

# A Survey and Analysis of the P3P Protocol's Agents, Adoption, Maintenance, and Future

Ian K. Reay, Patricia Beatty, Scott Dick, *Member, IEEE*, and James Miller

**Abstract**—In this paper, we survey the adoption of the Platform for Privacy Preferences Protocol (P3P) on Internet Web sites to determine if P3P is a growing or stagnant technology. We conducted a pilot survey in February 2005 and our full survey in November 2005. We compare the results from these two surveys and the previous (July 2003) survey of P3P adoption. In general, we find that P3P adoption is stagnant, and errors in P3P documents are a regular occurrence. In addition, very little maintenance of P3P policies is apparent. These observations call into question P3P's viability as an online privacy-enhancing technology. Our survey exceeds other previous surveys in our use of both detailed statistical analysis and scope; our February pilot survey analyzed more than 23,000 unique Web sites, and our full survey in November 2005 analyzed more than 100,000 unique Web sites.

**Index Terms**—E-commerce, P3P, privacy-enhancing technologies, Web page maintenance, Web survey.

## 1 INTRODUCTION

THE ease of data creation, manipulation, and aggregation in the online world has removed many of the previous barriers that protected a user's privacy [1], [2], [3]. Users are now confronted with unfamiliar issues such as having their online actions tracked and profiled [2], [4] and having their personal information sold to third parties without their knowledge [5], [6]. These topics are discussed in a report by the Canadian Internet Policy and Public Interest Clinic that describes the current state of information sharing in North America [7]. The report describes an industry active in selling, aggregating, analyzing, and enhancing customer information. These actions allow companies to enhance business efficiency and undertake more efficient marketing campaigns by allowing them to more accurately target individuals or groups. This information, when used inappropriately, can also have a significant effect on the personal privacy of data subjects. For instance, such information has been used in creating psychological profiles [8], terrorist profiling [5], and identity theft [9].

In an effort to provide guidance to companies about what types of data collection is reasonable, privacy principles have been developed by several governments and organizations. Examples of such principles include the Organisation for Economic Co-operation and Development (OECD)'s privacy principles [10], the Federal Trade Commission (FTC)'s fair information practice principles [11], the US Safe Harbor principles [12], and the principles contained within the European Union (EU)'s Data Protection Directive [13]. Since the OECD's principles have been adopted by all 30 member nations including the US and many EU member states, these principles have the widest

international support. These principles provide guidance to organizations and their patrons regarding limiting the collection of information, data quality, purpose for collection, limiting information use, security safeguards, openness regarding the data collector, data-subject participation, and accountability of the data collector. In an effort to assist both organizations and patrons in abiding by and enforcing these principles, a number of technologies and methods have been either proposed or developed [14], [15], [16], [17], [18], [19], [20], [21], [22]. Anonymization tools in particular (for example, JAP [20] and Tor [21]) are powerful user-side privacy protection tools. A comparative analysis of several privacy-enhancing technologies, with a focus on anonymization, is presented in [68].

This paper analyzes the current state of adoption of one of these technologies called the Platform for Privacy Preferences (P3P) project [14]. P3P has been proposed as a solution to the many problems that exist with human-readable privacy policies. These policies attempt to inform Web site patrons about privacy sensitive actions undertaken by the Web site. However, half of all Internet users may not have the required education to properly understand the legal terminology used [16], [23] or technologies discussed (such as cookies) [24] within these policies. P3P was proposed as a means to lessen the cognitive demands on Internet users, and version 1.0 has been adopted as an official World Wide Web Consortium (W3C) recommendation. This is done through the usage of P3P user agents that first locate and then analyze P3P documents posted on a server. Once analyzed, any relevant information can be displayed to the user in an understandable manner.

A symbiotic relationship exists between P3P agents and P3P documents where neither provides value without the other. The lack of P3P agent adoption in Web browsers is of paramount concern. The quality of Microsoft's Internet Explorer (IE)'s default P3P agent has been questioned [25], and Mozilla's Firefox browser does not contain a P3P agent [26]. In addition, a number of critics have outlined concerns with both the scope of P3P [19], [27], [28], [29] and what

• The authors are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada, T6G 2V4. E-mail: {ireay, beatty, dick, jnm}@ece.ualberta.ca.

Manuscript received 9 Feb. 2006; revised 5 Nov. 2006; accepted 8 Mar. 2007; published online 12 Apr. 2007.

For information on obtaining reprints of this article, please send e-mail to: [tdsc@computer.org](mailto:tdsc@computer.org), and reference IEEECS Log Number TDSC-0023-0206. Digital Object Identifier no. 10.1109/TDSC.2007.1004.

may be implied when P3P is used in its most minimal form [30]. P3P may also suffer from an unwillingness or inability of users to make privacy-preserving decisions [31], [32] and semantic inconsistencies [33]. Furthermore, P3P 1.1 has been released as a working group note [35] but is *not* proceeding to candidate recommendation status at this time. In contrast, Linn [34] suggested that improved tools will lead to a growth in P3P adoption.

Due to P3P's symbiotic nature, the server-side adoption of the protocol is vital to P3P's viability. Cranor et al. conducted Web surveys of P3P server-side adoption in May [36] and July 2003 [37]; hereafter, we will refer to the July 2003 survey as the AT&T survey. This survey claimed that P3P 1.0 [38], which became an official W3C recommendation in 2001, was experiencing an increase in adoption by Internet Web sites. To our knowledge, no further surveys have been published since July 2003. Our goal is to provide evidence depicting P3P's current state of adoption, changes in P3P adoption between July 2003 and November 2005, and P3P policy maintenance practices. This will be accomplished through comparisons between the AT&T survey, a pilot survey completed in February 2005, and our full survey completed in November 2005. Our results indicate that most areas of the Internet have seen P3P adoption stagnate. P3P policies frequently contain syntactic errors, and very few Web sites ever perform maintenance on their policies. In addition, our results indicate that P3P adoption is correlated with Web site popularity and P3P adoption varies significantly between languages and countries.

The remainder of this paper is organized as follows: In Section 2, we introduce the P3P protocol and review the available P3P user agents. In Section 3, we review the survey methodology of the AT&T survey, as well as our extensions and improvements. In Section 4, we present our results comparing the AT&T survey to our own. In Section 5, we present our analysis of our February 2005 pilot survey against our November 2005 full survey. In Section 6, we present our analysis of P3P document errors and maintenance. In Section 7, we present our analysis of P3P adoption across language and national boundaries. In Section 8, we present our results examining P3P adoption and error rates for the most popular Web sites on the Internet. Finally, Section 9 offers a summary and discussion of future work.

## 2 THE P3P PROTOCOL

The current Internet standard for automated handling of privacy policies is P3P 1.0 [14]. The P3P protocol utilizes XML files, whose semantic meaning is governed by a predefined XML schema provided by the W3C, to represent human-readable privacy policies in a machine-readable form. When users visit a Web site, the P3P files posted by Web sites are retrieved and compared against personal preferences defined by the human user via P3P user agents. The agent's responsibility is to retrieve the applicable policy, determine if conflicts exist with the user's preferences, and display these concerns to the user in a form they understand.

The P3P protocol also defines what is known as a P3P compact policy, which is a character string placed within the HTTP header of server responses. Compact policies were developed to limit the bandwidth needed when multiple cookies are set, each requiring a different P3P policy, as in the case of third-party cookies. Thus, instead of a browser requesting the full P3P policy every time a cookie is set, the browser only checks the HTTP header of the response.

P3P provides a machine-readable form of a Web site's privacy policy. A user agent must then analyze the policy content and display the relevant data in a manner that is understandable to the user. However, requirements governing functionality of the P3P agents were not considered in the P3P 1.0 specification (guidelines are included in P3P 1.1, Section 6). As a result, a wide variety of functionality is provided in default P3P agents and browser extensions.

Microsoft bundled a default P3P agent in Internet Explorer 6.x (IE 6.x). IE 6.x is limited to blocking cookies and displaying a human-readable form of the P3P policy. To our knowledge, there is no evidence indicating that IE 6.x's human-readable policies are less complex than the usual human-readable privacy policies found on Web sites. Netscape Navigator 7.x's default user agent possessed a cookie-handling functionality similar to IE 6.x and presents the human-readable policy in a bulleted format [25]. However, this P3P agent is no longer included by default in current Netscape or Mozilla browsers [26].

Three major P3P browser extensions or proxy services have been created. AT&T's Privacy Bird [39] is an IE extension that was developed by the AT&T survey's authors. A usability study [25] found that this agent generally outperformed IE 6.x's P3P agent and human-readable privacy policies. However, they also found the usability of their agent suffered from the lack of available P3P policies. With current P3P adoption rates, users rated the AT&T Privacy Bird agent 2.9 out of 5, where 5 is highly usable (5-point Likert scale). The second browser extension, Privacy Fox [40], was developed as an open source project for Firefox. The design of Privacy Fox was highly influenced by AT&T Privacy Bird. As of February 2006, the Privacy Fox project appears to be defunct. The third browser extension is the Joint Research Center (JRC) P3P Proxy [41], which filters Web requests made through a browser through an intermediate proxy service that requests, analyzes, and returns relevant information to the user's browser embedded in the Web page. A P3P policy editor tool is also available.

## 3 THE SURVEY METHOD

The goal of the AT&T surveys was to determine the extent of P3P adoption since its inception as a W3C recommendation in 2001. Our survey followed a similar but improved method in order to determine how P3P adoption has evolved between May 2003 and November 2005. Comparisons will also be undertaken between our February 2005 pilot survey and our November 2005 main survey. The February 2005/November 2005 comparison allows for a far more detailed analysis due to the limitations in data quality and quantity in the AT&T survey.

