Approximate Information Flows: Socially-based Modeling of Privacy in Ubiquitous Computing

Xiaodong Jiang, Jason I. Hong, James A. Landay

Group for User Interface Research
Computer Science Division
University of California, Berkeley
Berkeley, CA 94720-1776, USA
{xdjiang, jasonh, landay@cs.berkeley.edu}

Abstract. In this paper, we propose a framework for supporting socially-compatible privacy objectives in ubiquitous computing settings. Drawing on social science research, we have developed a key objective called the *Principle of Minimum Asymmetry*, which seeks to minimize the imbalance between the people about whom data is being collected, and the systems and people that collect and use that data. We have also developed *Approximate Information Flow* (AIF), a model describing the interaction between the various actors and personal data. AIF effectively supports varying degrees of asymmetry for ubicomp systems, suggests new privacy protection mechanisms, and provides a foundation for inspecting privacy-friendliness of ubicomp systems.

1 Introduction

Privacy is not an absolute notion. It is, rather, a highly fluid concept about controlling the dissemination and use of one's personal information, one that often involves tradeoffs with efficiency, convenience, safety, accountability, business, marketing, and usability. Although a precise definition of privacy seems elusive, a very revealing characterization was given by Columbia economist Eli Noam [22].

"Privacy is an interaction, in which the information rights of different parties collide. The issue is of control over information flow by parties that have different preferences over 'information permeability'."

New technologies have always led to new threats to privacy. Some recent examples include intrusive telemarketing, logging of employee web surfing, and remote monitoring of public areas with video cameras, just to name a few (see [11] for many more examples).

Although many people believe that ubiquitous computing (ubicomp) holds great promise, there are also many critics that believe that such technologies will exacerbate these and other privacy-related issues for four reasons. First, wide-scale deployment of tiny sensors, coupled with improvements in recognition and data mining algorithms, is allowing personal data to be invisibly captured and analyzed. For example, in January 2001, cameras were used to scan the faces of people at the NFL

Super Bowl, with face recognition algorithms used to automatically match captured images with a database of known criminals.

Second, ubiquitous network connectivity is breaking down existing physical and social boundaries in local settings, often creating a mismatch between perception and reality. For example, people meeting in a room may have certain expectations about privacy, but a video camera streaming out images from that room may create a disparity between those expectations and what is actually taking place.

Third, improved storage capabilities are making it easier to keep large amounts of data, making captured data accessible at places and times far removed from its original context. For example, Grudin points out that emails and newsgroup postings written long ago may come back and haunt the author years later [14]. It is not difficult to imagine similar scenarios with automatically gathered sensor data.

Fourth, the decreasing cost and increasing usability of ubicomp technologies is lowering barriers to entry, making them available to almost anyone. Soon, even school-age children may be able to use simple scripts to collect information from sensors, in much the same way that "script kiddies" run denial-of-service attacks on large company web sites.

Purely technological solutions, such as anonymizers and electronic cash, are quite appealing because they are automatic and self-enforcing. Very little oversight and interaction is needed on the part of the people using these technologies. However, we believe that such solutions by themselves can achieve privacy goals only in limited situations, that is, they are necessary but not sufficient. As noted by legal expert Lawrence Lessig, practical privacy is shaped by four strongly interacting forces: markets, social norms, legislation, and technology [19]. All four of these forces are needed to solve the information privacy problems we will face.

However, research (such as [26, 29]) has shown that *asymmetric information*—situations where one side of a transaction has better information than the other—can contribute to *externalities*, situations where the cost and consequences of one individual's actions are shouldered by other uninvolved parties. These externalities can significantly influence the first three of these forces, namely market, social, and legal. Specifically, the existence of asymmetric information and externalities prevents these forces from being fully applied to achieve desired privacy goals.

Our position is that, in order to successfully address privacy concerns, ubiquitous computing technology must be designed to minimize asymmetry and externalities, so that market, social, and legislative forces can be fully brought to bear. While this will not solve all privacy problems, addressing these issues with the right technological infrastructure will make it easier for people to make more informed decisions, to evolve social norms as to what is appropriate and what is not, to enforce legal requirements, and to detect privacy violations.

In this paper, we propose a framework for addressing privacy concerns in a ubiquitous computing setting, based on the four-layer OM-AM (Objectives, Models, Architectures, and Mechanisms) framework developed by Sandhu [27]. In the OM-AM framework, the Objectives and Model articulate what the requirements are, while the Architecture and Mechanisms address how to meet these requirements. Our framework as discussed in this paper is focused primarily on Objectives for privacy, with informal discussion of the information space Model we have developed. A more formal treatment of information spaces can be found in [16].

In Section 2, we describe a simple ubiquitous computing scenario that is reused throughout this paper. In Section 3, we draw on social science research on *asymmetric information* and *externalities*, developing a key objective called the *Principle of Minimum Asymmetry*. The goal of this principle is to minimize the imbalance between the people about whom data is being collected (the *data owners*) and the systems and people that collect and use that data (the *data collectors* and *data users*).

