Privacy in Pervasive Environments: Next Generation Labeling Protocols

MARK S. ACKERMAN

Department of Electrical Engineering and Computer Science and School of Information

University of Michigan

Ann Arbor, MI 48109, USA

ackerm @ umich.edu

Abstract. In pervasive environments, privacy is likely to be a major issue for users, and users will want to be notified of potential data capture. To provide notice to users, this paper argues for what it calls labeling protocols, technical mechanisms through which users can be informed of data requests and their consequences. Recent experiences with the Platform for Privacy Preferences (P3P) Project, an attempt to provide privacy mechanisms for the Web, suggest important lessons for the design of a next-generation labeling protocol that will be usable and useful in pervasive environments. This paper examines the P3P lessons and open issues with an eye to pervasive requirements.

Keywords. *privacy, pervasive environments, ubiquitous computing, Platform for Privacy Preferences, P3P, labeling protocols.*

Introduction

Next generation computational environments will be pervasive, ubiquitous, and perceptual. However, to gain wide-scale acceptance and adoption, a critical problem – privacy – must be ameliorated or solved.

In pervasive environments, privacy will become increasingly important and visible as a problem for users:

- Because of the envisioned large number of sensors, location-awareness, and fusion systems, the sheer volume of identifiable data will necessarily increase greatly.

- Pervasive environments, as envisioned, include perceptual interfaces that can recognize users and can detect pointing gestures and certain facial expressions. Export of image or acoustic information by independent agents in such an environment, especially when offered in an unfettered data marketplace, could allow third parties to recognize when a person was present in a situation.

- Pervasive environments offer the potential of tracking and capturing substantial portions of users' activities. Regardless of whether this is an actual threat, users' perceptions that such possibilities exist may deter them from using and adopting pervasive environments.

- Some of the envisioned outcomes of having this data; e.g., automatic configuration and personalized services, will be highly public and visible to users. Privacy failures will also be public and visible.

In short, users may feel increasingly unable to maintain their privacy. As a result, some have called privacy the killer threat to pervasive environments. Finding a solution to this real or perceived threat will be of considerable importance to the adoption of pervasive environments.

There are many ways to ameliorate the likely privacy problem in ubiquitous environments. For example, one could legally eliminate data transfers where the user was identifiable. However, this would remove a large number of potential applications. Another solution might be to make subsequent reuse of personal data illegal. This is unlikely in the current US political climate. Yet another way might be to provide users with the information they need about data requests and their consequences, and let them make informed decisions about when and how to release their personal data. For this to occur, the environment or the objects in the environment must somehow provide this information to the user. Technical mechanisms to facilitate these informed decisions will be critical in pervasive environments. In this paper, I will refer to a technical mechanism through which users can be informed of data requests and their consequences as a *labeling protocol*.

This paper's goal is to examine what will be required to construct usable and useful labeling protocols. By doing so, the paper will also construct an understanding of what will be required in pervasive environments to suitably facilitate individuals' privacy.

The paper proceeds as follows: After a brief examination of the social requirements in dealing with privacy, the paper will survey the various technical mechanisms that have been employed to date for facilitating privacy. One of these is the labeling protocol. The paper will go into some detail on the Platform for Privacy Preferences Project (P3P), the first and most significant example of a labeling protocol. While P3P is an attempt to find privacy mechanisms for the Web, its lessons will be important for next generation environments. Therefore, the paper first describes P3P as a technical mechanism. More importantly, it then discusses the lessons and issues in the P3P experience for future labeling protocol mechanisms. The paper concludes by examining these lessons in regard to pervasive environments and their new requirements.

The Social Side of Privacy

Before moving to technical mechanisms, it is important to first step back and examine what "privacy" is and how users conceptualize it. What is meant by "privacy"? A simple but useful definition of privacy is "the ability of an individual to control the terms under which their personal information is acquired and used [1]." As such, privacy is about individuals' capabilities in a particular social situation to control what they consider to be personal data. Loss of one's privacy when one does not expect it can be psychologically devastating; some people become incensed. Yet, an individual's privacy is always defined in conjunction with capabilities of others to transact business and even to control their own privacy, leading to an inherent tension. Privacy may have to be traded off in certain transactions, such as the access to credit or to maintain the quality of health care [2]. Indeed, societal needs may also transcend an individual's privacy concerns, as in the case of public health.

Computer users in general are quite concerned with their privacy. Most published reports are about Web or network use, but presumably concerns extend to other types of computer use as well. Fisher [3] reported "Forty-one percent of Web

buyers surveyed last year by Forrester Research of Cambridge, Mass., said they have contacted a site to be taken off their databases because they felt that the organization used their information unwisely. (pp. 20-21)." A Business Week/Harris Poll [4] found that over 40% of online shoppers were very concerned over the use of personal information, and 57% wanted some sort of laws regulating how personal information is collected and used. Culnan [1] argued that privacy concerns were a critical reason why people do not go online and why they provide false information online.