TABLE 1  
Lists of Web Sites Used in the AT&T Survey

List	Source	Date	Description	Currently Available
PFF Most Popular	Popularity rankings by Neilson/NetRatings	Oct-01	Contains 85 of the 100 busiest websites	No
PFF Random	Popularity rankings by Neilson/NetRatings	Oct-01	302 sites selected at random from the top 7,821 websites	No
PFF Refined Random	Popularity rankings by Neilson/NetRatings	Oct-01	209 sites selected at random from the top 5,625 websites	No
Netscore top 500	Popularity rankings by comScore media metrix	Jul-02	Top 500 websites	No
Key Measures	Popularity rankings by comScore media metrix	Jul-02	Top 500 websites	No
Alexa Global 500	Popularity rankings	Feb-03	Top 500 websites rated by Alexa	Yes
Froogle	Web Crawl of froogle.com	Apr-03	Businesses indexed in Froogle.com	No
Yahooligans	Web Crawl of yahooligans.com	Apr-03	Children's websites indexed on yahooligans.com	No
FirstGov	Compiled from firstgov.gov	Apr-03	The US government websites indexed on firstgov.gov	Yes
News	Web Crawl of news.google.com	Apr-03	News websites indexed on news.google.com	No

In an effort to minimize the threats to validity inherent within the survey, we have adopted the following policies (experimental design practice):

- Implementation bias must be minimized in favor of a faithful implementation of P3P.
- Error analysis must accurately define what constitutes an error in a P3P policy.
- Rigorous statistical analysis must be employed to provide confidence in survey results.
- Selection of Web sites to be surveyed must be undertaken in a definitive and reproducible manner.

Unfortunately, the original AT&T survey does not possess many of these characteristics, causing considerable problems as outlined in the following sections.

### 3.1 The AT&T Survey Method

Through the use of the AT&T Privacy Bird P3P agent, a series of Web sites belonging to multiple lists were surveyed in May 2003 to determine if they possessed P3P policies. The retrieved policies were then analyzed to determine the types of data used by each Web site.

#### 3.1.1 Site Selection

Cranor et al. [36] chose to select a series of Web sites that were popular or located on popular indexes or lists instead of implementing a random sampling methodology (Table 1). Their rationale for this originates from a desire to analyze P3P from a user's perspective.

Although we concur with this choice, there are several problems with the lists selected. Some lists (PFF Top 100, PFF Refined Random, and PFF Random) were constructed using rankings no longer publicly available; in addition, the authors of the lists arbitrarily removed "adult," "children," and "business-to-business" Web sites without providing a precise definition of what these terms mean or how their "removal" policy was implemented. The Key Measures and Netscore Top 500 lists are also no longer available, and the Netscore Top 500 list removed sites satisfying the criteria of "third-party sites including advertising networks," where third-party and advertising networks were not clearly defined. Further, the decision criteria for link selection in the Web crawler employed to gather the Froogle, Yahooligans, and News lists was not explained. This is important because a Web site may contain links that belong to a different category; for example, news.google.com links to impages.google.com, which is not a news site and may possess a different P3P policy.

#### 3.1.2 Collection of Data

The AT&T survey used the P3P analysis engine of the AT&T Privacy Bird P3P agent [39] that was previously implemented by the survey's authors. This engine provides tools for XML analysis and the definition of privacy preferences using the A P3P Preference Exchange Language (APPEL) rule-based language [42]. These abilities allowed for the following types of data to be collected: information pertaining to XML structural correctness, P3P document validity, frequency of P3P tag occurrence, and a categorization of policies based on perceived safety, subjectively defined by the AT&T Privacy Bird creators.

The use of this engine introduces an implementation bias that is incompatible with the implementations of the official P3P analysis tools provided by the W3C. The engine records errors if omissions or mistakes are found in the document that leads to the policy being unusable. If the policy contains errors that are not critical, then no error will be reported [36]. For this reason, AT&T Privacy Bird may be able to analyze policies that would cause other agents to fail. In addition to errors, implementation bias is introduced through the categorization of policies using the predefined APPEL rule sets in the engine. The rule sets in question categorize Web sites based on a perceived level of risk applied to certain actions on individual data items. However, this level of risk is highly dependent on factors outside of P3P's scope such as situational context [43].

#### 3.1.3 Analysis of Data

The majority of the data analysis discussed the general P3P adoption by list and relative frequency of P3P tags by list. The analysis pertaining to the general adoption of P3P detailed both the number of sites implementing P3P and the number of sites whose P3P documents were found to be unusable. However, inferences made upon differences in the observed data are questionable due to the lack of statistical analysis. The analysis concentrating on the relative frequency of P3P tags adds little value to the goals of our project, and for this reason, it is only mentioned to ensure a complete overview of the AT&T survey. Finally, it is noted that the survey did not report on the adoption of

TABLE 2  
Lists Comparable with the AT&T Survey

List	Source	Date collected	Description	Comparable list in AT&T survey
Alexa Global 500	Alexa.com	Feb 2005	Top 500 websites ranked on unique visitors and page views	Alexa Global 500
FirstGov	Firstgov.gov	Feb 2005	US government websites	First Gov
Alexa Top 100	Alexa.com	Nov 2005	Top 100 websites ranked on unique visitors and page views	PFF most popular
Alexa Top 7,821	Alexa.com	Nov 2005	Top 7,821 websites ranked on unique visitors and page views	PFF random
Alexa Top 5,625	Alexa.com	Nov 2005	Top 5,625 websites ranked on unique visitors and page views	PFF refined random
Top 400 Ecommerce	Internet retailer magazine	Nov 2005	Top 400 E-Commerce websites ranked on revenue	Froogle

P3P compact policies except to state an overall adoption rate for all analyzed lists. No discussion was provided describing the extent of the P3P compact policy adoption in individual Web site populations.

### 3.2 2005 Survey Methodology

The 2005 survey uses a site selection mechanism that is similar to the AT&T survey but utilizes an official policy verification tool provided by the W3C and will extend and improve upon the analysis undertaken in the AT&T survey by using a greater number of business or e-commerce lists, a greater number of Web sites surveyed, and a rigorous statistical analysis. This survey methodology was implemented in both the February 2005 pilot survey and our full November 2005 survey.

#### 3.2.1 Site Selection

Table 2 outlines the comparisons undertaken between the 2005 surveys and the AT&T survey. The Alexa Global 500 [44] and FirstGov [45] lists exist in both surveys, so no approximation is required. Alexa.com is a respected subsidiary of Amazon.com that collects traffic information that is then used by Amazon.com and other affiliated business partners to analyze Internet traffic patterns. The FirstGov list contains all US federal and state Web sites, allowing for an analysis of P3P adoption in an environment where machine-readable privacy policies must be provided. The remainder of the lists in Table 2 requires approximations. Web site rankings from Alexa.com are used to approximate the PFF lists as both are based upon popularity rankings. The ranking methods employed by Alexa.com and the PFF surveys do differ slightly; the PFF lists estimate Web site popularity by the number of unique visitors per month, whereas the Alexa rankings are the geometric mean of both the number of unique users per day and the number of unique pages the user viewed per day for a specified

TABLE 3  
Lists That Will Be Used in the Extended Analysis

List	Source	Date collected	Description
Top 300 E-Commerce	Internet retailer magazine	February 2005	Top 300 E-Commerce websites ranked on revenue for 2004
BBBOnline	BBBonline.org	February 2005	All websites/businesses that qualify for the BBBOnline reliability seal
Truste	Truste.org	February 2005	All websites/businesses that qualify for the Truste seal
Alexa Language Lists	Alexa.com	February 2005	20 languages, each containing 100 websites ranked on unique visitors and page views
Forbes 500	Forbes Magazine	February 2005	The Top 500 international companies ranked on revenue

time period. In addition, the exclusion of "adult," "children," and "business-to-business" Web sites from the PFF lists creates a further inconsistency between the two surveys since no sites were excluded in the 2005 survey. The Froogle list is approximated using the Top 400 E-commerce Web sites ranked by Internet Retailer Magazine [46]. To our knowledge, no other organization has amassed as large or as accurate a list as Internet Retailer Magazine. The Top 400 E-commerce list is ranked on revenue generated from the online sale of physical items. Although both of these lists are intended to target e-commerce sites, it is unclear as to whether the Froogle list is entirely populated by e-commerce sites; the Web crawler used to collect this list is never fully described. The comparison between the Froogle and Top 400 E-commerce list is further complicated by the fact that the Top 400 E-commerce list is ranked by online revenue (e-commerce sales), whereas the Froogle list does not use this financial metric.

Table 3 describes the lists that will allow an extension of our results beyond the AT&T survey. The Top 300 E-commerce list [47] was used in the February pilot survey because its replacement, the Top 400 E-commerce list, had not yet been released. The BBBOnline list [48] provides a collection of business Web sites who have registered for the Better Business Bureau's Web-reliability seal. A Web-reliability seal indicates a trusted third party (BBBOnline) has certified that the business practices of the Web site in question meet the third party's standards of ethical business practice (including privacy protection). BBBOnline's parent, the Better Business Bureau, has operated as a third-party recommendation organization in North America since 1912. The Truste list [49] targets a similar domain of business Web sites that have shown a previous interest in third-party certifications. Truste has operated as a nonprofit third-party recommendation service since 1997. Truste and BBBOnline are the two largest online recommendation organizations, and the addition of these lists allows for analysis of Web sites who have shown a previous desire to implement technologies that inform users about a Web site's business practices. The inclusion of the Forbes 500 [50], which is generally considered the most accurate listing of large

companies, allows for insight regarding P3P adoption in the largest companies in the world.