In Sections 4 and 5, we outline the model we have developed for describing the interaction between these actors and personal data, called *Approximate Information Flow* (AIF). This model embodies three different abstractions about personal data, each of which describe a different perspective on privacy. The first abstraction, a storage perspective, is *information spaces*, a concept that describes where the data is stored, how it is used, and how it flows to other information spaces. The second abstraction, a dataflow perspective, is the *lifecycle* of such data, described in terms of *collection*, *access*, and *second use*. Collection is the point at which data is gathered, access is the point at which data is initially requested and used, and second use is downstream usage by others after initial access. The third abstraction, an end-user perspective, is a set of *themes for minimizing asymmetry*, described in terms of *prevention*, *avoidance*, and *detection.* Prevention deals with eliminating conditions where personal data may be misused, avoidance deals with minimizing the risk involved with using personal data, and detection deals with discovering improper uses of personal data. Section 4 deals with the first abstraction, information spaces, while Section 5 deals with the data lifecycle and the themes for minimizing asymmetry.

In Section 6, we discuss how the AIF model can be used to support different degrees of information asymmetry in ubicomp systems, and how it can be utilized to specify socially-compatible privacy objectives, suggest new privacy solutions and enable new methods of privacy inspection and certification for ubicomp systems. We present related work in section 7 and then conclude in Section 8.

2 An Example Scenario

Throughout this paper we will use the following ubiquitous computing scenario. As we present new concepts, we will describe them in the context of this scenario.

Alice is visiting a city in a foreign country. She decides to go to a local store and rent Bob[1], a handheld electronic tourguide that displays nearby points of interest. The Bob system uses a combination of GPS and infrared beaconing to track Alice's location, both indoors and outdoors. Her location is wirelessly sent to a centralized server, so that other people she is traveling with can find her.

3 A Social Foundation for Privacy in Ubiquitous Computing

Our approach is founded upon social science research about the impact of information on social behavior. These studies span a wide range of social science

[1] In this case, Bob is a ubiquitous computing application instead of a person.

fields, including economics, sociology, social psychology, and public policy. Two key ideas linking these studies are that of asymmetric information and externalities. In this section, we first look at how asymmetric information may lead to externalities with negative impacts on privacy. We then propose a new design principle called the Principle of Minimum Asymmetry, which is the primary objective of our privacy framework. This section closes with a discussion on how the Principle of Minimum Asymmetry applies to ubiquitous computing systems.

3.1 Defining "Asymmetric Information"

Environments with *asymmetric information* describe situations in which some actors hold private information that is relevant to everyone. Research on the impact of asymmetric information on social behavior started with Akerlof's work on used-car markets [2], for which he was awarded the 2001 Nobel Prize in Economics, and with Berg's work on education as a ticket to better jobs [4]. The word *relevant* covers many possibilities. The private information can be directly relevant in the sense that it directly affects the payoffs of the players. For example, when a consumer buys a used car, it may be very difficult for him to determine whether or not it is a good car or a lemon. In contrast, the seller of the used car probably has a pretty good idea of its quality. The private information held by sellers, especially unscrupulous ones, may lead to a "malfunctioning of markets," for example, one that is dominated by lemons.

On the other hand, the private information can also be indirectly relevant in that it helps each actor to anticipate the behavior of others. For example, in a bicycle-theft insurance market, insurance companies have a deep interest in knowing the actions taken by bicycle owners. High-risk owners either do not bother to lock their bikes or use only a flimsy lock, making their bicycles more likely to be stolen. However, insurance companies have a hard time distinguishing between high-risk and low-risk owners. Again, this leads to a malfunctioning system because insurance companies do not have a strong motivation to insure this market, penalizing low-risk owners.

Although not the sole cause, the existence of significant asymmetries in both information and power between different parties engaging in social exchanges is a leading contributor to the emergence of externalities. The notion of an *externality* was originally invented by economists to denote all the connections, relations, and effects that agents do not take into account in their calculations when entering into a market transaction [10]. For example, a chemical plant that pollutes a river with toxic products produces a negative externality for all other people that use the river. The chemical plant calculates its decision to exploit the resource without taking into account the effects of its actions on other's activities. As a result, the interests of fishermen are harmed. To pursue their activity, the fishermen will have to make investments for which they will receive no compensation, such as spending more money to clean up their fish before it is sold.

Figure 1 shows a graphical version of the scenario we presented in Section 2. We now frame this example in terms of *data owners*, the people about whom data is being collected; *data collectors*, the systems and the people that collect information about data owners; and *data users*, the systems and the people that use this information.

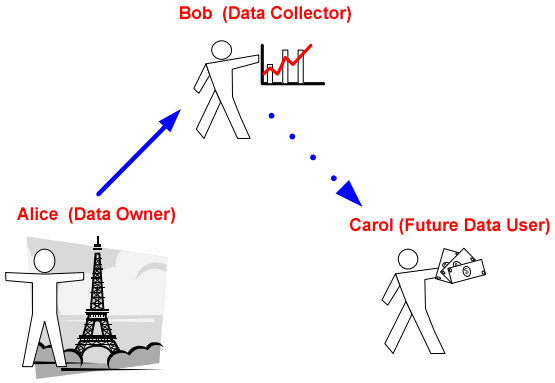

Fig. 1. Example scenario described in terms of externalities and asymmetric information. Bob has more information about Alice than vice versa, creating an asymmetry. Bob passing on Alice's data to Carol imposes an externality on Alice.

Alice (the data owner) is visiting a city in a foreign country. She rents the Bob system (a data collector), an electronic tourguide that uses GPS and infrared beaconing to track her location. The Bob system displays Alice's current location to her (so in this case, both Alice and Bob are acting as data users). At the same time, Alice's location is being sent to a centralized server, where it is permanently stored, ostensibly for performance profiling and for allowing Alice's friends find her. However, it turns out that the Bob system also collects this data for use by Carol, an advertising agent (making Carol a data user).