It should be noted that this concern about privacy is not unjustified. The 1998 FTC privacy study found that a majority of online businesses "had failed to adopt even the most fundamental elements of fair information practices. ([1], p. 8)." Indeed, relatively few consumers believe that they have much control over how personal information from online activity is used or sold by businesses [5]. Current business practices, consumer fears, and media pressure have combined to make consumers seriously concerned about privacy.

Tackling privacy, however, is not an easy matter. If nothing else, privacy discussions often turn heated very quickly: Some people consider privacy to be a fundamental right. For example, Davies [6] argues that it has become a squandered right. Others would create a marketplace for personal data. Still others argue against any form of privacy: Etzioni [7] argues that privacy is societally illegitimate. For the purposes of this paper, I will not favor any particular viewpoint, wishing to examine merely why privacy is a difficult but critical problem in pervasive environments. In this paper, both users and organizations may have legitimate but sometimes conflicting goals and agendas. This is in the nature of most societal issues.

More to the point, however, privacy is a hard problem because individuals wish to control their personal information in a very detailed and nuanced manner. Goffman [8] noted that people must control their presentation of self, their face, to others. People need to be able to control what others think of them, and find it disconcerting when they cannot. Even more, people find it disconcerting when the rules of everyday conduct appear to change, as they can with new technologies. In these situations, people may feel that they have been unfairly treated or that they have not received proper notice [1].

As a social requirement for system construction, it is important to note that individuals do not view the "privacy problem" uniformly. The population is segmented by types of concerns and degree of concern. First, people have differing types of concerns. Culnan and Armstrong [5] make the argument that people have two kinds of privacy concerns. First, they are concerned over unauthorized access to personal data from security breaches or the lack of internal controls. Second, people are concerned about the risk of secondary use; that is, the reuse of their personal data for unrelated purposes without their consent. This secondary use includes sharing with third parties who were not part of the original transaction. It also includes the aggregation of transaction data and other personal data to create a profile. Smith, Milberg, and Burke [9] raise two additional concerns: People have a generalized anxiety about personal data being collected, and people are also concerned over their inability to correct any errors.

Another way in which people differ is in their level of concern. Overall, the research literature describes a general anxiety and its extent, but there is some research providing more detail. A persistent finding is that it is useful to consider US consumers not as one homogenous group. Westin [10] found three separate groups: the marginally concerned, privacy fundamentalists, and the pragmatic majority. These groups differ significantly in their privacy preferences and attitudes. The marginally concerned group is mostly indifferent to privacy concerns; privacy fundamentalists, on the other hand, are quite uncompromising about their privacy. The majority of the US population, however, are members of the pragmatic majority. They are concerned about their privacy, but are willing to trade personal data for some benefit (e.g., customer service). These groupings have been consistent across studies (e.g., [11], [12]). (Spiekermann et al. divided the pragmatics into those who were concerned with revealing their identity and those who were more concerned about making their personal profiles available.) In Ackerman et al., these groups were 27% marginally concerned, 17% privacy fundamentalists, and 56% pragmatic majority. Spiekermann et al. noted a larger group of privacy fundamentalists and fewer marginally concerned in Germany. It should be noted that, despite these groupings, consumers still want adequate measures to protect their information from inappropriate sale, accidental leakage or loss, and deliberate attack [13]. Indeed, in Ackerman, Cranor, and Reagle [11],

the concerns of pragmatists were often significantly reduced by the presence of privacy protection measures such as privacy laws or privacy policies on Web sites

In summary, then, privacy has been shown to be important in long series of attitudinal surveys. While there is some discrepancy between people's reported preferences and their actual behavior, the vast majority of people wish to have protection. At a social-theoretic level, the "privacy problem" is important to people because of their strong need to control their presentation of self to others. Nonetheless, it is critical to note that people differed widely in their types and degrees of concern about their privacy. The majority weighs data release against the benefits, and presumably the consequences, of that release.

The following section surveys the various technical mechanisms that have been suggested to ameliorate these privacy concerns.

Privacy Technologies

A number of useful technical mechanisms have been developed for dealing with the "privacy problem". In this discussion, I bracket off the use of regulation and law, which has been argued as an adequate protection. Regulation and law are important design considerations, both as constraints and as enforcement mechanisms, and I will return to them below. Opinions vary considerably, but certainly a wide range of regulation has not kept up adequately with technical changes (see for example the problems with FTC regulation of media discussed in [14]). There is increasing sentiment that regulation is required for privacy protection (perhaps in combination with technical mechanisms), but regardless, the following technologies are still useful.

Technical mechanisms can be roughly broken into four broad categories. These categories include encryption and security mechanisms, anonymizing mechanisms, infrastructures, and labeling protocols. Some form of each is appropriate in pervasive environments, although existing examples are more for the Web and general information systems. The first three mechanisms will be discussed briefly in turn; this is followed with a lengthy discussion of labeling protocols and P3P in particular.

First, encryption and other security mechanisms provide some privacy capabilities. It must be noted that security is necessary but not sufficient for privacy. Even with the tightest security mechanisms, some disclosure will be required (e.g., to provide services to a specific person). On the other hand, one cannot control the dissemination and use of private data without secure transmission and storage. Therefore, security is necessary for privacy, but security is not sufficient to safeguard against subsequent use, to minimize the risk of sensor-based disclosure, or to reassure users.

