Privacy In Context

Mark Ackerman[1], Trevor Darrell[2], and Daniel J. Weitzner[3]

[1] *MIT/Laboratory for Computer Science*

[2] *MIT/Department of Electrical Engineering and Computer Science and MIT/Artificial Intelligence Laboratory*

[3] *MIT/World Wide Web Consortium*

Corresponding Author's Contact Information: Mark Ackerman, MIT/Lab for Computer Science, Room NE43-610, 200 Technology Square, Cambridge, MA, 02139, ackerman@lcs.mit.edu (or ackerman@ics.uci.edu).

Brief Authors' Biographies: Mark Ackerman is a research scientist at MIT's Project Oxygen and works on collaborative systems. He is also an associate professor in Information and Computer Science at UC Irvine (on leave). Trevor Darrell is a computer vision researcher with an interest in perceptual user interfaces; he is a menber of the Artificial Intelligence Lab and an Assistant Professor in the department of Electrical Engineering and Computer Science at MIT. Daniel Weitzner is director of the World Wide Web Consortium's Technology and Society activities.

ABSTRACT

Context-aware computing offers the promise of significant user gains - the ability for systems to adapt more readily to user needs, models, and goals. Dey et al presents a masterful step towards understanding context-aware applications. We examine Dey et al in the light of privacy issues -- that is, individuals' control over their personal data -- to highlight some of the thorny issues in context-aware computing that will be upon us soon. We argue that privacy in context-aware computing, especially those with perceptually-aware environments, will be quite complex. Indeed privacy forms a *co-design* space between the social, the technical, and the regulatory. We recognize that Dey et al is a necessary first step in examining important software engineering concerns, but future research will need to consider how regulatory and technical solutions might be co-designed to form a public good.

CONTENTS

1. **INTRODUCTION**
2. **USERS AND PRIVACY IN CONTEXT-AWARE APPLICATIONS**
3. **POLICY-DRIVEN DESIGN REQUIREMENTS**
4. **OPEN RESEARCH ISSUES**

5. INTRODUCTION

Context-aware computing offers the promise of significant user gains - the ability for systems to adapt more readily to user needs, models, and goals. Computational systems have clearly lacked the vast amount of state data required to handle even relatively simple customizations, leaving systems without the ability to handle an individual's level of desired social nuance and control (Ackerman 2000). Collaborative systems, for example, with their need for fine-grained control over personal data could profit enormously from additional context.

Dey, Salver, and Abowd (2001 [this special issue]) presents a significant step towards understanding context-aware applications. Here, we examine Dey et al in the light of privacy issues - that is, individuals' control over their personal data. We recognize that Dey et al is a necessary first step in examining important software engineering concerns, and first steps cannot solve every research question: Here, we wish to use Dey et al's viewpoint to highlight some of the thorny issues in context-aware computing that will be upon us soon.

According,we briefly summarizes Dey et al's privacy stance, one made both explicitly and implicitly in the paper, and then deepens it to fit other pervasive, context-aware environments. This is followed by a discussion of the technical issues in privacy control in context-constructing applications, especially those derived from perceptual data. Next, the regulatory and legal oversights that exist and may need to exist are briefly discussed. The piece ends with a discussion of co-evolutionary design, where technical possibilities may result from the co-occurrence of regulatory assistance. Throughout, we take examples from not only Dey et al, but also from Project Oxygen at MIT as well as other large-scale, pervasive environments.

6. USERS AND PRIVACY IN CONTEXT-AWARE APPLICATIONS

Privacy is intrinsically bound up with control - who controls what information as well as the applications and systems that construct and disseminate that information. The control of the various sensors and services is not well specified in Dey et al., although often applications are implied to be under the control of a "user". In this Dey et al basically views software as a resource with a specific owner, usually an individual.

However, the control over systems determining the presence of individuals in a room or other location, on the other hand, is not specified at all. This is a critical point. In general, in a context-aware environment with a suite of context-aware applications, an individual operates within many social environments, and social environments may make use of many individuals' data. This view argues that the software environment, consisting of many systems, may be in a complex relation to the person Dey et al considers the user. In other words, Dey et al does not clearly distinguish among the various types of contexts within which a user resides or the number of contexts. These

are significant concerns when dealing with one's private data. Consider the following, all of which will have different privacy characteristics for users:[1]

1. An individual wishes to connect over a mobile, wireless network. This requires only partial identification data; that is, a user merely needs to provide the relevant certificates that he or she can connect. The only context data that can be derived are how many people are in a specific location; no individuals can be identified. In this case, data are derived by the system's owner, but there is no additional control over some user's private data. It is also possible to prevent pseudo-identifiability, where the user cannot be mapped to an individual but can be identified as the same user of a system, through the proper use of certificates. In other words, the user need not trust the owner of the system. Privacy is maintained by allowing the user to disseminate only the necessary data, which cannot be used to identify the user.