In this case, Bob and Carol know much more than Alice does about how any collected data will be used. Furthermore, Alice has little control over how her data will be used, by whom, and for how long. This gives rise to an asymmetry of information between Alice, Bob, and Carol.

Also, when Bob and Carol engage in data exchanges about Alice, they may create a negative impact on Alice without ever involving her, imposing a negative externality on her. For example, Carol might send unwanted spam to Alice based on where she went. As we will discuss in the next section, the existence of such asymmetric information has a significant negative effect on economic, social, and legislative dealings with the privacy problem.

3.2 The Effects of Asymmetry on Privacy

Legal expert Lawrence Lessig has pointed out that practical privacy is shaped by four strongly interacting forces: markets, social norms, legislation, and technology [20]. However, research has shown that asymmetric information and any resulting externalities are significant factors on the first three of these forces. Specifically, the existence of asymmetric information and externalities prevents these forces from being fully brought to bear to address privacy concerns.

With respect to market forces, economists have used asymmetric information and externalities to successfully explain a wide range of market behaviors, from the labor market to the health care market. More recently, leading economists have used these

tools to investigate the failures of the personal information market. Varian [29] points out that allowing third parties to buy and sell information imposes an externality on individuals, since the affected individuals are not directly involved in those transactions. Laudon [18] points out that the current crisis in the privacy of personal information is also a result of market failure. This failure has resulted in enormous asymmetries in both power and information, as well as large negative externalities for individuals. Noam [22] continues this line of thought, arguing that both a certain degree of information symmetry among the transacting parties, as well as a stable market free from large externalities, is necessary for a healthy market for personal information to succeed. In essence, market failures occur largely because the existence of asymmetric information imposes significant cognitive costs on individuals involved in data exchanges, and negatively impacts their ability to make informed decisions. This problem will only be exacerbated in ubiquitous computing environments due to the proliferation of data collection, thereby increasing the number of decisions one has to make regarding data exchanges.

With respect to social forces, externalities have been used to study the emergence of social norms [15]. Social norms are cultural phenomena that prescribe and proscribe behavior in specific environments, the emergence of which are key to trust formation and privacy concerns. Externality-based studies of social norms have found that norms largely arise to overcome negative externalities or to promote positive externalities in well-defined groups. For example, the fact that a neighbor plays loud music at 3 AM in the morning creates a negative externality for his neighbors. However, social norms have evolved to the extent that most people would turn down the volume or shield their windows when playing music that early in the morning. Violating social norms makes one vulnerable to social sanctioning, such as increased isolation in the community or decreased cooperation from neighbors when their help is needed. However, social sanctioning is contingent on easy detection of violations of social norms. In future ubicomp environments, individuals may have little knowledge or control about how their data may be used and by whom. A negative cost, such as being stalked by an unwanted suitor, may be imposed at a much later time, making it practically impossible to detect how a privacy violation happened or why. Under these conditions, the presence of asymmetric information has made externalities much harder to overcome.

With respect to legislative forces, legal scholars and public policy specialists have also considered the impact of asymmetric information and externalities on legislative approaches to privacy. Law professor Pamela Samuelson has recently discussed problems with applying property rights legislation to personal information [26]: privacy may not be achievable unless the default rule of the new property rights regime limits transferability of property rights. Furthermore, the presence of significant asymmetric information makes it very difficult for the average person to judge the risks of selling his property rights in personal data. Samuelson therefore proposed licensing of personal information as an alternative means to legally protect privacy. However, she also acknowledges the futility of any licensing regimes unless individuals are informed about and can exert real control over their personal data.

In the previous example of data exchanges between Alice, Bob, and Carol (see Figure 1), the existence of asymmetric information makes it hard for Alice to assess privacy risks associated with data sharing with Bob. For example, Alice's personal

data is sent by Bob to Carol without Alice's knowledge or control. This asymmetry makes privacy violations more immune to social and legal sanctioning that would otherwise be possible through legislation and development of social norms.

3.3 The Principle of Minimum Asymmetry

The presence of asymmetric information and negative externalities are at the heart of the information privacy problem. Negative externalities are often much harder to overcome in environments with significant asymmetry in both information and power between different parties.

Our position is that the role of any technical approach in addressing privacy concerns should be to minimize the asymmetry between data owners on one side, and data collectors and data users on the other. Based on these observations, we have derived the following principle for achieving privacy in ubicomp called the *Principle of Minimum Asymmetry*.

Principle of Minimum Asymmetry

A privacy-aware system should minimize the asymmetry of information between **data owners** and **data collectors and data users**, by:

- **Decreasing** the flow of information from data owners to data collectors and users
- **Increasing** the flow of information from data collectors and users back to data owners

3.4 Implications for Privacy-aware Ubiquitous Computing

The goal of the Principle of Minimum Asymmetry is to reduce information asymmetry within a given application context, which will facilitate market, social, and legal recourses in addressing privacy concerns. Returning to our scenario, the source of asymmetry comes from the fact that Bob collects a great deal of information about Alice, while Alice knows very little about how Bob uses that information.

Applying the Principle of Minimum Asymmetry, we can either decrease the information flow out from Alice, or increase the flow of information back to Alice. Examples of decreasing the information flow out include anonymizing or pseudonymizing Alice's data, increasing the granularity of the location information, decreasing the rate at which location information is sent back to the server, and increasing the control over who can access the data and under what conditions. Some of these techniques, such as anonymization, can be applied either before the data is stored by Bob or before the data is sent onwards to Carol. Examples of increasing the information flow back to Alice include logging of all accesses about Alice's location, notification when someone accesses Alice's location information, and clear feedback on what kind of information is being stored.