The second broad category is the so-called Privacy Enhancement Technologies (PETs). These include a variety of anonymizing and de-identifying mechanisms. Anonymous and pseudo-anonymous remailers, or tools to hide the identity of an email's sender, include for example Mixmaster. These are often quite useful for reporting abuse, crime, or human rights violations, where one must cloak one's identity for safety. Another example, digital cash, affords users the capabilities of real cash in that it cannot be traced and does not leave a digital trace. One such form of digital cash has been ECash [15]. Digital cash, however, has not caught on. Non-traceable identifiers for Web use have included the AT&T Crowds [16] and the ZKS Freedom systems. For providing anonymity at the network layer, routing mechanisms such as Onion routing [17] have been proposed. Other PETs and anonymizing mechanisms are discussed in Cranor [18].

There are often situations where anonymity is not appropriate or disclosure is required. Anonymous remailers, for example, have been abused by people providing copyrighted materials (although this is hotly debated). Social cohesion and norms in e-communities are thought to be reduced by the presence of anonymous users. Therefore, a class of so-called Gentle PETs [19] offers pseudo-anonymity instead of anonymity, where users can be traced in case of illegal actions but generally are anonymous. Many, however, still have the same social issues. In addition, many times some disclosure is required or warranted – the user may wish a product shipped to her address, or the user may trust the recipient.

Recently, substantial research has been done on providing anonymity in location-aware services for pervasive environments. Systems such as Cricket [20] allow

users to determine their position without providing it to a location service. Other work has provided algorithms for cloaking location.

The third broad category consists of middleware layers to facilitate the construction of privacy-aware software systems. This line of research is relatively recent. Hong [21] is perhaps the most recent and most advanced. He proposes an architecture that combines aspects of data flow and blackboard architectures for sensor-based environments. Of particular interest is that each datum can be tagged with privacy preferences. Moreover, Hong's system supports what he terms optimistic privacy control, where privacy abuses are detected on the basis of log files. This allows open sharing, but with some protection. His system also includes pessimistic privacy control, or restriction on access, as well as mixed-initiative control, where users are asked for permission. The nature of Hong's privacy preference tags is not described in [21]; however, it appears only relatively simple preferences are currently supported. To my knowledge, no other infrastructure is as advanced.

The final privacy mechanism is one that I term here a labeling protocol. Again, in order for there to be informed consent, especially in situations where anonymity is impossible or not desired, some mechanism must exist to both describe and announce the required or preferred personal data, the data collection's scope, and the data collection's intended use and consequences. Labeling protocols, then, are required to provide a vocabulary for detailing what the collected personal data might be and potentially to announce their collection or intended collection. If such labeling protocols will be an important part of pervasive privacy solutions, it is important to discuss in depth the major attempt at a labeling protocol, the P3P project [18, 22]. The next section does so.

P3P: A Labeling Protocol

The Platform for Privacy Preferences Project (P3P) by the World Wide Web Consortium (W3C) is the major labeling protocol to date. In this section, the concern will be to discuss how P3P works and what lessons and open issues P3P offers to future privacy solutions. This technical overview will be followed by a discussion of P3P's lessons.

P3P is an ongoing effort to create a privacy standard for the Web. It was initially touted as a large-scale solution to privacy for e-commerce and other Web services. Now, P3P is best considered as attempting to be a partial privacy solution; it is now a straightforward labeling protocol. (See [23] and [18] for a history.) P3P's target is restricted to the Web.

P3P allows services and individual users to come to agreements on the release of personal data. It also allows users to understand, to some extent, what the consequences of their disclosure will be. Essentially, P3P deals with how people manage their private information with regard to other people, companies, and institutions:

> The goal of P3P is to enable users to exercise preferences about Web sites' privacy practices. P3P applications will allow users to be informed about Web site practices, delegate decisions to their computer agent when they wish, and tailor relationships with specific sites ([24]).

P3P is both an architecture and a protocol. P3P's basic architecture, as defined in architecture specifications (e.g., [25], [22]), consists of a two-way relationship between a Web-based service and some user agent. (See Figure 1.)

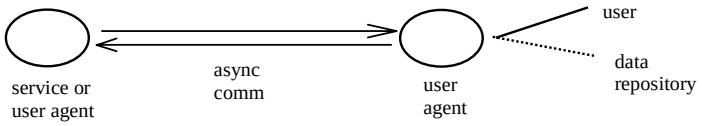

Figure 1: P3P Architecture

The user agent (or set of agents) acts on behalf of the user. The user agent may consist of a manually controlled interface, an intelligent agent (or set of agents), or a third-party. The data repository, which can be used to store personal data, is optional; if it exists, it might be embedded or moved to a third-party intermediary.