2. Two individuals walk into a conference room, and the room wishes to adjust the temperature appropriately. The system that controls the room may wish to determine whether these two users have previously used the room and set the temperature. Other systems, such as file caching systems, also wish to know who is entering the room. A room agent asks the two users for identifying information.

   This scenario differs from the above scenario in two ways. First, the users have been asked to provide identifying information. In this case, the temperature application may be satisfied with the user providing a data model that can appropriately state the user's preferences but cannot be used as identifiable data. However, the temperature application would have to be written to handle this substitution. The Dey et al. paper provides a usable architecture for doing so. On the other hand, the file caching application clearly needs identifying information. Second, the downstream consequences for providing private data are unknown or unspecified. A user may trust the file caching application because of the norms of operating system usage and system administration (although this varies somewhat by organization). There is no way for a user to know further uses of identifying information provided to the temperature application. Furthermore, other applications, unknown and unseen to a given user, could be operating in the room, and any identifying or private data provided to these applications may, again, have unknown consequences and further uses.

   In this scenario, some privacy relief can be given to the user by allowing the user to provide partial, non-identifying data. The major privacy effects come from there being no mechanism to tell the user what context-gathering systems

[1] We recognize that individuals vary widely in their concern over privacy and information control. Many studies (including Ackerman et al 1999) have found three general clusters in the American population - privacy fundamentalists, privacy pragmatists, and privacy unconcerneds. We simplify here for the argument's clarity, but any full analysis would need to consider each cluster separately.

are present and what they intend to do with any collected data. However, there are several mitigating circumstances. The user is aware that data have been requested, because it must be requested explicitly. (It is not clear, however, that if users are bombarded with hundreds of requests by dozens of systems that they will actually provide individual attention to each request.) Furthermore, a user still has control over the dissemination of that data at the point of request (but only at that point). Finally, the request and its consequences occur within some organizational or institutional setting, and this allows some social regulation in the form of rules or norms.

3. We can consider the same type of system – two individuals walking into a meeting room with a temperature application, file caching application, and unknown systems – but in this case, the identification occurs automatically using peripheral interfaces. We will consider the technical possibilities in more detail below. However, this scenario differs from the above in that the user is never aware that any private data have been requested or provided. The consequences of providing the data are again unknown. This is the most problematic case, since there does not appear to be any privacy remedy or relief technically. However, the identification – in this case – occurs within some organizational or institutional setting, and this allows some social regulation again.

4. The final scenario is having a mobile user walking down the street. An image-based perceptual system recognizes the mobile user, matching the user, for example, from a picture taken when cashing a check. Again, the user is not aware that any private data have been requested and does not know the downstream consequences of being identified. We might assume in this case that a context is being constructed, but it is not for the immediate benefit of the user. Benefits accrue not to the user, but to the data gathering organization as well as downstream consolidators. Indeed, control over the derived data, and the systems construct and inferring context, lie completely outside the user and perhaps social convention.

While it might be argued that Dey et al's framework works when there is an explicit device for each user, as well as the physical possession (e.g., of an active badge or PDA) to determine ownership, many authors have envisioned perceptual interfaces which track users' position in a room, can recognize them when they return, and can detect pointing gestures and certain facial expressions. The Microsoft Easyliving project uses a network of stereo cameras to track users as they move in a room, and uses this information to supply geometric context to an application (Brumitt et al 2000). Several researchers have built perceptually enabled kiosks, which react directly to a user's speech or gesture (Rehg 1997; Darrell et al 2000). Face recognition and expression tracking are an active area of research in desktop perceptual user interfaces (Pentland 2000), and several commercial products for face identification have recently been introduced (e.g., www.visionics.com). Systems that combine CCTV surveillance and face recognition have been sold, but so far only for outdoor law enforcement applications (Privacy Digest 2000).

> Left unchecked, there can be major problems with privacy and passively sensed audio-visual context. Export of image or acoustic information by independent agents in an environment, when offered in an uncontrolled marketplace of digital information, can lead to a scenario where any third party can compile a complete record of a person's daily activity, a scenario termed by Agre as the "privacy Chernobyl" (Agre 1999, Agre and Rotenberg 1997).

These scenarios should be a sharp reminder that social settings into which context-aware applications will be introduced as well as the social arrangements in which context-aware applications will be employed are likely to be varied and complex. Dey et al's scenarios are explicitly simplified for explanatory and rhetorical purposes, but this simplification carries with it a utopian reduction that glosses over two important social considerations. First, one person's contextual awareness is another person's lack of privacy (Hudson and Smith 1996). One person's sensor is another person's spy. Second, control over the acquisition and dissemination of contextual data is not likely to be straightforward. Users will not want to actively participate in controlling their private data (because of the sheer volume, especially at the sensor data level), and in any case, may not be allowed active participation and choice in many cases.