Adding some or all of these mechanisms would allow Alice to have a better understanding of what the privacy risks are and to make more informed decisions. These mechanisms would also make it easier to seek market, social, and legal recourses. An example of applying market forces is that people can publish reviews of

competing tourguide systems if they have a better understanding of how personal information collected by the systems are used. An example of social forces is that people might be less likely to access someone else's information intrusively if that access will be logged and will also notify the data owner. An example of legal forces is that notifications and logs of how an individual's data is being accessed can be used to foster accountability and detect violations of any laws.

It is important to note that minimum asymmetry is a relative notion rather than an absolute one. Some degree of information asymmetry will always exist. For example, a person in authority by definition will have more knowledge about data use then an average person. Law enforcement and management reasons may even render some level of asymmetry desirable. Furthermore, different degrees of asymmetry will be shaped by a wide variety of application design goals, including efficiency, convenience, safety, accountability, usability, business, marketing, and privacy. So the question is not how to eliminate asymmetric information in its entirety but how to strike a balance to achieve a more equitable distribution of risk and protection for a given application context. In the next section, we describe Approximate Information Flow, a model for describing the interactions between actors and personal data that can incorporate varying degrees of asymmetry.

4 Approximate Information Flow: Information Spaces

In this section, we describe Approximate Information Flow (AIF), a novel model for privacy-aware ubiquitous computing architectures that embodies the Principle of Minimum Asymmetry. The information flow is called "approximate" because data representing the same content can be acquired with different levels of confidence, transferred at different levels of accuracy, and live for different periods of time. Each of these factors has varying implications for privacy.

"Model" is an overused term that has been used to describe everything from a philosophical standpoint to a particular implementation method. The AIF privacy model we describe in this paper is close to the Model concept in Sandhu's OM-AM framework [27]. Rather than specifying a particular method for enforcing privacy, our AIF privacy model supplies key sets of abstractions describing information flow within a system of people and computers.

The first abstraction is *information spaces*, which is a collection of data delimited by physical, social, or activity-based boundaries. Personal data is stored in and used within an information space, and may flow to other information spaces. The second abstraction describes the lifecycle of personal data, consisting of *collection*, *access*, and *second use*. The third abstraction is a set of themes for minimizing asymmetry, consisting of *prevention*, *avoidance*, and *detection*.

Although these three abstractions seem different, they are actually different facets of the same thing. Information spaces describe the collection, management, and sharing of personal information from a storage perspective. In contrast, the data lifecycle describes this from a dataflow perspective, and the set of themes for minimizing asymmetry describe this from an end-user perspective.

In this section, we focus on describing the first abstraction, information spaces. In section 5, we combine the second and third abstraction to create a new design space for categorizing privacy protection mechanisms, and in section 6 we show how AIF can be utilized to support varying degrees of information asymmetry in ubicomp.

4.1 Information Spaces

The central notion of AIF is that of information spaces. Information spaces are repositories of personal data owned by data owners, data collectors, or data users. Each of these *principals* might represent a specific person, a particular device, or even a smart room infrastructure managing the activities within that room. The data contained in an information space might be about the principals (e.g., a person's location) or an ongoing activity (e.g., there is a meeting in this room). There are three important privacy-sensitive properties of data contained in an information space:

- **Persistence of data**: Persistence refers to the lifetime of data and whether its quality should degrade over time. For example, a video recording of a class may only be allowed to live until the end of the current semester.
- **Observational accuracy of data**: The more features a data item contains about its owner, the more "accurate" it is. For example, a context-aware phone forwarding application might need to know precisely which room someone is in, while a map service would need just the building. As another example, a video file might be blurred to different extents depending on need. As a third example, a person's location might be updated every second, every ten seconds, or every sixty seconds.
- **Observational confidence of data**: Observational confidence measures the uncertainty of data. The unreliable nature of most sensors and the increasingly prominent recognition-based interface has made it almost impossible to collect any data with 100% certainty. For example, if a sensor can only be 50% sure about one's location, release of such data might not be as risky as if it were 90% sure.

Information spaces are not necessarily bound to physical locations, devices, or the way data is managed. Data collected in one's home and private office may belong to the same information space, even though they actually reside at different physical locations, on different devices, and are managed differently. There are three different types of *boundaries* that can serve to delimit an information space:

- **Physical boundaries:** Physical boundaries separate one information space from another through physical locations. For example, you might have one information space for all information collected in your office and another for your home.
- **Social boundaries:** Social boundaries separate one information space from another through social groups. For example, all the data created in a work group could be defined to be jointly owned by the group, no matter where the data is or how it is created.
- **Activity-based boundaries:** Activity-based boundaries separate information spaces from one another through activities the space owners are involved in. For example, all conversations during John's public speech belong to an information space owned by the general public, while his after-speech chats with members of audience do not.

Each information space also has specific privacy-sensitive operations that can be applied to the data within that space. We define five types of such operations below. Logging and user notification are implicitly associated with all of these operations.

- **Addition/Deletion/Update:** Addition, deletion, and update are the same familiar operations performed routinely on databases.
- **Authorization/Revocation:** Principals use authorization and revocation to change ownership and release policies regarding data in their information spaces.
- **Promotion/Demotion:** Promotion increases privacy risks by making data live longer, become more accurate, or be observed with a higher level of confidence. Demotion does exactly the opposite.
- **Composition/Decomposition:** Principals can combine data from different sources or split data into separate pieces. For example, location data can be combined with activity data to give an idea of whether a person is busy or not.
- **Fusion/Inference**: Higher-level information can be inferred from raw data, such as inferring an ongoing meeting using vision-based analysis of activities in a room.