The P3P protocol functions within this architecture to send proposals (i.e., a service's data requests and privacy statements) and to match them with users' privacy preferences:

> P3P is designed to help users *reach informed agreements with services* (Web sites and applications that declare privacy practices and make data

> requests). As the first step towards reaching an agreement, *a service sends a machine-readable P3P proposal* …, in which the organization responsible for the service declares its identity and privacy practices….
>
> Proposals can be *automatically parsed* by user agents such as Web browsers and compared with privacy preferences set by the user. *If there is a match between service practices and user preferences, a P3P agreement is reached.* Users should be able to configure their agents to reach agreement with, and proceed seamlessly to, services that have certain types of practices; users should also be able to receive prompts or leave when encountering services that engage in potentially objectionable practices. Thus, users need not read the privacy policies at every Web site they visit to be assured that information exchanged (if any) is going to be appropriately used. ([24])

User agents are required to have an embedded trust engine in order to match preferences and proposals. With P3P, the "appropriate" release of data is to be defined by the user, as described by the user's privacy preferences. These preferences are then matched against the services' data collection and use policies.

The site's policy statements (see below) are matched against the user preferences. The user preferences themselves can be stated in any manner; they may be machine learnt or explicitly set by the user, for example. Another manner of setting the preferences might be to use third-party preferences and then tailor them. The APPEL interchange language was designed to facilitate this use.

Automatically transferring data occurs in the current version only when the data is clickstream or otherwise readily available. While the original intention of P3P was to automatically negotiate the release of personal data to services that followed appropriate privacy practices, it was found that users did not want the automatic transfer of most personal data [11].

P3P Proposals

P3P proposals are extensively described in Cranor [18] as well as in the W3C specifications (e.g., [22]). P3P proposals are XML-based statements that include "slots" for access to identifiable data, dispute resolution mechanisms, desired

data, consequence of data release, purpose of data collection, recipients of the data, and data retention policy. These slots have extremely limited options for values. For example, a P3P statement can have as its data retention disclosure only the values of no-retention, stated-purpose, legal-requirement, business-practices, and indefinitely. The site can have a human-readable retention policy as well, but the machine-readable portion is limited to these five values.

An example of a P3P proposal can be seen in Table 1. This proposal states that the TheCoolCatalog.com Web site collects data for personalization and their own development. They keep the data indefinitely and do not provide access. They also provide urls for obtaining information about dispute resolution, a link to their third-party oversight, and a English-text privacy policy statement.

```
<POLICY xmlns="http://www.w3.org/2000/P3Pv1" entity="TheCoolCatalog,
   123 Main Street, Bethesda, MD 20814, USA">
   <DISPUTES-GROUP><DISPUTES resolution-type="independent"
     service="http://www.PrivacySeal.org" description="PrivacySeal.org"
     image="http://www.PrivacySeal.org/Logo.gif"/></DISPUTES-GROUP>
   <DISCLOSURE discuri="http://www.TheCoolCatalog.com/PrivacyPractice.html"/>
   <STATEMENT>
     <RECIPIENT><ours/></RECIPIENT>
     <PURPOSE><custom/><develop/></PURPOSE>
     <RETENTION><indefinitely/></RETENTION>
     <DATA-GROUP>
       <DATA name="user.gender"/>
       <DATA name="user.home." optional="yes"/>
     </DATA-GROUP>
   </STATEMENT>
```

Table 1: A sample P3P policy statement from a Web site. Adapted from http://www.w3.org/Consortium/Offices/Presentations/P3P/20.html.

P3P Technologies

A number of auxiliary technologies have been constructed for P3P. Aside from policy generators and trust engines, prototype user interfaces (e.g., AT&T Privacy Bird [18]) and user agents have been constructed. One set of augmentative agents consists of the Privacy Critics system [26]. These agents note discrepancies in privacy policies, check privacy statements against third-party databases of problems, and provide other useful assistance to users. I believe that a similar mechanism may also be of use in pervasive environments.

P3P's use has also been extended. Langheinrich [27] proposes using P3P in "privacy beacons" to enable users to know when sensor environments will capture private data (such as location or physical presence).

In summary, as Hockheiser [23] noted, P3P is an attempt to construct

> …a standardized, machine-readable format for these [privacy] policies, along with a protocol for finding them. By making privacy policies easier to find and understand, P3P might help users find the information that they need to make informed decisions regarding sites they visit. ([23], p. 283)

P3P As Success

P3P has been adopted with some success. While there is a perception, especially in the policy community, that P3P was late for its window of opportunity, many major sites label themselves with P3P, and Web browsers (e.g., Internet Explorer 6.0+, Netscape 7) use P3P. On the other hand, adoption by smaller sites has been slower. (See [28] for adoption patterns and rates.)

P3P did take a significant amount of time to design and deploy. In retrospect, some of this was perhaps inevitable. P3P was the first social protocol [24]. As such, it had adoption issues that come with new technologies. For users to adopt P3P, sites had to label their practices; for sites to adopt, users had to use suitable user agents and trust engines. In addition, regulators had to feel comfortable with P3P (especially in Europe). It still remains for courts to consider whether P3P is enforceable. Nonetheless, if it were not for the expectation failure that P3P would "solve" the privacy problem on the Web, P3P would be seen entirely as an adoption success.