When perceptual interfaces and perceptually-derived context become commonly used in pervasive computing environments, as we believe they will, it will require a combination of social, legal, and technical consideration to establish a reasonable degree of privacy. What might that reasonable degree be?

2. POLICY-DRIVEN DESIGN REQUIREMENTS

The history of privacy law all around the world illustrates the fundamental importance of individual control over one's own personal data. Since the early 1970s, when democratic governments around the world began developing privacy rules to address the perceived privacy threats posed by large mainframe databases, the notion of user control has features prominently in all regulatory schemes (HEW 1973, OECD 1980, European Union 1995, FTC 2000). The most recent US privacy guidelines have been issued by the Federal Trade Commission (FTC 2000) cover four basic areas:

- **Notice**: The individual should have clear notice of the type of information collected, its use, and an indication of third parties other than the original collector who will have access to the data.

- **Choice:** The ability to choose not to have data collected.

- **Access:** The ability for the data subject to see what personal information is held about him/her, to correct errors, and to delete the information if desired.

- **Security:** Reasonable measure taken to secure (both technically and operational) the data from unauthorized access.

These guidelines do not have the force of law in the US. Privacy regulation in Europe and other countries with comprehensive 'data protection' laws, also add other areas of regulatory interest to these principles. Nevertheless, these Fair Information Practices do suggest a baseline for privacy rules that all system designs ought to consider.

Context-aware systems certainly pose a new set of privacy challenges not previous considered by existing regulatory frameworks. As we have noted, the passive collection and logging of information related to users' activities, information access patterns, movement, etc., all raise serious privacy concerns. Notably, recent changes in US law have come to recognize some of the leading edge privacy threats posed by new information environments. Amendments in US wiretapping law in 1994 granted special protections for 'transactional records' of user network usage data. This change in law was prompted by a desire to protect the privacy of email and Web server access logs. Even in 1994, the US Congress saw that records of the Web pages visited and the email messages sent and received by users could reveal highly sensitive, personal information. The protections enacted in 1994 gave a higher level of protection against law enforcement access, through police wiretap procedures, though not to commercial or private sector use of these data.

Nevertheless, how to extend privacy initiatives and technical prototypes into a contextually aware, perceptual interface world is not clear. We have noted the new technical possibilities in the various scenarios above. We should view the US Congress' increased concern over these records as a clear indication of the public policy imperative to offer protection against unauthorized access to log and sensor data. Below are the open research issues in providing users with control over the personal information about them.

3. OPEN RESEARCH ISSUES

At MIT's Project Oxygen, we are developing a range of technologies that may gather context information about a user via perceptual means, and we are simultaneously investigating how to give the user control over how his or her image or expression are used.

First, it is likely that there will be considerable variation in the rules people wish to have govern perceptual context and privacy. Different social environments will lead to different levels of expectation of audiovisual privacy, and users will have varying levels of desired control. Notice requires a framework for expressing privacy policies between a user and the environment and among users in an environment. This follows from the pioneering work of the Platform for Privacy Preferences (P3P) (http://www.w3.org/TR/P3P) by the World Wide Web Consortium. P3P enables Web sites to translate their privacy practices into a standardized, machine-readable format that can be retrieved automatically and easily interpreted by a user's browser. P3P clients then can compare the privacy statement with the user's preferences. Thus, P3P is the type of labeling protocol required, but its use to date has been only for the exchange of private data in

Web usage. It will need to be extended for contextually aware and perceptual environments.

Second, there are a number of HCI and user interface problems to be resolved. The requirement for notice, especially when accompanied by a need for consent, can be overwhelming to users. It also requires that users disrupt their activities (Ackerman 2000). In a contextually-based environment, such notice can not only occur an order of magnitude more than cookie notices, such a disruption also defeats the basic goal to make contextually-aware environments unobtrusive. Early implementers of user privacy tools have noted that P3P-compliant user interfaces have the challenge of presenting often complex information about a privacy policy in a manner that is easily understandable by the user, who may care about her privacy but may have a low tolerance for being diverted from the primary task at hand to consider the privacy context of a given transaction. It is not clear how many users will invest 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. New efforts will be required to find basic user interface mechanisms for relatively unobtrusive notice. A greater understanding of the general HCI requirements for privacy (e.g., how individual differences affect use) will also be required.