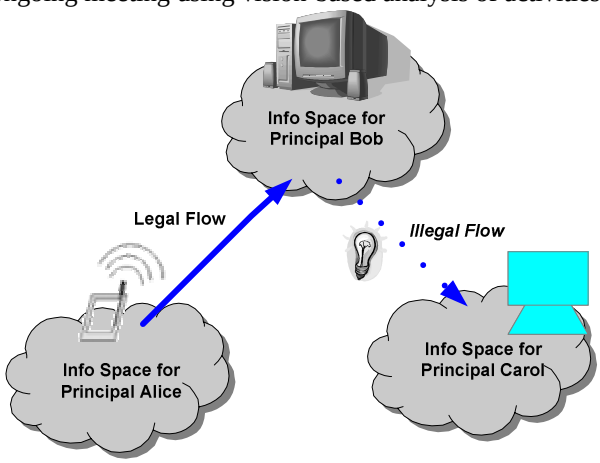

Fig. 2. Example scenario described in terms of information spaces and flows. Since Alice has authorized it, an information flow from Alice to Bob's centralized server is a legal flow. However, since it is not authorized, a flow from Bob to Carol is illegal.

Figure 2 shows the scenario presented in Section 2, framed in terms of information spaces. Alice is visiting a city in a foreign country. She rents the Bob system, an electronic tourguide. In this instance, the Bob system acts as its own information space. When Alice's location is sent to a centralized server, Alice's data flows from Bob's information space to the centralized server's information space and is added there. When Carol accesses the centralized server, Alice's data flows from the centralized server to Carol's information space and is added there.

Policies can be applied to decrease the information flow out from Alice. For example, Alice's data can be demoted by anonymizing it before it is sent to the centralized server or before it is sent onward to Carol. Alternatively, Alice can specify that no one is authorized to access her data. Policies can also be applied to increase

the information flow back to Alice. For example, the Bob tourguide system can be instrumented to notify Alice if her data flows out of that information space.

4.2 Discussion of the Information Spaces Abstraction

An information space is properly viewed as a semantic construct around which information flow control policies can be formulated. These control policies are embodied in various components of AIF, such as the definition of information spaces, privacy restrictions of operations allowed within an information space, and restrictions of legal flows between different information spaces. These components collectively determine allowable usage policies for personal data.

The idea of an information space draws its insights from architectural design and environmental psychology. To a large extent, the architectural design of a physical space shapes the activities that take place within it. For example, Agre has discussed the relationship between activities and places using the example of a theater [1]. A theater as a space reflects a set of social relations between the actors and the audience, between those who have been admitted into the seating areas and those who have not, and between the people with expensive tickets and the people with cheap tickets. The theater also assigns every activity to a place, such as dressing in the dressing room, performing on the stage, watching from the seats, and buying tickets in the lobby.

As in physical spaces, information spaces seek to provide similar structural resources and constraints for organizing a complex and privacy-sensitive social process. However, information spaces are not separated only by physical boundaries, as deployment of ubicomp technologies has effectively disrupted a clean mapping between activities and places. For example, an incoming call to your cell phone may disturb the entire audience in a theater. As such, this abstraction serves as the foundation for a new privacy-aware system architecture of virtual spaces.

5 Approximate Information Flow: Data Lifecycle and Themes for Minimizing Asymmetry

In the previous section, we described information spaces, the first abstraction in AIF. This section describes the second and third abstractions. The second abstraction describes the lifecycle of personal data, consisting of *collection*, *access*, and *second use*. The third abstraction is a set of themes for minimizing asymmetry, consisting of *prevention*, *avoidance*, and *detection*. In this section, we describe these abstractions, and combine them to create a new design space for categorizing privacy protection mechanisms in ubiquitous computing.

5.1 Data Lifecycle: Collection, Access, and Second Use

According to an August 2000 Pew Internet & American Life Project Poll, 86% of Internet users favor “opt-in” privacy policies requiring companies to ask people for permission to use their personal information. Another 84% were “concerned” with

"businesses and people you don't know getting personal information about you and your family" [23]. This last result is corroborated by a March 2000 Business Week/Harris Poll, which showed that 88% of people wanted Web sites to ask for permission before sharing one's personal information with others [7].

The observation here is that people are concerned not only about how their data is collected and used, but also who else gets to use the data. Thus, the exclusive focus on initial data access of current solutions is inherently inadequate, and there is no reason to believe that there will not be similar results with respect to ubiquitous computing.

For this reason, we have developed the data lifecycle, which looks at what happens to data over time, and the different privacy concerns that can arise at these times. We have separated this lifecycle into three phases: collection, access, and second use. Important privacy decisions are made in each of these stages.

Collection refers to the point at which data is gathered. Important decisions made at this phase include what data should be collected, in what form, how long it will persist, how accurate the data is, and how confident the system is in the data. In our scenario, collection occurs when Bob gathers data about Alice's location from GPS.

Access refers to the point at which data is initially requested. Important decisions made at this phase include what data can be accessed, how accurate and how confident the data is, who should be allowed to access it, for what purpose, and under what circumstances. In our scenario, access takes place when Bob uses Alice's location information to display nearby points of interest, and when Bob uploads Alice's data to the centralized server.