Finally, a major limitation of the AT&T survey is its bias toward P3P adoption in English-language Web sites. The addition of the Alexa Language lists [44] will allow us to explore P3P adoption in a broader linguistic and cultural context that to our knowledge has not been attempted in any previous survey. If it is found that P3P adoption is biased toward a particular culture or language, this problem must be remedied since it is unreasonable to expect all privacy conscious individuals to speak a single language.

Although our lists cannot be considered representative of the general Internet, our lists do represent sectors of the Internet where important privacy concerns exist. Popular Web sites (Alexa Global 500 and Alexa Language Lists) have the opportunity to influence online privacy since their actions affect large numbers of users. Business Web sites (Forbes, Top 300/400 E-commerce, BBBOnline, and Truste) may implement user profiling, data aggregation, and differential pricing. Finally, government Web sites (FirstGov) often utilize highly sensitive information (social security numbers and health/financial information), which if mishandled can have serious consequences to a user.

### 3.2.2 Collection of Data

Instead of using the AT&T Privacy Bird engine, our survey used the official P3P validator tool [51], provided by the W3C, to harvest P3P policies and extract relevant information. This choice is due to the lack of strict adherence to the P3P protocol by the AT&T Privacy Bird engine regarding handling policies with errors [36]. Although this lack of adherence may be desirable in a fault-tolerant P3P agent, it is not advantageous when one is attempting to scientifically analyze P3P adoption.

### 3.2.3 Analysis of Data

For each of the points of interest between the two surveys, we formally investigate the change in the representative dichotomous variables—we will report on the relative significance between each of these relationships. We decided that a common approach to this analysis was preferable, and during the initial design of the survey, we had limited insight into the likely values that each of the dichotomous variables would possess—and could not rule out a series of low frequency values. Given these assumptions, it was believed that an application of Fisher's exact probability test was more appropriate than a chi-square test (with or without Yates' correction), which has difficulties accommodating low-frequency values and does not naturally lend itself to directional formations. Having said this, we note that, in general, Fisher's exact probability test possesses a lower statistical power than the corresponding chi-square test and, hence, our analysis can be viewed as conservative in this respect [52]. Finally, due to what we see as conflicting causal evidence on adoption drivers for P3P, all statistical tests are formulated as two-tailed nondirectional tests, and  $\alpha$  is set at the traditional 0.05 level.

Due to our desire to investigate P3P adoption in a general sense, the analysis of dichotomous variables must be expanded beyond individual lists. This analysis is complicated by the existence of sample sizes varying from 300 to 20,263 Web sites; a simple amalgamation of unique

TABLE 4  
P3P Adoption in Lists That Do Not Require an Approximation

Date	Firstgov		Alexa Global 500	
	July-03	Nov-05	July-03	Nov-05
Total Sites	344	366	500	500
Total Accessible	338	359	495	489
Availability P-value	1.0000		0.2064	
Total P3P enabled	7	107	92	79
P3P enabled P-value	0.0001		0.3548	

Web sites would bias the results toward the large lists. We decided to employ resampling techniques to create representative amalgamated (from multiple lists) distributions of the union of individual lists by randomly sampling each list 10,000 times. This process results in the construction of a Gaussian distributed description of the entire population [53]. Each population was amassed through the selection of an equal number of Web sites from each list, chosen randomly and without replacement from their respective list. Due to the limited guidance in the relevant literature regarding how large a sample should be, we arbitrarily decided to collect 200 Web sites from each list. This choice allows for a random selection from all lists to occur, and when we tested our results against other sample sizes, we found little difference in results. Duplicated Web sites in the lists were removed once the representative populations were created; this choice ensured that the created populations are realistic (no duplicates) and that lists whose Web sites were selected last were not overrepresented.

The distributions representing the dichotomous variables contained in these populations are analyzed through one-way analysis of variance (ANOVA) ( $\alpha = 0.05$ ) and Cohen's d effect size (including bias-corrected equivalents) [54] methods. The use of statistical significance tests alone is unreasonable due to the one-way ANOVA's sensitivity to sample size, which conflicts with the need for large populations in resampling techniques [55]. Cohen's d effect size allows for an analysis independent of sample size indicating the magnitude of the experimental effect. Since there is no method for calculating Cohen's d effect sizes directly from the contingency tables used during the application of Fisher's exact test, we will first calculate an odds ratio effect size [56] that are then converted into a Cohen's d effect size using the method described by Hasselblad and Hedges [57]. Our analysis will follow Cohen's suggestion that  $d \pm 0.2$  indicates a small effect,  $d \pm 0.5$  is a medium effect, and  $d \pm 0.8$  is a large effect [54].

## 4 COMPARISON OF THE AT&T AND 2005 SURVEYS

The comparison between our November 2005 survey and the AT&T survey will be completed in two parts. The analysis is divided due to the list approximations that led to a decreased confidence in results.

Table 4 examines full policy adoption in the lists whose domains can be considered the same with a high degree of confidence. The high degree of confidence stems from both the sources being identical and the lack of any statistically significant difference in availability. The results indicate a disparity in full policy adoption between US government

TABLE 5  
Analysis of Lists Requiring an Approximation for Comparison between 2003 and 2005

	Froogle	Top 400 E-Commerce	PFF Top 100	Alexa Top 100	PFF Refined Random	Alexa Top 5,625	PFF Random	Alexa Top 7,821
Total Sites	1017	400	85	100	209	5,625	302	7821
Total Accessible sites	1010	394	84	100	195	5519	286	7672
Accessible sites p-value	0.2111	1.0000			0.0001		0.0004	
Total sites P3P enabled	133	121	26	17	29	540	35	669
P3P enabled p-value	0.0001	0.0353			0.0275		0.00438	

Web sites and the most popular Web sites on the Internet. Government Web sites have seen a statistically significant ( $p < 0.0001$ ) large ( $d = 1.6537$ ) increase in full policy adoption. The increase in adoption is likely a result of the enacting of the E-governance Act of the US [58]. This act requires government Web sites to provide their privacy policies in a machine interoperable manner. In contrast, no statistical evidence exists indicating a change in adoption for the Alexa Top 500 Web sites ( $p < 0.3548$ ).

The analysis in Table 5 assumes that the samples (May 2003 and November 2005) are drawn from the same underlying distribution and can be considered a representative sample of the distribution. Given that the AT&T survey either chose to exclude certain "categories" of data or did not provide a suitable description of their collection techniques, it is debatable as to the validity of the assumption as this is likely to have skewed the sample. Hence, we urge caution when interpreting the results from Table 5. A further concern is with regard to Web site accessibility between competing lists skewing the results. This is analyzed in line 2 of Table 5. As can be seen, statistically significant differences exist in the number of accessible sites for the comparisons between PFF refined random/Alexa 5,625 ( $p < 0.0001$ ) and PFF random/Alexa

7,821 ( $p < 0.0004$ ); however, the associations are unlikely to have a substantial effect as they exhibit effect sizes of  $d = 0.2591$  and  $d = 0.2235$ , respectively, indicating a relatively small effect.

If one assumes that the distributions are suitably similar, then there is evidence indicating that all four groups of Web sites experienced a statistically significant change in full policy adoption (line 5). Further ad hoc analysis shows that an increase in full policy adoption occurred for e-commerce sites (Froogle/Top 400 E-commerce), whereas a decrease has occurred in the remaining groups. We are unable to provide any explanation for the divergence of these results.

## 5 COMPARISON OF THE FEBRUARY AND NOVEMBER 2005 SURVEYS

Table 6 summarizes how P3P adoption has changed between February and November 2005. Our confidence in the results is high due to the lists being relatively independent (Table 7), with most of the overlaps being sufficiently under 10 percent. In addition, no statistically significant differences in Web site availability exist except for the BBBOnline and FirstGov lists. In the case of BBBOnline, although the difference is statistically significant, the association is unlikely to have a substantial effect as it only exhibits a small effect size of  $d = 0.1446$ . To avoid problems of small cell frequencies when calculating the odds ratio effect size for the accessibility of Web sites for the FirstGov list, we will add 0.5 to the observed frequencies in accordance with accepted practice [59]. The resulting large ( $d = 1.5036$ ) operational effect detected in the FirstGov list suggests a major change occurred, and although 98.1 percent of the sampled Web sites were still able to be contacted, caution is urged when interpreting these results.

The FirstGov ( $p < 0.0220$ ) and BBBOnline ( $p < 0.0004$ ) lists were the only lists that experienced a statistically significant difference in adoption of full P3P policies. Through ad hoc analysis, full policy adoption was found to be increasing for both FirstGov (likely due to the enacting of the E-governance Act) and BBBOnline. We know of no rationale for the differences between BBBOnline and the other lists. BBBOnline was also the only list that experienced a statistically significant change in P3P compact policy adoption.