Even more importantly than adoption, the P3P effort shows what a labeling protocol can accomplish. With the Web, a labeling protocol and its resulting architecture are critical: Users need to be informed of the data requests and consequences. P3P is, at the least, a necessary first step. As argued above, for users to have privacy, they must be warned of potential dissemination of their personal data. Indeed, for adequate opt-in or opt-out procedures, a user must be notified of the exact request. This requires some protocol like P3P. In addition, for users to feel as though they are in control, they must have an adequate understanding of any consequences. For this, they must understand to whom they are releasing data, the purposes for which the data are requested, and the redress they might have. Again, this requires some protocol like P3P.

Only a labeling architecture and protocol can help in this situation, and P3P shows that such an architecture and protocol could be constructed. This is a critically

important lesson. For the rest of this paper, it is assumed that these protocols and architectures are possible, but that they will need to be refined and augmented for future pervasive environments.

Therefore, distilling the tradeoffs, limitations, and open issues from the P3P effort is essential to understanding what needs to be done for next generation labeling protocols. The critics of P3P have been quite vocal, and they have pointed out some important issues. Many other issues became obvious through the effort itself. These open issues indicate the changes and continued research needed for next-generation labeling protocols. These critiques, open issues, and lessons are examined next.

P3P Critiques, Open Issues, and Lessons

P3P has been critiqued from a number of directions, but here I will consider three types of critiques. The first involve the political-economic context of the P3P effort. Another set of critiques basically questions the usability of the system overall. These usability critiques present important lessons and open issues. In addition, there are a number of issues and lessons that arose from the design process itself. After briefly describing the political-economic critiques, this section will discuss the usability and design process issues and lessons.

Political-economic Critiques

The first set of critiques examines P3P in its political-economic context. In some critiques, P3P is perceived as a political "tool" of corporate interests that wanted to stall or foreclose privacy regulations on the Internet [29]. In this view, these corporate interests were unwilling or uninterested in cooperating fully, and P3P was only a tactic for delaying attempts to properly regulate the Internet.

Other political-economic critiques have examined P3P as potentially creating a marketplace or negotiated space of personal data. This line of critique follows from P3P's nature: P3P is a "social protocol", in Cranor and Reagle's [30] terms; it is an attempt to create a computer-based protocol to augment social interaction. As such, it is necessarily political, in that any attempt to regulate or facilitate social interaction has political and power consequences. P3P carries with it a

number of implications about privacy: how "privacy" should be approached, what the fundamental "components" of "privacy" should be, and what kinds of attention should be paid to these components.

Many advocates (e.g., [29]]) and researchers feel that privacy is a fundamental right instead of a commercial interest. P3P appears to stand in harsh contrast; these advocates see P3P as promoting a potential marketplace for personal data, eroding other viewpoints. Furthermore, certain aspects of privacy are thought to be important by these advocates to furthering privacy as a right. Some, especially access and subsequent use, are thought to be absent from P3P. (One might argue whether including such aspects was realistic in a conservative-dominated US.)

These criticisms essentially object to P3P on ideological grounds. These grounds may or may not be valid with regard to privacy on the Web. But if they are valid, then it is likely that labeling protocols will also further a potential marketplace for personal data in pervasive environments.

Labeling protocols, like many technologies, can be used in a variety of political scenarios. Here, we have assumed the utility of informing users and letting them decide how to act. In the current US political environment, substantial regulation is not feasible, and mechanisms using labeling protocols may be the best that can be achieved. For the remainder of the paper, issues concerning the political-economic context and implications will be bracketed off from further consideration.

Usability Critiques and Lessons

Another set of critiques focuses on the usability of P3P. These critiques are more central to the technology itself and therefore must be examined in depth. The usability concerns come from a number of directions.

One usability criticism is that P3P's vocabulary and structure may be too complex for naïve users to incorporate and use [23]. Hockheiser, for example, felt that P3P's vocabulary might appear to users as though the terms are common and understandable. However, the terms, as used by services, have meanings that are restricted or not obvious. In his critique, the vocabulary is structured around the processes and needs of services, rather than the empirically-derived preferences of

users. (There was, in actuality, some testing of the terms and preferences in [31].) Hockheiser does point out that P3P suffers from an inherent tradeoff between the simplicity needed by users and the complexity required for a concrete and grounded understanding of preference and policy terms. This is a tension that must be managed with P3P, labeling protocols in general, and many user-centric technologies, since it involves the conflict between a vocabulary that is succinct and controllable versus a vocabulary that is completely explanatory. Hockheiser also notes usability issues for the user agents (or user clients, in his terms).

Hockheiser derives other usability arguments from [31, 32]. In this analysis, privacy is a wicked problem (in the computer science sense of "wicked", meaning an ill-formed, intractable problem). Overall, the vocabulary and user agent issues are but symptoms of the general problem of user control with privacy technologies.

