu (L.F. Cranor), egelman@cs.cmu.edu (S. Egelman), shengx@cmu.edu (S. Sheng), am40@andrew.cmu.edu (A.M. McDonald), abdur@duvel.ir.iit.edu (A. Chowdhury).

1567-4223/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.elerap.2008.04.003

agents. In Section 3, we present our study methodology. In Section 4, we measure P3P deployment among a number of different sets of websites. In Section 5, we compare the deployment rates we measured with previous studies and present data we collected by monitoring P3P policy additions, deletions, and changes to answer questions about P3P deployment trends. In Section 6, we present our analysis of the content of P3P policies to answer questions about the level of privacy protection offered on the Internet today. In Section 7, we investigate the accuracy of P3P policies to determine how reliable they are and the extent to which they are being kept up to date. In Section 8, we compare the content of P3P policies with the content of human-readable policies at websites that do not have P3P to gain insights into the representativeness of the privacy policies of P3P-enabled websites. Finally, we discuss our conclusions in Section 9. We conclude that while P3P adoption has been slow to date, the number of sites adopting P3P is increasing, and P3P adoption is strongest for e-commerce and US government websites. We show that there are a large number of errors in P3P policies, but most of these errors do not prevent user agents from making accurate assessments of a website's overall privacy level. We also show that P3P policies are generally representative of all website privacy policies and therefore provide a useful data source for website privacy policy studies.

2. The platform for privacy preferences (P3P)

The platform for privacy preferences (P3P1.0) Recommendation [11] was issued by the W3C in April of 2002. It has been implemented in two popular web browsers and in a number of other P3P user agents. The W3C has also issued "notes" describing A P3P Preference Exchange Language (APPEL) [10] and P3P1.1 [9]. APPEL is a language for representing user preferences about P3P policies. P3P1.1 includes a variety of extensions and clarifications to the P3P1.0 Recommendation and documents suggested wording for presenting P3P policy information to end users in English.

P3P was created to increase understanding of website privacy policies. However, it is not without its critics. Some claim that industry pushes for self-regulation prevent the US from passing a comprehensive privacy law and leave users with far weaker alternatives [28]. Others claim that P3P is hard to implement, lacks enforcement provisions, and will never have enough adopters for it to gain momentum [19]. While some valid concerns have been raised, we believe that P3P needs to be examined within the context of the current privacy policy environment in which a P3P policy is as legally valid as its natural language counterpart [13]. In this paper, we address the issue of adoption and do not cover these other concerns, which are addressed in other papers [34].

In this section, we describe the P3P1.0 Recommendation, some of the P3P user agents currently available, and the Privacy Finder P3P search service we developed.

2.1. P3P 1.0

P3P1.0 specifies an XML syntax for privacy policies, a protocol for user agents to locate P3P policies on websites, and a syntax for compact policies sent in HTTP response headers.

2.1.1. P3P syntax

P3P policies are computer-readable XML documents that provide the name and contact information for the website (<ENTITY> element), the types of information that may be collected (<CATEGORIES> element), how information may be used (<PURPOSE> element), how information may be shared (<RECIPIENT> element), information about an individual's ability to access their own information in the site's records (<ACCESS> element), data retention policies (<RETENTION> element), and options for dispute resolution (<DISPUTES> element). A set of multiple choice options are defined for most of these elements, although human-readable fields are also provided to allow for more detailed explanations of privacy practices. In addition, attributes can be used to indicate whether a particular purpose or recipient is always required or whether an opt-in or opt-out policy applies. P3P policies may also contain a <NON-IDENTIFIABLE> element if a site does not store personally identifiable data or a <TEST> element if the policy has been posted for testing purposes only.

The P3P language is extensible, allowing new elements to be added as needed. These new elements may be labeled as required or optional, indicating whether or not it is safe for a user agent to ignore them if it does not know what they mean.

The W3C runs a P3P validation service that can be used to check the syntax of P3P policies and to make sure P3P files have been setup properly on a website. The Perl code for validation is freely available [30].

2.1.2. Locating P3P policies

P3P1.0 specifies the format for *policy reference files* that indicate the location of P3P policies on a website and the parts of the website to which they apply. Most websites have just one policy for the entire site; however, some have multiple policies that cover different files or directories on the site. Once a P3P user agent has obtained a policy reference file, it has the information it needs to locate the relevant P3P policy.

Websites have three options for notifying user agents about the location of their policy reference files. The first option is to place the policy reference file in a standard *well-known location*: `/w3c/p3p.xml`. The second option is to add an HTTP response header that advertises the location of the policy reference file. The third option is to embed an HTML or XHTML <link> tag in their HTML content.

The well-known location is the most popular and easiest to implement of these methods (77% of the P3P-enabled sites we visited for this study use the well-known location). However, it requires access to a particular directory on the web server, which is not an option for some website operators.

2.1.3. P3P compact policies

Compact P3P policies consist of a series of tokens transmitted in a P3P HTTP header along with a cookie. The purpose of the compact policy is to enable the web browser to make a quick decision about whether to accept a cookie. The compact policy is only a summary of the site's larger policy, but in many cases is enough for the a user agent to make a decision about a cookie. Every site that uses a compact policy is also required to maintain a full P3P policy so that if more information is needed, the full policy can be analyzed by the user agent. Compact policies consist of a series of three-letter and four-letter tokens separated by spaces. These tokens can represent the multiple choice fields of the following P3P elements: <ACCESS>, <CATEGORIES>, <DISPUTES>, <NON-IDENTIFIABLE>, <PURPOSE>, <RECIPIENT>, <REMEDIES>, <RETENTION>, and <TEST>.

2.2. P3P user agents

Microsoft's Internet Explorer 6 (IE6) was one of the first P3P user agents available. IE6 allows users to specify personal privacy preferences by selecting from one of the browser's built-in privacy settings or by importing a privacy settings file. These settings are used to specify conditions under which cookies should be blocked or restricted on the basis of their P3P compact policies. IE6 does not consider full P3P policies in its decisions. A small icon is displayed when cookies have been blocked or restricted, but there is no persistent indicator to provide P3P-related information in IE6. IE6 also