TABLE 6  
Comparison of February 2005 and November 2005

	Alexa Global 500		BBBOnline		FirstGov		Forbes 500		Top 300 E-Commerce		Truste	
	Feb	Nov	Feb	Nov	Feb	Nov	Feb	Nov	Feb	Nov	Feb	Nov
Total Sites	500	500	20263	20263	366	366	500	500	300	300	1317	1317
Total Accessible Sites	480	489	19322	19059	366	359	493	496	298	293	1277	1286
Accessible p-value	0.1433		0.0001		0.0152		0.5466		0.1768		0.3359	
Total sites with full policy	77	79	844	980	81	107	24	27	87	87	192	202
P3P enabled p-value	1.0000		0.0004		0.0220		0.7740		0.9282		0.6614	
Total sites with compact policy	60	68	770	1045	11	17	12	15	89	82	149	149
Total compact policy p-value	0.5694		0.0001		0.2517		0.6971		0.6505		0.9510	
Total sites with full and compact policy	49	46	458	630	9	14	7	8	59	49	91	94
P3P compact policy p-value	0.7460		0.0001		0.2956		1.0000		0.3404		0.8789	

TABLE 7  
Overlap between Lists in November 2005

	Alexa Global 500	FirstGov	Top 300 E-Commerce	Forbes 500	Truste	BBBOnline
Alexa Global 500	-	75.00% (3) [4] {0.80%}	38.89% (7) [18] {3.60%}	38.46% (5) [13] {0.60%}	29.41% (20) [68] {13.60%}	33.33% (6) [18] {3.60%}
FirstGov	75.00% (3) [4] {1.09%}	-	0.00% (0) [0] {0.00%}	0.00% (0) [0] {0.00%}	0.00% (0) [0] {0.00%}	0.00% (0) [0] {0.00%}
Top 300 E-Commerce	38.89% (7) [18] {6.00%}	0.00% (0) [0] {0.00%}	-	23.08% (3) [13] {4.33%}	23.53% (8) [34] {11.33%}	56.10% (23) [41] {13.67%}
Forbes 500	38.46% (5) [13] {2.60%}	0.00% (0) [0] {0.00%}	23.08% (3) [13] {2.60%}	-	37.50% (3) [8] {1.60%}	28.57% (2) [7] {1.40%}
Truste	29.41% (20) [68] {5.16%}	0.00% (0) [0] {0.00%}	23.53% (8) [34] {2.58%}	37.50% (3) [8] {0.61%}	-	19.51% (32) [164] {12.45%}
BBBOnline	33.33% (6) [18] {0.09%}	0.00% (0) [0] {0.00%}	56.10% (23) [41] {0.20%}	28.57% (2) [7] {0.03%}	19.51% (32) [164] {0.81%}	-

Ex. The overlap between Alexa Global 500 and FirstGov consists of 4 sites. Of these 4 sites, 3 are P3P enabled, leading to a P3P adoption rate of 75% in the overlap. Additionally, Four websites exist in both Alexa Global 500 and FirstGov, thus the percentage of the FirstGov list that overlap with Alexa Global 500 is  $4/366 = 1.09\%$  and the percentage of Alexa Global 500 websites that overlap with FirstGov is  $4/500 = 0.80\%$ .

It is surprising that many Web sites adopt full P3P policies without compact policies since rational organizations would only adopt a technology when suitable incentives exist. Given the lack of usable P3P user agents or laws requiring P3P adoption for private organizations, the most significant incentive to adopt P3P appears to be the

IE 6.x cookie blocking feature. Our results, however, suggest that some other motivation appears to exist. However, due to limitations in our data set, we are unable to definitively ascertain what these motivations may be. The results in Table 6 also suggest that the P3P protocol is often abused since many Web sites using compact policies do not post a corresponding full policy. This result brings into question the inclusion of P3P compact policies in the P3P specification since the only method available to determine if a Web site is adhering to the P3P protocol is to attempt to request the full P3P policy, thereby negating any performance improvements. P3P 1.1 includes guidelines for compact policy processing that include rules for checking the validity of compact policies.

Table 7 indicates another unexpected result, suggesting that when a Web site exists in more than two lists in Table 6, it is often more likely to possess a P3P document. Unfortunately, analysis beyond this informal statement would be questionable due to the generally small magnitude of the overlaps. This analysis is only provided for completeness and to point out a peculiar effect that we are currently unable to explain.

Since no generalizations of P3P adoption can be determined from the results in Tables 6 and 7, a resampling-based analysis will now be conducted (see Table 8). In this analysis, four groups were developed from the lists analyzed in Table 6. These groups are based upon general characteristics inherent to the Web sites belonging to the lists, and due to space constraints, only full P3P policies will be analyzed. We urge caution when interpreting our groups as true indicators of their respective Internet fields. It is possible that our groups lack important subsets of Internet sites. In addition, our groups may suffer from an overemphasis of lists such as FirstGov that could have unique P3P adoption drivers (E-governance Act). The "All Lists" group is an amalgamation of all six lists contained in Table 6. The results for this group represent the general P3P adoption over all surveyed lists. The "Nonlegislated" group is comprised of all Web sites not required by law to adopt P3P or a similar substitute (all lists except FirstGov). This group will provide insight as to whether Web site administrators choose to adopt P3P on their own accord. The "Business" group is comprised of all Web sites belonging to Forbes 500, Top 300 E-commerce, BBBOnline, and Truste. These lists have a business focus,

TABLE 8  
General P3P Adoption in Various List Groupings Using Resampling Techniques

		All Lists		Non-legislated		Business		Seal adopting	
		Feb 05	Nov 05	Feb 05	Nov 05	Feb 05	Nov 05	Feb 05	Nov 05
Number of generated populations		10000		10000		10000		10000	
Mean number of P3P policies		175.46	194.8	131.44	135.13	103.75	107.1	37.98	40.21
Std deviation		8.46	8.96	7.66	7.80	6.80	7.05	5.40	5.52
Std errors		0.085	0.090	0.077	0.078	0.068	0.071	0.054	0.055
F-value for change in number of P3P policies		24640.478		1137.956		1171.800		832.363	
p-value for change in number of P3P policies		0.0001		0.0001		0.0001		0.0001	
Cohen's d effect size <sup>1</sup> for change in number of P3P policies		2.22		0.48		0.41		0.48	
Std error of the effect size estimate		0.02		0.01		0.01		0.01	
Confidence interval	Lower	2.18		0.45		0.46		0.38	
	Upper	2.25		0.51		0.51		0.44	

NOTE: 10,000 populations were created for all domains.

TABLE 9  
Percentage of P3P Policies Containing Structural Errors

		Feb-05	Nov-05	p-value
Alexa Global 500	Full	20.78%	22.78%	0.8470
	Compact	40.00%	35.29%	0.5894
BBBOnline	Full	60.55%	69.29%	0.0001
	Compact	13.51%	10.24%	0.0378
FirstGov	Full	9.88%	13.08%	0.6478
	Compact	27.27%	17.64%	0.6525
Forbes 500	Full	29.17%	25.93%	1.0000
	Compact	8.33%	7.14%	1.0000
Top 300 E-Commerce	Full	27.59%	33.33%	0.5102
	Compact	12.36%	12.20%	1.0000
Truste	Full	26.56%	29.21%	0.5760
	Compact	33.56%	30.87%	0.7101

where the Web site's functionality allows for the disseminating of company information and/or directly transacting with customers. For analysis of sites that are explicitly transaction based, refer to our analysis in Table 6 regarding Top 300 E-commerce. Our final group is the "Seal Adopting" group (BBBOnline and Truste).

Table 8 indicates that the only group that experienced a significant operational alteration ( $d > 0.8$ , large) was the "All Lists" group; this group experiences a significant increase in adoption between the control (February 2005) and treatment (November 2005) groups. The confidence intervals indicate that this large effect contains very little random variation due to experimental effects or sample limitations and hence can be considered a highly reliable figure. However, when we look at the Nonlegislative group, we see that the operational alteration decreased substantially to a small/medium ( $0.2 < d < 0.5$ ) effect indicating that much of the effect in the "All Lists" group is attributable to the Web sites within the FirstGov list. The other groups experience effect sizes broadly inline with this "Nonlegislative" group and, hence, we might hypothesize that outside of the (US) governmental sector, P3P implementation rates were relatively stable throughout the study. Confidence intervals for these groups indicate that these results are reasonably robust to experimental variation.

## 6 P3P DOCUMENT ERRORS/MAINTENANCE BETWEEN FEBRUARY 2005 AND NOVEMBER 2005

While completing the analysis in Sections 4 and 5, a large number of errors were encountered in all lists (Table 9). The errors summarized in Table 9 only include errors that are strict violations of the P3P XML schema or basic XML document structure; this does not include inconsistencies between P3P policies and their human-readable counterparts. The analysis of inconsistencies between P3P policies and their human-readable counterparts is a complex task requiring extensive knowledge of applicable legislation, case law, and expert opinion, which is beyond the scope of this paper. Although Table 9 only indicates the prevalence of structural errors, no statistically significant changes were detected except in the case of the BBBOnline group. Further

ad hoc analysis indicates that the BBBOnline group experienced an increase in the error rates for full policies but a decrease in the error rates in compact policies between February and November 2005. For full policies, the operational effect of the statistically significant difference was small/medium ( $d = 0.212$ ). For compact policies, the statistically significant difference was found to have a small operational effect ( $d = 0.173$ ), suggesting that little change occurred over the time period.

It is interesting to note that the prevalence of errors can vary widely between full and compact policies. It is especially surprising that P3P compact policies were more likely to be in error than full policies for the Alexa Global 500, FirstGov, and Truste groups. Compact policies are simpler (a string of text) and do not need to conform to stringent XML formatting rules. We are currently at a loss to explain these results.