Second use refers to use and sharing of data after initial access has been made. Second use also includes passing data from one party, who might have authorized access to the data, to another party, who might not. Consequently, data owners often have very little control over second use. Important decisions made at this phase include who else should be able to access the data, what they can do with it, and whether they should be allowed to share it even further with others. In our scenario, second use occurs when Carol accesses Alice's data from the centralized server.

5.2 Themes for Minimizing Asymmetry: Prevention, Avoidance, and Detection

We have categorized technology-based privacy protection mechanisms into three themes: prevention, avoidance, and detection. *Prevention* seeks to ensure that undesirable use of private data will not occur. For an information space, such mechanisms attempt to prevent possible misuse (1) by reducing the persistence, accuracy or confidence of data to an acceptable level, or (2) by eliminating privacy-risky operations. For example, pseudonymization or query randomization techniques are used to reduce data accuracy. Video data can be masked to reduce its confidence measurement, and data can be frequently garbage-collected to reduce its persistence. In addition, computation performed on encrypted data has the goal of eliminating any privacy-risky operations. *Avoidance* seeks to minimize the risks and maximize the benefits associated with data exchanges by carefully considering the context in which they take place. This is often done through properly informing the end-user of privacy risks. Examples of avoidance through humans include explicit end-user consent to specific policies and notification on transmission of personal data. In contrast,

detection assumes that some undesirable use will occur, and seeks to find such incidents in the hope that privacy violators will be held accountable. Though not directly related to ubiquitous computing, an example of detection is putting your phone number into a search engine to see where it is listed. Another example is credit rating services, such as Equifax [12], which, for a fee, keeps track of changes to a person's credit rating and notifies them whenever someone accesses that data.

Each of these themes reduces asymmetry of information in different ways. Prevention decreases and controls the flow of information from data owners to data collectors and users. Avoidance simultaneously decreases the flow of information out and increases the flow of information in. Detection increases the flow of information back to data owners about collection, access, and second use of their data.

5.3 A Design Space for Privacy Solutions in Ubiquitous Computing

Combining the data lifecycle with the set of themes for minimizing asymmetry leads to a new design space for categorizing privacy protection mechanisms (see Figure 3). In this design space, anonymization and pseudonymization technologies (such as [5]) are preventative measures that can be used on collection or access. These techniques can also be applied in second use, though there is no guarantee for data owners that this will actually be done. In contrast, the Platform for Privacy Preferences (P3P) [9], a system that tells web users how their information will be used, would be an avoidance measure that is used mainly during the collection phase.

Traditional role-based access control models (RBAC) [13] offer an elegant solution to the problem of managing complex access control rule sets. Instead of all-or-nothing

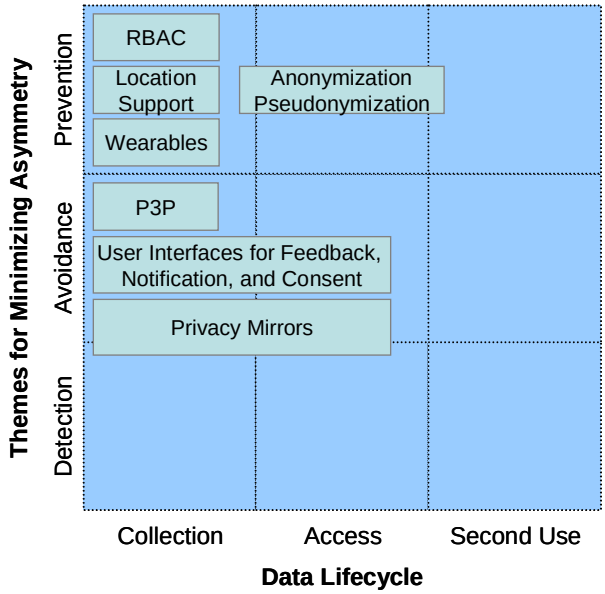

Fig. 3. A Design Space of Privacy Solutions in Ubiquitous Computing

privileges of super-user and normal user, organizations can create varying levels of privilege and assign them to different people. Covington et. al. [8] extended traditional RBAC models with "environment roles" to provide support for data access decisions in ubicomp environments. For example, a system could make access decisions based on the requestor's current location, in addition to the requestor's role. Such a scheme would be a preventative measure affecting access of data.

Different architectural styles can also be described in terms of this design space. For example, the Cricket system [24] beacons out location information to end-users, operating more as a location-support system instead of a location-tracking system. Similarly, researchers at EuroPARC have avoided storing user location information in a central repository, keeping it instead at that person's own workstation [30]. Other researchers have suggested using wearable computers to store sensitive personal information [25]. These architectures are preventative measures for collection because the owner of the data has full control over the computer storing the data.

Researchers have also focused on designing privacy interfaces that provide users with appropriate feedback about data collection. For example, Bellotti and Sellen have done some early work on how to provide such feedback in collaborative multimedia environments [3]. Their system provides many cues to help end-users avoid risky situations during collection. In general, privacy user interfaces for feedback, notification, and consent are focused on avoidance at collection and access, and could also be used at second use.

More recently, Nguyen has proposed a new privacy interface called the Privacy Mirror [21] for ubiquitous computing environments. This interface provides feedback to end-users, showing them what information is being collected, and what information has been accessed and by whom. Privacy mirrors provide feedback and detection mechanisms to help end-users avoid risky situations at collection and initial access.