This intractability comes from fundamental social requirements. First, if we follow Goffman [1961], users want to control their "face" with respect to different groups, persons, and institutions. People make decisions about what information to release to others everyday. Within a technology, however, this requires controlling their information transfer in two dimensions – in one dimension by each recipient of the personal data (i.e., all potential recipients, perhaps including any interaction effects among recipients) and in the other dimension, by each datum (i.e., all possible pieces of private information, as defined by the user). This is essentially an infinite two-dimensional space.

As well, except in unusual circumstances, people do not have to stop to deliberate about what to say or withhold. Nor do we need to interrupt conversations to laboriously switch modes. To require users, in the middle of a social interaction, to change modes or set properties to control their personal data is likely to disrupt the user experience.

The P3P protocol, accordingly, presents a difficult HCI challenge. The P3P protocol currently allows for the expression of up to ten dimensions, not just two. With some important exceptions, these ten dimensions within P3P incorporate most details of everyday data exchange. But, one can easily assert that no one knows how to construct a suitable user interface for such a ten-dimensional protocol. Without a completely accurate grouping mechanism (or some manner

of collapsing categories in a meaningful way), few users will be able to correctly categorize a situation without errors. Fewer yet may take the time to categorize, since normal social activity does not require this explicit categorization. Machine learning technologies have yet to be robust and precise enough for use with privacy, since if the learning is wrong even once, the user will likely walk away from the system. Moreover, one can also assert that no one knows how to construct a user interface that is suitably flexible but does not require the user to interrupt social interaction.

Labeling protocols are not unique in this. Elsewhere I have argued that there are no current technical mechanisms to straightforwardly mechanize the naturally occurring, everyday social activity of handling personal information in its entirety. This social-technical gap ([31, 32]) leads to information systems that are often brittle. They are not sufficiently capable of handling human nuance, flexibility, and ambiguity. Given the difficulty with information handling in general, we can expect a long road to finding adequate HCI mechanisms for handling privacy.

To summarize, several usability issues, then, remain for further work. More work will be required on suitable mechanisms for user control. In P3P, because of the nature of Web privacy (as explained in the previous section), satisfactory user control was often difficult if not impossible to achieve. In a Web environment, users are in a position where they must handle an essentially infinite information space, which places an impossible burden on user interface mechanisms. For future labeling protocols, the user control mechanisms must be simplified to the point where their complexity can be handled by most, if not all users, but their utility is still valuable [31]. The great concern, of course, is that they will be simplified to the point where their utility is minimal. Further research work will be required to find suitable user control mechanisms. The success for next generation user interfaces and next generation protocols will likely be found in constraining this problem appropriately.

In addition, P3P requests, if not handled proactively and adequately by a user agent, would be disruptive to normal social interaction. P3P provides the raw materials for users' decisions about their personal data, but either they need to provide attention to the P3P statements or suitable user mechanisms will be required. As will be argued below, the sheer volume of requests in a pervasive environment is likely to be

overwhelming to most users. This is not merely a user interface problem; the problem is conditioned as well by the underlying social requirement that there be apparently seamless changes between social states and roles. While perhaps more obvious in hindsight, P3P efforts wrestled with this problem, but did not offer a solution. New labeling protocols will need to continue to struggle with this problem.

Other Lessons and Open Issues

Several non-usability lessons for future protocols can also be garnered from the P3P experience. These involve the design process for P3P as a social protocol.

Firstly, the P3P design inhabits the intersection of law and technology. P3P contains mechanisms for understanding who is responsible for the data and what would happen to the data. These mechanisms are critical for users to believe in the efficacy of the system and its help in maintaining their privacy. In the P3P design process, a significant tension existed between the precision required for computational support and the ambiguity required for commercial and legal approval. More importantly, the Web environment adds enormous complexity to describing legal entities (often across geographical and regulatory boundaries), consequences (resulting in very broad categories), and redress. A very large effort in the P3P design went into resolving the issues of responsibility and in dissemination, but some parts of P3P still ended up with overly broad categorizations. In future protocols, one would prefer to reduce the complexity, when unnecessary in new environments such as pervasive situations, and to increase the granularity of categorization, when helpful in more constrained situations.

In addition, the P3P effort showed that future labeling protocols should start as research projects, with their longer time frames and their ability to prototype. P3P suffered from a too-early level of public scrutiny. It began as an industry-led W3C effort as an outgrowth of an earlier W3C project (PICS). Early participants included privacy advocates, W3C staff, and some corporate interests (e.g., direct marketers). It grew quickly, as the W3C reached out, to include W3C corporate members (e.g., computer companies and banks) and interested parties (e.g., academics). The P3P effort was by its nature semi-public. The specifications are written within working groups, and these working groups are limited to W3C

members and invited participants. When a specification is ready, however, there is a period of public comment before the specification is finalized by a vote of W3C members. In general, W3C follows the process of setting standards. In addition to this normal process, considerable effort was made to include privacy advocates (of all degrees) and government entities (e.g., privacy commissioners). This all was quite beneficial.

However, the perceived need for an immediately available technology by the largely commercial members of the W3C led to severe pressures on the specification process. The technology was over-promised and over-hyped, given the available time frame and the early stage of understanding what such a protocol might be and what it might have to include. There was no gestation time for the specifications, nor time to run empirically-based evaluations of alternative designs.