Given the relatively large number of policies that are invalid (Table 9), a more detailed yet rigorous analysis of these errors is conducted (Tables 10, 11, and 12). Table 10 describes how P3P full and compact policies are being adopted and whether they contain errors. Table 11 describes the extent that noncorrective maintenance is being undertaken and, finally, Table 12 describes the extent that corrective maintenance is undertaken. The largest improvement in full P3P policies occurred in the Forbes 500 list, where one out of seven invalid policies were improved (Table 12), and from this result, it would appear that few invalid full or compact P3P policies are being corrected. In addition to the lack of corrective maintenance, we detected very little noncorrective maintenance of P3P policies (Table 11). A P3P policy (full or compact) is deemed to have undergone noncorrective maintenance if any information elements in the document changed while remaining valid in both surveys. The low frequency of noncorrective maintenance precludes any application of statistical tests in identifying major changes. Through a visual inspection of the modifications, it was found that many of the changes involved the types of data collected, purpose for collection, and length of retention information. It would appear reasonable to expect such changes as business practices within organizations change.

Due to space limitations, we reserve a rigorous analysis of P3P error types for our future work. The nature of XML documents complicates any analysis of errors since violations can occur both in the XML document structure and in the XML schema that governs policy content. When errors occur in the XML document structure, errors occurring later in the document are possibly masked since the syntactic interpretation of the document may change. Similarly, violations of the XML schema governing the content of policies span a wide range of potential issues that would require a significant space to rigorously describe.

In order to generalize our findings, resampling techniques using the groups defined in Table 8 were again employed to analyze the occurrence of *valid* (Table 13) and *invalid* (Table 14) full policies. Table 13 indicates that the "All Lists" group experienced a large ( $d > 0.8$ ) increase in *valid* full policy adoption, which can be stated with a high degree of confidence (lower confidence interval = 1.29). In contrast, all other groups experienced either small ( $d > -0.2$ ) ("Seal Adopting" Web sites) or small/medium ( $-0.5 < d < -0.2$ ) decreases in *valid* full policy adoption.

TABLE 10  
Total Number of Sites with Valid, Invalid, and No P3P Full/Compact Policies as of February 2005

Status of Full Policy	Status of Compact Policy	Alexa Global 500	BBBOnline	FirstGov	Forbes 500	Top 300 E-Commerce	Truste
Valid	Valid	20 (4.2%)	112 (0.6%)	7 (1.9%)	5 (1.0%)	39 (13.1%)	41 (3.2%)
Valid	Invalid	17 (3.5%)	8 (0.0%)	1 (0.3%)	1 (0.2%)	2 (0.7%)	29 (2.3%)
Valid	None	24 (5.0%)	213 (1.1%)	65 (17.8%)	11 (2.2%)	22 (7.4%)	71 (5.5%)
Invalid	Valid	12 (2.5%)	327 (1.7%)	0 (0.0%)	1 (0.2%)	18 (6.0%)	21 (1.6%)
Invalid	Invalid	0 (0.0%)	11 (0.0%)	1 (0.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Invalid	None	4 (0.83%)	173 (0.9%)	7 (1.9%)	6 (1.2%)	6 (2.0%)	30 (2.3%)
None	Valid	14 (2.9%)	227 (1.2%)	1 (0.3%)	5 (1.0%)	21 (7.0%)	37 (2.9%)
None	Invalid	7 (1.5%)	85 (0.4%)	1 (0.3%)	0 (0.0%)	9 (3.1%)	21 (1.6%)
None	None	382 (79.6%)	18166 (94.0%)	283 (77.3%)	464 (0.94%)	181 (60.7%)	1027 (80.4%)

For example, "Nonlegislated" ( $d = -0.29$ ) and "Business" ( $d = -0.35$ ) groups experienced small to medium decreases. Again, these results can be stated with a 95 percent certainty due to the small range of the confidence interval. We are at a loss to explain why full policy adoption is so low for "Seal Adopting" sites when they have shown a previous inclination to adopt technologies with a similar purpose. We can only speculate that this may be due to a lack of requests for such documents from Web site patrons.

The results in Table 14 (*invalid* full policy adoption) indicate that a medium/large ( $0.5 < d < 0.8$ ) increase of invalid full policies occurred in "Seal Adopting" Web sites ( $d = 0.68$ ). However, in the cases of "All Lists," "Nonlegislated," and "Business" groups, large ( $d > 0.8$ ) increases in *invalid* full policy adoption were found with a high degree of confidence. We are at a loss to explain these error rates due to the free and publicly available W3C P3P validator tool, which analyzes P3P policies for structural errors. This effect has some interesting parallels with recent surveys conducted by Tien et al. [60] and Huynh and Miller [61] who investigate defect rates found on Web servers. Both surveys show that the defects are dominated by 404 errors, which are primarily missing hypertext links. Again, these error types are completely preventable as "link

checkers," which can automatically discover these errors, are widely available. Again, both surveys were unable to provide a causal explanation as to why these defects remain uncorrected.

TABLE 12  
Number of Sites Modifying, Adopting, or Dropping Their P3P-Enabled Status

	Alexa Global 500	BBBOnline	FirstGov	Forbes 500	Top 300 E-Commerce	Truste
<b>Full Policies</b>						
Invalidated their policy (Valid $\rightarrow$ Invalid)	1	4	4	0	3	1
Fixed their policy (Invalid $\rightarrow$ Valid)	1	1	1	1	1	3
<b>Compact Policies</b>						
Invalidated their policy (Valid $\rightarrow$ Invalid)	0	2	0	0	1	0
Fixed their policy (Invalid $\rightarrow$ Valid)	1	5	0	0	0	2
Dropped a valid policy (Valid $\rightarrow$ Dropped)	6	58	1	2	7	9
Dropped an invalid policy (Invalid $\rightarrow$ Dropped)	2	44	0	0	0	2
Created a valid policy (No P3P $\rightarrow$ Valid)	6	39	26	4	4	9
Created an invalid policy (No P3P $\rightarrow$ Invalid)	4	222	4	1	4	8

TABLE 11  
Number of Sites Performing Noncorrective Maintenance of P3P Policies

	Alexa Global 500	BBBOnline	FirstGov	Forbes 500	Top 300 E-Commerce	Truste
Total number of P3P enabled sites in both Feb and Nov with valid P3P policies	54	261	66	15	53	131
Number of valid full or compact policies that were modified and remained valid.	0	5	5	0	1	2

TABLE 13  
Analysis of *Valid* P3P Policy Adoption in Various Internet Domains Using Resampling Techniques

	All Lists		Non-legislated		Business		Seal adopting	
	Feb 05	Nov 05	Feb 05	Nov 05	Feb 05	Nov 05	Feb 05	Nov 05
Number of generated populations	10000		10000		10000		10000	
Mean number of Valid P3P policies	133.89	144.08	94.24	92.35	72.73	70.71	25.44	25.32
Standard Deviation	7.51	7.87	6.54	6.58	5.72	5.81	4.45	4.43
Standard Errors	0.075	0.079	0.065	0.066	0.057	0.058	0.044	0.044
F-value for change in number of valid P3P policies	8777.429		416.116		615.732		3.470	
p-value for change in number of valid P3P policies	0.0001		0.0001		0.0001		0.063	
Cohen's d effect size for change in number of valid P3P policies	1.32		-0.29		-0.35		-0.03	
Standard error of the effect size estimate	0.02		0.01		0.01		0.01	
Confidence interval	Lower		-0.32		-0.38		-0.05	
	Upper		1.29		-0.26		-0.32	

TABLE 14  
Analysis of *Invalid* P3P Policy Adoption in Various Internet Domains Using Resampling Techniques

	All Lists		Non-legislated		Business		Seal adopting	
	Feb 05	Nov 05	Feb 05	Nov 05	Feb 05	Nov 05	Feb 05	Nov 05
Number of generated populations	10000		10000		10000		10000	
Mean number of Invalid P3P policies	41.57	50.72	37.20	42.78	31.02	36.39	12.54	14.89
Standard Deviation	4.71	5.05	4.51	4.73	4.18	4.48	3.30	3.58
Standard Errors	0.047	0.051	0.045	0.047	0.042	0.045	0.033	0.036
F-value for change in number of invalid P3P policies	17479.022		7285.113		7706.579		2313.170	
p-value for change in number of invalid P3P policies	0.0001		0.0001		0.0001		0.0001	
Cohen's d effect size for change in number of valid P3P policies	1.87		1.21		1.20		0.68	
Std error of the effect size estimate	0.02		0.02		0.02		0.01	
Confidence interval	Lower		1.84		1.18		1.17	
	Upper		1.91		1.24		1.23	

The results in Tables 13 and 14 indicate that the increases in P3P adoption that were observed in Table 8 were primarily a result of invalid P3P policy adoption, with the exception of the sites within the FirstGov list. We, in fact, have found a decrease in the deployment of valid P3P policies in all groups that did not include the FirstGov list.

in these results is enhanced as there is no statistical evidence indicating changes in P3P adoption in any of the lists between the February and November 2005 surveys.

These observed differences in P3P adoption could also conceivably arise due to different cultural or legal influences that could motivate Web site operators to more readily adopt P3P rather than others. In order to undertake such an analysis, the nation each Web site in the Alexa Language list is hosted from was identified through reverse Internet Protocol (IP) lookups using the Linux "host" program. Once retrieved, the IP addresses were compared to a database purchased from IP2Location [62], which maps IP addresses to a particular nation. IP2Location states that their accuracy is above 95 percent. Through this method, it was determined that the Web sites contained in the Alexa Language lists originated from 49 nations. Some nations such as the US host Web sites in a variety of languages. For instance, 84 English, 47 Arabic, 26 Greek, and 26 Spanish Web sites are hosted from the US. Web sites using these languages constitute 60 percent of all Web sites hosted from the US. Additionally, since many nations have few Web sites being hosted from them, a Pareto analysis with a cutoff of 90 percent was used to reduce the number of nations under analysis, resulting in nations with fewer than 30 Web

## 7 P3P ADOPTION ACROSS LANGUAGES AND JURISDICTIONS

The Alexa Language lists provide an opportunity to determine if full policy adoption is highly influenced by changes in language or jurisdiction. Until recently, the P3P specification was only available in English, potentially limiting P3P adoption. Additionally, since P3P relies upon existing legislation for enforcement and the actions Web sites can undertake may be limited by such legislation, it would appear reasonable to expect that P3P usage should vary between jurisdictions.