A number of technical possibilities exist for ameliorating privacy concerns in contextually-aware environments that have only been touched on in Dey et al. These may offer some help with perceptual interfaces. While we have illustrated the privacy dilemmas posed by context-aware systems, we believe that the power of context-aware tools can also be used to help ease the privacy threats that these systems pose. One possibility is that context-aware computing could aid the negotiation over privacy agreements. As one example, a notion of location-dependent-state with collaborative filtering could offset the user interface difficulty with raw P3P. Context-aware user privacy widgets could help with the user interface challenge by considering what privacy choice the user made the last time she entered this particular context, what privacy choices have the user's trusted colleagues made in the current context, and what privacy choices did the user make when dealing with the same data collector though in an otherwise different context. Another possibility is that assemblages of privacy critics, using location information, could warn users when critical personal data are about to be revealed (Ackerman and Cranor 1999). Nonetheless, we expect the HCI challenge to remain large, and we hope that privacy will become a critical HCI research thrust.

Concomitant with the above technical possibilities for notice and consent is the requirement for security. For example, audio/visual expressions from users should be encrypted before being exported or stored, and the decryption key should be made available only to that user. A physically-grounded key could be distributed such that users must be physically present in the room to access it. As well, a user-specific *session* key could be obtained from the perceptual context widget by a user while physically in the room; allowing any data, which may be later distributed, to be watermarked with this key and thereby creating traceability.

Finally, in addition to any technical solutions and ameliorations, the architecture must contain features that verify-- if not guarantee -- that the privacy promise made with regard to particular data have been and will be respected. This is not satisfied merely by details of 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. We wish to shift the transactional cost for privacy from each user to the public. This can be done only through some regulatory or legal infrastructure that is erected to bolster and facilitate technical solutions. We will need to understand, as a research problem, how regulatory and technical solutions might be *co-designed* to form a public good.

The Oxygen privacy research goal, then, is to head off the privacy Chernobyl envisioned by Agre through a combination of technical and regulatory co-design initiatives. In Oxygen, we have started the development of perceptual context prototypes with the observed user's privacy uppermost in mind. In our work, we are establishing design principles for indoor environments as well as basic infrastructures that can be used to implement those principles. This will not require a radical departure from architectures like Dey et al., but we believe keeping privacy uppermost will greatly facilitate the use and adoption of those architectures.

NOTES

Support. This work has been supported by the Oxygen Alliance partners at MIT.

Authors' Addresses: Mark Ackerman, MIT/Lab for Computer Science, 200 Technology Square, Cambridge, MA, 02139. Email: ackerman@lcs.mit.edu. Trevor Darrell, MIT/Artificial Intelligence Laboratory, 200 Technology Square, Cambridge, MA, 02139. Email: trevor@ai.mit.edu. Daniel Weitzner, MIT/World Wide Web Consortium, 200 Technology Square, Cambridge, MA, 02139. Email: djweitzner@w3.org.

REFERENCES

Ackerman, M. S. (2000). The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility. *Human Computer Interaction*, 15, pp. 179-203.

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

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

Agre, P. (1999) Red Rock Eater Mailing List; notes and recommendation 14, http://commons.somewhere.com/rre/1999/RRE.notes.and.recommenda14.html

Agre, P, and Rotenberg, M. (eds). (1997) Technology and privacy: the new landscape. Cambridge: MIT Press.

Brumitt, B., Meyers, B., Krumm, J., Kern, A.,and Shafer, S. (2000) EasyLiving: Technologies for Intelligent Environments. *Proceedings of Handheld and Ubiquitous Computing*, September 2000. http://www.research.microsoft.com/easyliving/

Darrell, T., Gordon, G., Harville, M., and Woodfill, J., (2000) Integrated person tracking using stereo, color, and pattern detection. *International Journal of Computer Vision* , 37(2), pp. 175-185.

Dey, A. K., Salber, D., Abowd, G. D. (2001). A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-Computer Interaction*, 16, xxx-xxx.

European Union. (1995) Directive on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

FTC (US Federal Trade Commission). 2000 Fair Information Practice Guidelines. *Privacy Online: Fair Information Practices in the Electronic Marketplace* (http://www.ftc.gov/reports/privacy2000/privacy2000.pdf).

HEW (United States Department of Health, Education and Welfare). (1973) Code of Fair Information Practices.

Hudson, S. E., and Smith, I. (1996). Techniques for Addressing Fundamental Privacy and Disruption Tradeoffs in Awareness Support Systems. *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW'96)*: 248-257.

OECD (Council of the Organization for Economic Cooperation and Development). (1980) Guidelines for the Protection of Privacy and Transborder Flows of Personal Data

Pentland, A. (2000) Perceptual User Interfaces: Perceptual Intelligence. *Communications of the ACM*, 43 (3), pp 35-44.

Privacy Digest. (2000) The "Mandrake" project, as described in the PRIVACY Forum Digest, 07.18 (http://www.vortex.com/privacy/priv.07.18) and 07.19 (http://www.vortex.com/privacy/priv.07.19).

Rehg, J.M., Loughlin, M., Waters, K. (1997) Vision for a Smart Kiosk. *Proceedings of the 1997 IEEE Conference on Computer Vision and Pattern Recognition* (CVPR '97), pp. 690-696.