Our formulation suggests that previous ubicomp privacy research has explored only a small portion of the entire design space, primarily preventative mechanisms for collection and initial access of data. We believe this narrow focus is inadequate for addressing privacy concerns since preventative privacy solutions rely on many assumptions about system configuration and user behavior which may not always hold. For example, pseudonymization is effective only if identity cannot be easily inferred from user behavior. As another example, location support [24], where the infrastructure tells you where you are instead of tracking you, is effective only if every user carries a personal device capable of performing local computations based on their location. For this reason, we believe that ubiquitous computing privacy research needs to be expanded to explore other areas in the design space.

6. Supporting Varying Degrees of Information Asymmetry

In this section, we discuss how the AIF model can support varying degrees of information asymmetry in ubicomp systems. In the AIF model, the information space abstraction effectively defines "privacy zones" delimited by physical, social or activity-based boundaries. Approximate information flows that come into or out from an information space determine the degree of information asymmetry in that space

through the amount of imbalance they introduce. Not only are the degrees of asymmetry affected by the presence of information flows, they are also affected by the privacy-sensitive features of data that these flows carry. In the AIF model, data stored in an information space contains descriptions of privacy-sensitive properties such as persistence, accuracy, and confidence. These properties affect the amount of identifiable information contained in a flow, thereby affecting the flow itself.

As we have seen, the three themes for minimizing asymmetry affect information flows between spaces in different ways. They can either decrease the information flow out from an information space (i.e., prevention or avoidance), or increase the flow of information back into a space (i.e., avoidance or detection). For each information space, we can combine these themes in different ways to support different degrees of information asymmetry.

Moreover, when applied to data during different stages of their lifecycle, even the same asymmetry-minimizing theme has different effects on the flow of information into or out from an information space. For example, in some cases, detection of second use privacy violations can offer stronger privacy protection than detection of initial access privacy violations, because second use violations may not always be apparent. In AIF terms, applying detection measures during second use of data further increases the flow of information back into a data owner's information space by providing her with more knowledge about privacy violations.

The ability of AIF to support varying degrees of information asymmetry in ubicomp systems can be utilized in different ways. The first use of AIF is to describe socially-compatible privacy objectives and requirements. On one hand, AIF defines "privacy zones" using the information space abstraction, and information asymmetry in terms of approximate information flows between information spaces. On the other hand, end-user privacy objectives are framed in terms of desired degrees of information asymmetry for the information spaces they own. As such, the AIF model is able to specify the diverse privacy objectives required by ubicomp.

The second use of AIF is to suggest new privacy solutions in the design space that minimizes information asymmetry for a given application context. As can be seen from Figure 3, so far there have not been any strong solutions for second use or detection of incidents of privacy violations. Preventing and avoiding illegal second use of personal data in the general case is very hard. For example, it is difficult to imagine any realistic mechanism to prevent people from memorizing sensitive data for later use once they see it. For this reason, we believe it is important to work on detection mechanisms in addition to prevention and avoidance.

The third use of AIF is that it provides an initial framework for inspecting and certifying privacy-friendliness of complex ubicomp systems. In the online world, the TRUSTe "trustmark" privacy seals [28] are used to certify privacy-friendly web sites. However, this paradigm cannot be easily transferred to ubicomp environments. It is not clear what would constitute the basic units for privacy certification, nor how such certification can be done. The AIF model provides a first step in developing powerful methods for privacy inspection and certification in ubicomp environments. It allows decomposition of complex ubicomp systems into basic units of privacy inspection, that is information spaces. And each information space can be inspected by analyzing the degree to which a desirable degree of information asymmetry it supports.

7 Related Work

As suggested by Langheinrich, privacy solutions in ubiquitous computing have so far been largely ad-hoc and specific to the system at hand [17]. He therefore proposed to develop privacy-aware ubiquitous computing systems based on the framework of *fair information practices (FIP)* [31], a significant policy development in the 1960s and 1970s that influenced privacy legislation worldwide.

FIP are a collection of abstract philosophical statements about privacy protection, such as transparency, collection limitation, use limitation and accountability. Each of these varies significantly in different real world contexts. For example, one can expect much less transparency and more accountability in a military than in a civilian environment. Real ubicomp systems need to be able to support a diverse array of applications with different degrees of transparency, limitation, and accountability. In other words, while the FIP provide us a basic understanding of important principles involved in privacy protection, our work seeks to offer a concrete way to support the diverse and application-specific privacy objectives in a rigorous framework.

Bellotti and Sellen have developed a framework for designing feedback and control in collaborative multimedia environments [3]. Such frameworks are useful, but they primarily exist to prescribe a design process for determining a desirable degree of asymmetry for a given application context. In contrast, the goal of AIF is to provide a model that can be easily configured to support varying degrees of asymmetry for a diverse array of application contexts.

Brin has proposed the "transparent society" as a framework for addressing privacy problems in general [6]. His argument is that instead of pushing towards stronger privacy mechanisms, information should be more freely shared among all of us, with everyone watching each other. In essence his proposal is to completely eliminate information asymmetry by granting everyone equal access. As we have discussed earlier in the paper, complete elimination of asymmetry is highly unlikely, and undesirable in many cases. In comparison AIF provides a more realistic framework by focusing on supporting varying degrees of asymmetry.

8 Conclusions and Future Work

In this paper, we have framed privacy in terms of the flow of information, with privacy itself being about control over this flow. One of the existing problems with privacy is that there is often a large asymmetry of information, with one party having much more information about another party. Such asymmetry often creates a negative externality, imposing a burden on people without their consent.