Fig. 1 describes how differences in full policy adoption exist between the 100 most popular Web sites for 20 languages as of November 2005. Statistically significant differences exist between English-language sites and all other languages except Dutch and French. Our confidence

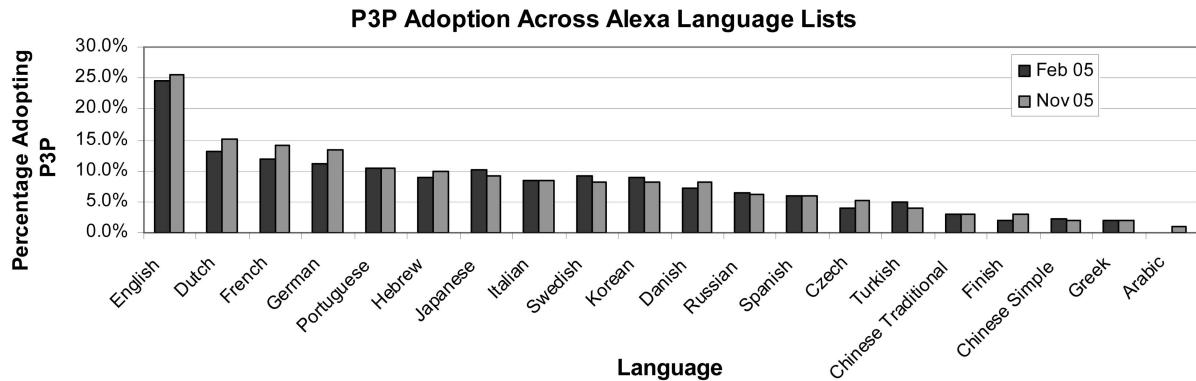


Fig. 1. P3P adoption by language.

sites being discarded. The results in Table 15 indicate that P3P adoption varies between jurisdictions, and although this analysis cannot exclude language as an explanatory factor in P3P adoption, it does suggest that P3P is not well suited to particular legal or cultural domains since its adoption ranges from 32.3 percent in the United Kingdom to 0.00 percent in China. These results suggest that there may be a need for the application of culturally sensitive design practices [63] in any future modifications of the P3P protocol.

## 8 P3P ADOPTION AND ERRORS VERSUS POPULARITY

After the pilot survey was completed, our results indicated that P3P adoption might be biased toward popular Web

sites (the AT&T survey found that P3P adoption increases with popularity). In addition, we were curious to see if the observed error rates continued in the general Internet population. As of October 2005, Alexa provided their rankings of the top 100,000 Web sites on the Internet. The addition of this list allows us to answer these questions regarding P3P adoption and errors. Figs. 2 and 3 represent a centile-by-centile analysis of P3P full and compact policy adoption in the Alexa 100,000 list. Each centile is composed of 1,000 Web sites grouped by rank. Caution is urged when interpreting the results in Figs. 2 and 3. A correct portrayal of the results would require a bar or scatter plot graph since the values derived from the centiles are discrete events. We chose to portray the graph as a line graph to ensure readability for the reader. The analysis in Fig. 2 indicates that P3P full policy adoption is correlated with Web site popularity with a nonlinear relationship. P3P adoption rates start at approximately 15 percent for the most popular sites. Adoption then drops to a relatively consistent rate of about 3 percent for less popular Web sites. Errors do not appear to follow the same nonlinear relationship of general P3P adoption but instead follow a relatively consistent rate of about 1 percent of Web sites.

Fig. 3 describes the adoption of P3P compact policies. In general, the adoption curve of P3P compact policies follows

TABLE 15  
Analysis of P3P Policy Adoption by Country

	Number of sites available		Number adopting p3p		Percentage adopting P3P	
	Feb	Nov	Feb	Nov	Feb	Nov
United States	297	298	34	38	11.4%	12.8%
Sweden	102	101	10	9	9.8%	8.9%
South Korea	94	93	8	7	8.5%	7.5%
Israel	90	90	6	7	6.7%	7.8%
Czech Republic	88	87	4	5	4.5%	5.7%
Finland	88	87	2	3	2.3%	3.4%
Japan	88	88	9	8	10.2%	9.1%
China	85	94	0	0	0.0%	0.0%
Denmark	83	82	6	7	7.2%	8.5%
Turkey	83	83	4	3	4.8%	3.6%
Russian Federation	79	79	5	5	6.3%	6.3%
Germany	77	77	10	12	13.0%	15.6%
Italy	75	76	5	4	6.7%	5.3%
Hong Kong	65	64	1	1	1.5%	1.6%
France	63	64	6	7	9.5%	10.9%
Greece	61	60	2	2	3.3%	3.3%
Canada	57	57	5	6	8.8%	10.5%
Brazil	57	58	4	5	7.0%	8.6%
Spain	34	34	2	2	5.9%	5.9%
United Kingdom	32	31	11	10	34.4%	32.3%

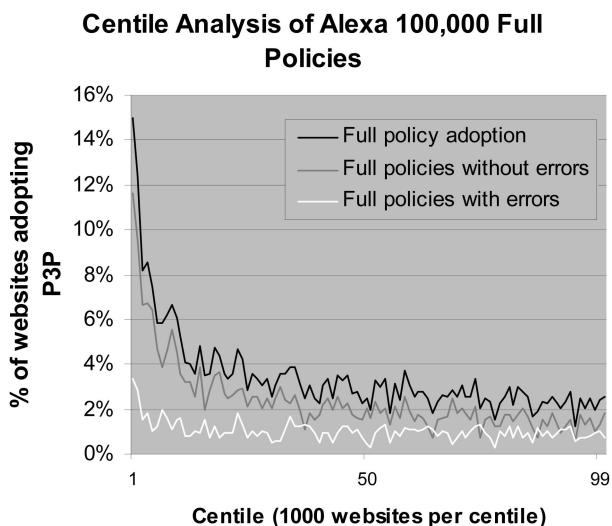


Fig. 2. Full policy adoption for Alexa 100,000 Web sites.

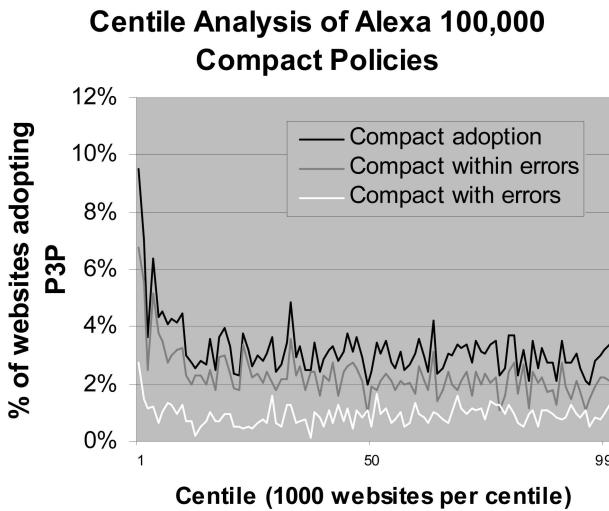


Fig. 3. Compact policy adoption for Alexa 100,000 Web sites.

the adoption of P3P full policies with the only major deviation occurring in the most popular Web sites on the Internet. This divergence in P3P full/compact policy adoption suggests that the motivation to adopt P3P may be different between the most popular Web sites on the Internet and the remainder.

## 9 CONCLUSIONS AND FUTURE WORK

Our results paint a sometimes contradictory and concerning image of P3P adoption. Evidence exists indicating that specific Internet domains have seen large increases in P3P adoption (FirstGov and Froogle/Top 400 E-commerce) since the 2003 AT&T survey. However, only FirstGov continues to show increasing adoption of valid P3P policies between February 2005 and November 2005. Although this increase is a positive indicator for P3P, one has to wonder why more US government Web sites have not adopted P3P, given that they have had more than three years to adopt P3P since the enacting of the E-governance Act. To our knowledge, there is no other suitable substitute that would satisfy the requirements of the E-governance Act. It would appear reasonable to argue that P3P adoption would increase if P3P had been explicitly named in the E-governance Act. Our data set is, however, insufficient to allow us to comment on the effects of naming such a technology in legislation.

Our results also bring into question whether P3P can exist in a self-regulatory environment. The analysis of nonlegislated Web sites indicates a decrease in valid P3P policies and an increase of invalid P3P policies. When one takes into account the lack of corrective maintenance of invalid P3P full and compact policies, it appears that companies have little incentive to provide quality documents in a self-regulated environment. This is especially surprising in the case of P3P compact policies since it would appear to be in the organization's best interest to ensure that third-party cookies are not blocked. The only domain that experienced an increase in valid full P3P policies was the FirstGov list, which is a legislated domain.

The lack of observed maintenance should also bring into question the information contained in P3P full and compact documents. There is little evidence indicating that Web sites

ever fix their erroneous policies. In addition, there appears to be little or no effort to update P3P full and compact policies over a nine-month time span. It would seem unreasonable to think that business practices change so infrequently for Internet companies.