New social protocols in general and new labeling protocols specifically will need to have additional time. One way to do this is to create such protocols as research projects; the research community can then vet alternative designs and underlying mechanisms. The resulting technology may be more florid and abstruse than strictly necessary, but just as the Web resulted from the efforts of the hypertext community, a new widely-deployed privacy protocol could result from such a research effort. In addition, a number of research alternatives could serve as a resource for policy-makers in their relatively short window of interest.

To summarize all of these lessons, P3P allows users and services to agree upon the conditions of data release and dissemination. This was a success, which shows that labeling protocols can be effective and useful. However, P3P's current state needs to be augmented (and some would say fixed) with resolution of issues in usability – particularly user control and seamlessness – as well as statements about responsibility and dissemination. Furthermore, additional time needs to be allowed for a maturation of any future protocols.

The next section argues that privacy will get only worse in pervasive environments, that a labeling protocol will be required, and that the lessons from P3P carry to these next generation environments.

Privacy in Pervasive Environments

As important as privacy has been for Web activity, it will be even more important for the acceptance of pervasive environments. Ubiquitous computing will exacerbate a problem already present and critical in users' minds.

To date, no technical mechanisms exist within pervasive architectures to notify individuals about what is being requested (or taken) and the effect of that provision. Some labeling protocol will be required. A next-generation labeling protocol, as did P3P, will at least reduce or eliminate the non-symmetric barter for private information – the individual can at least determine what is desired and what will be done with the data. No such protocol currently exists for pervasive environments, although there is substantial research interest. A next generation labeling protocol, designed for pervasive environments, is necessary.

However, new stresses will exist for this next generation labeling protocol:

- In a context-aware environment with a suite of context-aware applications, an individual will operate within many social and organizational contexts, and surrounding social and organizational environments may make use of many individuals' data. A pervasive software environment, consisting of many systems, may be in a complex relation to "the user." The P3P effort showed how difficult it was to concretely demark the responsible entities and consequences. This will need to be tackled head-on for large-scale ubiquitous environments. (However, it may be possible to restrict the problem initially in intra-organizational settings.)

- If only because of the volume of data requests, most users will have an increasingly difficult time controlling their data, given the increased volume, invisibility of the technology, and uncertain regulatory environment. The time needed to investigate the impact of various systems on their privacy rights and then take steps necessary (even where available) to protect their privacy is likely to be daunting. The P3P effort showed that new efforts will be required to find basic user interface mechanisms for notifying users.

- Individuals will need help in determining what to do. Relatively few people understand the full implications of data transfer; this is especially true as the ubiquitous computing world begins. A new civility, with shared assumptions about social obligations and relationships, will take some time to come together. Within the

current regulatory and political environment, it is not clear that any privacy promise made with regard to particular data will be respected. This cannot be satisfied merely by cryptography; that is, cryptographic security is a necessary but not sufficient condition for privacy. At this point, the cost to an individual for determining and correcting a privacy mishap (intentional or accidental) is too high. People will need new tools to weigh the truthfulness and the consequences of privacy situations and of privacy statements. In addition, the P3P experience says that some sort of enforcement – organizational, regulatory, or legal – will be key.

If, as argued above, labeling protocols are required for pervasive environments, and as argued above, there are new stresses on even the P3P mechanisms, additional time and efforts are required to find proper solutions. This again argues for working out any new labeling protocols while pervasive environments remain research prototypes.

Conclusions

P3P is the first "social protocol" to address privacy technically. Best seen as a labeling protocol for describing the consequences of disclosing personal information, P3P's goal is to help users control their privacy on the Web according to their own preferences. Some similar mechanism will be required for ubiquitous computing. At times, identification and authentication is required, and some mechanism for signaling the consequences must be created for pervasive environments. The P3P project raised significant technical and socio-technical issues that must be addressed again in the next generation labeling protocol. However, with the additional time made possible by creating and evaluating numerous research alternatives and prototypes, it is entirely possible that we can help users control their privacy in the upcoming pervasive environments.

Acknowledgements

Many people have helped in my understanding both of privacy and pervasive environments. I wish to thank (in no particular order) Lorrie Cranor, Joseph Reagle, Ralph Swick, Danny Weitzner, Maya Bernstein, Ari Schwartz, Atul Prakash, and the many participants in the P3P project, the W3C, and MIT/Project Oxygen. I would also like to thank the three anonymous reviewers of this paper and Chris Schmandt for their help with this paper. My deep appreciation goes to

the late Michael Dertouzos for his support of my examining privacy in Project Oxygen.

This work was sponsored in part by the MIT Project Oxygen partners and the Intel Corporation.