Beatty et al. [64] have recently analyzed P3P adoption on e-commerce Web sites using Roger's technology adoption factors as an analytical framework. In this paper, the authors concluded that P3P user agents have poor triability and observability and are incompatible with many user's past experiences. The results of this survey suggest that this analysis needs to be extended to include adoption factors from an organizational viewpoint.

The usability of P3P also appears questionable. AT&T Privacy Bird users find the system to be of questionable value given the current adoption rates. Our results indicate that this is unlikely to change since the only identified growth in usable P3P policies occurred in a legislated domain. To our knowledge, the only domain of Web sites that are legislated to provide machine-readable privacy policies are those belonging to the US federal and state governments. Without a significant increase in P3P adoption, the usability of these agents from an end-user perspective appears questionable at best.

P3P 1.1 may have the potential to remedy some of these issues, as we have pointed out. However, W3C has decided not to pursue candidate recommendation status at this time. Their rationale is that there is "insufficient support" from Web browser implementers for developing P3P 1.1 user agents [14]. This determination appears to reflect our findings that P3P adoption has stagnated in a niche position; it appears that browser implementers simply do not have enough market incentive to expend the resources needed to develop and integrate P3P 1.1 user agents. Furthermore, the additional data elements of P3P 1.1 are at best a mixed blessing, given the high error rates in P3P 1.0 documents.

This research is limited by its concentration only upon the server-side adoption of the P3P protocol. Although information describing the extent that users are requesting these policies and the agents they use would undoubtedly provide insight, such an analysis is beyond the scope of this paper and would require a large sample of Internet users covering large cultural and geographic distances.

In conclusion, our results appear to indicate that P3P is the Internet's privacy standard in name only, since the vast majority of Internet Web sites lack valid P3P policies. At this point in time, we would have to conclude that P3P offers little assistance to the user. However, if legislation similar to the E-governance Act requiring a machine-readable form of the privacy policy were to be enacted over a broad section of the Internet, adoption of valid P3P policies might increase. If this occurs, P3P agent usefulness may improve.

Future work we plan to undertake in this area includes a more comprehensive analysis of errors in full and compact policies. If it is determined that the errors were the result of poor policy development tools or issues with the P3P schema, the potential exists to improve the protocol either through tool development (for example, the JRC Policy Editor) or future versions of the protocol itself. Further, if technologies such as P3P, which rely upon users to make reasonable choices regarding their privacy, are to succeed, researchers must address the

findings of AcQusiti and Grossklags [31] and Spiekermann et al. [32] who found that users often trade long-term privacy for short-term benefits. If users can be persuaded to make reasonable choices, then technologies such as P3P may prove valuable. Another improvement that may be required is the development of user agents that continue to function even if P3P policies are not widely available. This might be accomplished through the collection of other observable server characteristics/documents or the analysis of third-party information such as blacklists and reputation mechanisms.

It would also be of value to determine the extent that Web sites adhere to the privacy principles adopted by their respective jurisdictions. Such an analysis could conceivably be accomplished through the application of tools such as the JRC EU legislation analysis tool [41]. This tool could also be configured to allow for testing of other groups of Web sites such as those belonging to the US safe harbor program or those belonging to other nations that have adopted privacy legislation such as Canada and Japan. Such an analysis may prove to be highly valuable since a growing body of evidence indicates that adherence to these laws/programs is questionable at best [65], [66], [67]. If such an analysis is feasible, then P3P or a derivative of P3P may provide a significant value as a tool for governments and organizations to remotely determine adherence to guidelines and legislation.

The analysis of P3P policies posted by organizations dealing with particularly sensitive information may also prove valuable. If members of the financial, health care, and data mining industries could be reliably identified, P3P could be used as a mechanism for remote analysis of their data-handling practices, thereby aiding public policy decisions.

Finally, our results also suggest that the adoption of P3P may vary across cultural lines. If supported, this finding suggests that the design and development of P3P must begin to take into account cultural perceptions of privacy if they are to be implemented across the Internet.

## ACKNOWLEDGMENTS

The authors wish to thank Alexa.com for graciously providing them with a copy of their Top 100,000 list. This research was supported in part by NSERC under Grant G121210906 and by Alberta Innovation and Science under Grant G230000066.

## REFERENCES

- [1] H. Kargupta, A. Joshi, K. Sivakumar, and Y. Yesha, *Data Mining: Next Generation Challenges and Future Directions*. MIT Press, 2004.
- [2] B. Malin, "Betrayed by My Shadow: Learning Data Identity via Trial Matching," *J. Privacy Technology*, 2004.
- [3] H. Nissenbaum, "Protecting Privacy in an Information Age: The Problems of Privacy in Public," *Law and Philosophy*, vol. 17, pp. 559-596, 1998.
- [4] J. Turow, L. Feldman, and K. Meltzer, "Open to Exploitation: American Shoppers Online and Offline," Annenberg Public Policy Center, Univ. of Pennsylvania, June 2005, [http://www.annenbergpublicpolicycenter.org/Downloads/Information\\_And\\_Society/Turow\\_APPC\\_Report\\_WEB\\_FINAL.pdf](http://www.annenbergpublicpolicycenter.org/Downloads/Information_And_Society/Turow_APPC_Report_WEB_FINAL.pdf).
- [5] A.I. Antón, Q. He, and D.L. Baumer, "Inside JetBlue's Privacy Policy Violations," *IEEE Security and Privacy*, vol. 2, 2004.
- [6] J. Rich, "Internet Service Provider Settles FTC Privacy Charges," Fed. Trade Commission, <http://www.ftc.gov/opa/2005/03/cartmanager.htm>, 2005.
- [7] "On the Data Trail: How Detailed Information about You Gets into the Hands of Organizations with Whom You Have No Relationship," The Canadian Internet Policy and Public Interest Clinic, Apr. 2006.
- [8] H. Black, "On-Line Data Brokers," Office of the Privacy Commissioner of Canada, [http://www.privcom.gc.ca/legislation/let/let\\_051118\\_e.asp](http://www.privcom.gc.ca/legislation/let/let_051118_e.asp), 2005.
- [9] J. Rich, "ChoicePoint Settles Data Security Breach Charges; to Pay \$10 Million in Civil Penalties, \$5 Million for Consumer Redress," Fed. Trade Commission, <http://www.ftc.gov/opa/2006/01/choicepoint.htm>, 2006.
- [10] "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data," [http://www.oecd.org/document/18/0,2340,en\\_2649\\_34255\\_1815186\\_1\\_1\\_1,00.html](http://www.oecd.org/document/18/0,2340,en_2649_34255_1815186_1_1_1,00.html), 1980.
- [11] M.K. Landesberg, T.M. Levin, C.G. Curtin, and O. Lev, "Privacy Online: A Report to Congress," Fed. Trade Commission, <http://www.ftc.gov/reports/privacy3/toc.htm>, 1998.
- [12] "Safe Harbor Privacy Principles," US Dept. of Commerce, <http://www.export.gov/safeHarbor/SHPRLNCIPLESFINAL.htm>, 2000.
- [13] "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data," *Official J. L No. 281*, 23 Nov. 1995.
- [14] "The Platform for Privacy Preferences (P3P) Project," <http://www.w3.org/P3P/>, 2006.
- [15] P. Resnick and J. Miller, "PICS: Internet Access Controls without Censorship," *Comm. ACM*, vol. 39, pp. 87-93, 1996.
- [16] C. Jensen and C. Potts, "Privacy Policies as Decision-Making Tools: An Evaluation of Online Privacy Notices," *Proc. Conf. Human Factors in Computing Systems (CHI '04)*, 2004.
- [17] "OASIS eXtensible Access Control Markup Language (XACML) TC," OASIS, [http://www.oasis-open.org/committees/tc\\_home.php?wg\\_abbrev=xacml](http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xacml), 2006.
- [18] M. Schunter and C. Powers, "The Enterprise Privacy Authorization Language (EPAL 1.1)," IBM, <http://www.zurich.ibm.com/security/enterprise-privacy/epal/>, 2003.
- [19] H. Hochheimer, "The Platform for Privacy Preferences as a Social Protocol: An Examination within the U.S. Policy Context," *ACM Trans. Internet Technology*, vol. 2, pp. 276-306, 2002.
- [20] JAP Team, "JAP—Anonymity and Privacy," [http://anon.inf.tu-dresden.de/index\\_en.html](http://anon.inf.tu-dresden.de/index_en.html), 2007.
- [21] "Tor: Anonymity Online," Electronic Frontier Foundation, <http://tor.eff.org/>, 2007.
- [22] M.K. Reiter and A.D. Rubin, "Anonymous Web Transactions with Crowds," *Comm. ACM*, vol. 42, pp. 32-48, 1999.
- [23] A.I. Antón, J.B. Earp, Q. He, W. Stufflebeam, D. Bolchini, and C. Jensen, "Financial Privacy Policies and the Need for Standardization," *IEEE Security and Privacy*, vol. 2, pp. 36-45, 2004.
- [24] C. Jensen, C. Potts, and C. Jensen, "Privacy Practices of Internet Users: Self-Reports versus Observed Behavior," *Int'l J. Human Computer Studies*, vol. 63, pp. 203-227, 2005.
- [25] L.F. Cranor, P. Gunuru, and M. Arjula, "User Interfaces for Privacy Agents," *ACM Trans. Computer-Human Interaction*, 2006.
- [26] T. Ledacky, H. Dhurvasula, and J. Loeb, "The Platform for Privacy Preferences (P3P)," Mozilla, <http://www.mozilla.org/projects/p3p/>, 2006.
- [27] "Pretty Poor Privacy: An Assessment of P3P and Internet Privacy," Electronic Privacy Information Center, <http://www.epic.org/reports/prettypoorprivacy.html>, 2000.
- [28] K. Lee and G. Speyer, *Platform for Privacy Preferences Project (P3P) and Citibank*. Citibank Advanced Development Group, 1998.
- [29] R. Thibadeau, "A Critique of P3P: Privacy on the Web," <http://dollar.ecom.cmu.edu/p3pcritique/>, 2000.
- [30] "Privacy on the Internet—An Integrated EU Approach to On-line Data Protection," Article 29—Data Protection Working Party, European Commission 5063/00/EN/FINAL WP 37, Nov. 21, 2000.
- [31] A. AcQuisti and J. Grossklags, "Privacy and Rationality in Individual Decision Making," *IEEE Security and Privacy*, vol. 3, pp. 26-33, 2005.
- [32] S. Spiekermann, J. Grossklags, and B. Berendt, "E-Privacy in 2nd Generation E-Commerce: Privacy Preferences versus Actual Behavior," *Proc. Third ACM Conf. Electronic Commerce*, 2001.
- [33] T. Yu, N. Li, and A.I. Antón, "A Formal Semantics for P3P," *Proc. ACM Workshop Secure Web Service (SWS '04)*, 2004.

- [34] J. Linn, "Technology and Web User Data Privacy—A Survey of Risks and Countermeasures," *IEEE Security and Privacy*, vol. 3, pp. 52-58, 2005.
- [35] L. Cranor, B. Dobbs, S. Egelman, G. Hogben, M. Langheinrich, M. Marchiori, M. Presler-Marshall, J. Reagle, M. Schunter, D.A. Stampy, and R. Wenning, *The Platform for Privacy Preferences 1.1 (P3P 1.1) Specification*, World Wide Web Consortium (W3C) note, <http://www.w3.org/TR/2006/WD-P3P11-20060210/Overview.html>, 2006.
- [36] L.F. Cranor, S. Byers, and D. Kormann, "An Analysis of P3P Deployment on Commercial, Government and Children's Web Sites as of May 2003," trade report, Fed. Trade Commission, May 2003.
- [37] S. Byers, L.F. Cranor, and D. Kormann, "Automated Analysis of P3P-Enabled Web Sites," *Proc. Fifth Int'l Conf. Electronic Commerce*, 2003.
- [38] L. Cranor, M. Langheinrich, M. Marchiori, M. Presler-Marshall, and J. Reagle, *The Platform for Privacy Preferences 1.0 (P3P 1.0) Specification*, World Wide Web Consortium (W3C) recommendation, <http://www.w3.org/TR/P3P/>, 2002.
- [39] "Privacy Bird," <http://www.privacybird.com/>, 2006.
- [40] F. Arshad and S. Sheng, "Privacy Fox," <http://privacyfox.mozdev.org/>, 2006.
- [41] JRC P3P Resource Centre, Joint Research Center, <http://p3p.jrc.it/>, 2006.
- [42] L. Cranor, M. Langheinrich, and M. Marchiori, *A P3P Preference Exchange Language 1.0 (APPEL 1.0)*, World Wide Web Consortium (W3C) note, <http://www.w3.org/TR/P3P-preferences/>, 2002.
- [43] D. Julta and Y. Zhang, "Maturing E-Privacy with P3P and Context Agents," *Proc. IEEE Int'l Conf. E-technology, E-commerce, and E-services*, 2005.
- [44] Alexa Traffic Rankings, <http://www.alexa.com/>, 2006.
- [45] FirstGov, <http://www.firstgov.gov/>, 2006.
- [46] "Internet Retailer Top 400 Guide," <http://www.internetretailer.com/top400/>, 2005.
- [47] "Internet Retailer Top 300 Guide," <http://www.internetretailer.com/top300/>, 2004.
- [48] BBBOnline, Better Business Bureau, <http://www.bbbonline.org/>, 2006.
- [49] Truste, <http://www.truste.org/>, 2006.
- [50] "Forbes 2000 Largest Public Companies," [http://www.forbes.com/2006/03/29/06f2k\\_worlds-largest-public-companies\\_land.html](http://www.forbes.com/2006/03/29/06f2k_worlds-largest-public-companies_land.html), 2006.
- [51] P3P Validator, <http://www.w3.org/P3P/validator.html>, 2002.
- [52] J.T. Casagrande, M.C. Pike, and P.G. Smith, "An Improved Approximate Formula for Calculating Sample Sizes for Comparing Two Binomial Distributions," *Biometrics*, vol. 34, pp. 483-486, 1978.
- [53] P.I. Good, *Resampling Methods: A Practical Guide to Data Analysis*, third ed. Birkhauser, 2006.
- [54] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, second ed. Lawrence Earlbaum Assoc., 1988.
- [55] J. Miller, "Statistical Significance Testing: A Panacea for Software Technology Experiments," *J. Systems and Software*, vol. 73, pp. 183-192, 2004.
- [56] M.W. Lipsey and D.B. Wilson, *Practical Meta-Analysis*, vol. 49. Sage Publications, 2001.
- [57] V. Hasselblad and L.V. Hedges, "Meta-Analysis of Screening and Diagnostic Tests," *Psychological Bull.*, vol. 117, pp. 167-178, 1995.
- [58] US E-government Act of 2002, Public Law No. 107-347-DEC. 17 2002, 2002.
- [59] J. Copas and D. Jackson, "A Bound for Publication Bias Based on the Fraction of Unpublished Studies," *Biometrics*, vol. 60, pp. 146-153, 2004.
- [60] J. Tian, S. Rudraraju, and L. Zhao, "Evaluating Web Software Reliability Based on Workload and Failure Data Extracted from Server Logs," *IEEE Trans. Software Eng.*, vol. 30, pp. 754-769, 2004.
- [61] T. Huynh and J. Miller, "Further Investigations into Evaluating Web Site Reliability," *Proc. Fourth Int'l Symp. Empirical Software Eng.*, 2005.
- [62] IP2Location, <http://www.ip2location.com/>, 2006.
- [63] N. Zakaria, J.M. Stanton, and S.T.M. Sarkar-Barney, "Designing and Implementing Culturally-Sensitive IT Applications," *Information Technology and People*, vol. 16, 2003.
- [64] P. Beatty, I. Reay, S. Dick, and J. Miller, "P3P Adoption on E-commerce Websites: A Survey and Analysis," *IEEE Internet Computing*, 2006.
- [65] N.E. Bowie and K. Jamal, "Privacy Rights on the Internet: Self-Regulation or Government Regulation," *Business Ethics Quarterly*, vol. 16, 2006.
- [66] "Compliance with Canadian Data Protection Laws: Are Retailers Measuring Up?" The Canadian Internet Policy and Public Interest Clinic, [http://www.cippic.ca/en/bulletin/compliance\\_report\\_06-07-06\\_\(color\)\\_cover-english.pdf](http://www.cippic.ca/en/bulletin/compliance_report_06-07-06_(color)_cover-english.pdf), 2006.
- [67] M. Markel, "Safe Harbor and Privacy Protection: A Looming Issue for IT Professionals," *IEEE Trans. Professional Comm.*, vol. 49, 2006.
- [68] S. Gritzalis, "Enhancing Web Privacy and Anonymity in the Digital Era," *Information Management and Computer Security*, vol. 12, no. 3, pp. 255-288, 2004.



**Ian K. Reay** received the BSc degree in computer engineering from the University of Alberta in 2004. He is currently working toward the MSc degree at the University of Alberta and is a software developer with M-Tech Information Technology. His research interests lie in privacy preservation, computer security, and identity management systems.

**Patricia Beatty** received the BSc degree in computer engineering from the University of Alberta in 2004. She is currently pursuing the MSc degree, also in computer engineering, at the University of Alberta. Her research interests are trust and privacy in e-commerce systems.



**Scott Dick** received the BSc degree in 1997, the MSc degree in 1999, and the PhD degree in 2002, all from the University of South Florida (USF). His PhD dissertation received the USF Outstanding Dissertation Prize in 2003. He has been an assistant professor at the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alta, since 2002. His research interests are in fuzzy logic, machine learning, data mining, and their interdisciplinary application to software reliability and veterinary epidemiology. He has published more than 20 papers in refereed journals and conferences. He is a member of the ACM, the IEEE, the ASEE, and the Fuzzy Systems Technical Committee of the IEEE Computational Intelligence Society and an associate member of Sigma Xi.



**James Miller** received the BSc and PhD degrees in computer science from the University of Strathclyde, Scotland. During this period, he worked on the ESPRIT project GENEDIS on the production of a real-time stereovision system. Subsequently, he worked at the United Kingdom's National Electronic Research Initiative on Pattern Recognition as a principal scientist, before returning to the University of Strathclyde to accept a lectureship and, subsequently, a senior lectureship in computer science. Initially, during this period, his research interests were in computer vision, and he was a co-investigator on the ESPRIT 2 project VIDIMUS. Since 1993, his research interests have been in software and systems engineering. In 2000, he joined the Department of Electrical and Computer Engineering at the University of Alberta as a full professor and, in 2003, became an adjunct professor at the Department of Electrical and Computer Engineering, University of Calgary. He is the principal investigator in a number of research projects that investigate software verification, validation, and evaluation issues across various domains, including embedded, Web-based, and ubiquitous environments. He has published more than 100 refereed journal and conference papers on software and systems engineering (see [www.steam.ualberta.ca](http://www.steam.ualberta.ca) for details on recent directions), currently serves on the program committee for the IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), and sits on the editorial board of the *Journal of Empirical Software Engineering*.