References

1. Culnan, M. J. Protecting Privacy Online: Is Self-Regulation Working? 2000, 19(1): 20-26.

2. Clarke, R. Introduction to Dataveillance and Information Privacy, and Definition of Terms. http://www.anu.edu.au/people/Roger.Clarke/DV/Intro.html,

3. Fisher, S. Privacy By Design. 2001, 23(27): 20-22.

4. Harris Poll. *Online Privacy: A Growing Threat.* Business Week. 96, 2000.

5. Culnan, M. J. and P. K. Armstrong. Information Privacy Concerns, Procedural Fairness and Impersonal Trust: An Empirical Investigation. 1999, 10(1): 104-115.

6. Davies, S. G. "Re-Engineering the Right to Privacy: How Privacy Has Been Transformed from a Right to a Commodity." in Agre and Rotenberg. *Technology and Privacy: The New Landscape*. MIT Press, Cambridge, MA, 1997.

7. Etzioni, A. *The Limits of Privacy*. Basic Books, New York, 1999.

8. Goffman, E. *The Presentation of Self in Everyday Life*. Anchor-Doubleday, New York, 1961.

9. Smith, H. J., S. J. Milberg and S. J. Burke. Information Privacy: Measuring Individuals' Concerns about Organizational Practices. 1996, June: 167-196.

10. Westin, A. F. *Harris-Equifax Consumer Privacy Survey 1991*. Equifax, Inc., Atlanta, 1991.

11. Ackerman, M. S., L. Cranor and J. Reagle. Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences. *Proceedings of ACM Conference in Electronic Commerce*, 1999: 1-8.

12. Spiekermann, S., J. Grossklags and B. Berendt. E-privacy in 2nd Generation E-Commerce: Privacy Preferences versus Actual Behavior. *Proceedings of ACM Conference on Electronic Commerce*, 2001: 38-46.

13. Dhillon, G. S. and T. T. Moores. Internet Privacy: Interpreting Key Issues. *Information Resources Management Journal*, 2001, 14(4): 33-37.

14. Neuman, W. R., L. W. McKnight and R. J. Solomon. *The Gordian Knot: Political Gridlock on the Information Highway*. MIT Press, Cambridge, MA, 1997.

15. Chaum, D. Security without identification: transaction systems to make big brother obsolete. *Communications of the ACM*, 1985, 28(10): 1030 - 1044.

16. Reiter, M. K. and A. D. Rubin. Anonymous Web transactions with Crowds. *Communications of the ACM*, 1999, 42(2): 32 - 48.

17. Goldschlag, D. M., M. G.Reed and P. F.Syverson. Onion Routing for Anonymous and Private Internet Connection. *Communication of the ACM*, 1999, 42(2): 39-41.

18. Cranor, L. F. *Web Privacy with P3P*. O'Reilly, Cambridge, MA, 2003.

19. Clarke, R. Of Trustworthiness and Pets: What Lawyers Haven't Done for e-Business. http://www.anu.edu.au/people/Roger.Clarke/EC/PacRimCL01.html,

20. Priyantha, N. B., A. Chakraborty and H. Balakrishnan. The Cricket location-support system. *Proceedings of 6th annual international conference on mjobile computing and networking*, 2000: 32 - 43.

21. Hong, J. An Architecture for Privacy-Sensitive Ubiquitous Computing. *Proceedings of MobiSys 2004*, 2004: forthcoming.

22. Cranor, L., M. Langheinrich, M. Marchiori, M. Presler-Marshall and J. Reagle. The Platform for Privacy Preferences 1.0 (P3P1.0) Specification. MIT/World Wide Web Consortium, http://www.w3.org/TR/P3P/,

23. Hockheiser, H. The Platform for Privacy Preference as a Social Protocol: An Examination Within the U.S. Policy Context. *ACM Transactions on Internet Technology*, 2002, 2(4): 276-306.

24. Cranor, L. and J. Reagle. The Platform for Privacy Preferences. *Communications of the ACM*, 1998, 42(2): 48-55.

25. Ackerman, M. S. *General Overview of the P3P Architecture*. 1997.

26. Ackerman, M. S. and L. Cranor. Privacy Critics: UI Components to Safeguard Users' Privacy. *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI'99)*, 1999: 258-259.

27. Langheinrich, M. A Privacy Awareness System for Ubiquitous Computing Environments. *Proceedings of Ubicomp 2002*, 2002: www.inf.ethz.ch/~langhein.

28. Byers, S., L. F. Cranor and D. Kormann. Automated Analysis of P3P-Enabled Web Sites. *Proceedings of Fifth International Conference on Electronic Commerce (ICEC 2003)*, 2003: 326-338.

29. Catlett, J. Open letter to P3P developers & replies. *Proceedings of Tenth conference on Computers, freedom and privacy: challenging the assumptions*, 2000: 157-164.

30. Cranor, L. F. and J. Reagle Jr. "Designing a Social Protocol: Lessons Learned from the Platform for Privacy Preferences." in MacKie-Mason and Waterman. *Telephony, the Internet, and the Media*. Lawrence Erlbaum Associates, Mahwah, NJ, 1998.

31. Ackerman, M. S. The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility. *Human-Computer Interaction*, 2000, 15(2-3): 179-204.

32. Ackerman, M. S. "The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility." in Carroll. *HCI in the New Millennium*. Addison-Wesley, New York, 2001.