To address this problem, we proposed the Principle of Minimum Asymmetry, which seeks to rectify this imbalance, either by decreasing the flow of information out or increasing the flow of information in. The ultimate goal of this principle is not to provide a purely technological solution to privacy, but to make it easier for market, social, and legal forces to be applied to address legitimate privacy concerns. This principle can be applied to ubicomp systems as a design goal for privacy.

We also introduced Approximate Information Flows (AIF) as a model for describing the various actors involved when dealing with personal data. AIF has three key abstractions, each of which describe the collection, management, and sharing of personal information from different perspectives. The first abstraction, a storage perspective, is information spaces, which are repositories of personal data. Operations can be applied to data within an information space, and data flows between different information spaces. The second abstraction, a dataflow perspective, is the data lifecycle, which consists of collection, access, and second use. Each of these phases represents a different way of how data is used, and each affects privacy differently. The third abstraction, an end-user perspective, is a set of themes for minimizing asymmetry, which consists of prevention, avoidance, and detection. These three abstractions can be used to analyze the degree of asymmetry in ubiquitous computing applications.

By combining the last two abstractions, the data lifecycle and the set of themes for minimizing asymmetry, we introduced a new design space for privacy technologies. We have also described how the AIF model can be used to support different degrees of information asymmetry in ubicomp systems, and how it can be utilized to specify socially-compatible privacy objectives, suggest new privacy solutions, and enable new methods of privacy inspection and certification for ubicomp systems.

Our framework is a four-layer one, including Objectives, Models, Architectures, and Mechanisms. In this paper we have focused primarily on Objectives for privacy. In [16] we developed a more formal model of information spaces for privacy control in ubicomp systems. Currently we are developing a suite of new privacy mechanisms based on the information space model, as part of a general infrastructure for context-aware computing.

Acknowledgements

We would like to thank John Canny, Anind Dey, Jen Mankoff, Scott Lederer, and all anonymous reviewers for their invaluable help in shaping this work.

References

1. Agre, P., *Changing Places: Contexts of Awareness in Computing.* Human-Computer Interaction, 2001. **16**(2-4): p. 177-192.
2. Akerlof, G., *The Market for "Lemons":Quality Uncertainty and the Market Mechanism.* Quarterly Journal of Economics, 1970.
3. Bellotti, V. and A. Sellen. *Design for Privacy in Ubiquitous Computing Environments*. In *The Third European Conference on Computer Supported Cooperative Work (ECSCW'93).* 1993. Milan, Italy: Kluwer Academic Publishers.
4. Berg, I., *Education and Jobs: The Great Training Robbery*. 1970, New York: Praeger.
5. Biskup, J. and U. Flegel. *Threshold-Based Identity Recovery for Privacy Enhanced Applications*. In *7th ACM Conference on Computer and Communications Security (CCS 2000)*. 2000. Athens, Greece: ACM.
6. Brin, D., *The Transparent Society*. 1998, Reading, MA: Perseus Books.

7. BusinessWeek, *Business Week/Harris Poll: A growing threat*. 2000. http://www.businessweek.com/2000/00_12/b3673010.htm
8. Covington, M.J., Long, W., Srinivasan, S., Dey, A.K., Ahamad, M., Abowd, G.D. *Securing Context-Aware Applications Using Environment Roles*. In *6th ACM Symposium on Access Control Models and Technologies (SACMAT 2001)*. 2001.
9. Cranor, L., et al., The Platform for Privacy Preferences 1.0 (p3p1.0) Specification. 2000. http://www.w3.org/TR/P3P/
10. Ellis, H.S. and W. Fellner, *External Economies and Diseconomies.* American Economic Review, 1943. **33**: p. 493-511.
11. Epic.com, http://www.epic.org/privacy/,
12. Equifax, http://equifax.com/
13. Ferraiolo, D., J.A. Cugini, and D.R. Kuhn. *Role- Based Access Control (RBAC): Features and Motivation*. In *Eleventh Annual Computer Security Applications Conference*. 1995.
14. Grudin, J., *Desituating Action: Digital Representation of Context.* Human-Computer Interaction (HCI) Journal, 2001. **16**(2-4): p. 269-286.
15. Horne, C., *Sociological Perspectives on the Emergence of Norms*, in *Social Norms*, K. Opp, Editor. 2001, Russell Sage: New York.
16. Jiang, X. and J. Landay, *Modeling Privacy Control in Context-aware Systems Using Decentralized Information Spaces.* to appear in IEEE Pervasive Computing, 2002. **1**(3).
17. Langheinrich, M. *Privacy by Design - Principles of Privacy-Aware Ubiquitous Systems*. In Proceedings of *Ubicomp 2001*. 2001. Atlanta, GA.
18. Laudon, K.C., *Extensions to the Theory of Markets and Privacy: Mechanics of Pricing Information*, in *Privacy and Self-Regulation in the Information Age*. 1997, US Department of Commerce.
19. Lessig, L. *The Architecture of Privacy*. In *Taiwan NET'98*. 1998. Taipei, Taiwan.
20. Lessig, L., *Code and Other Laws of Cyberspace*. 1999, New York NY: Basic Books.
21. Nguyen, D.H. and E.D. Mynatt. *Privacy Mirrors: Making Ubicomp Visible*. In *Human Factors in Computing Systems: CHI 2001 (Workshop on Building the User Experience in Ubiquitous Computing)*. 2001. Seattle, WA: ACM Press.
22. Noam, E.M., *Privacy and Self-Regulation: Markets for Electronic Privacy*, in *Privacy and Self-Regulation in the Information Age*. 1997, US Department of Commerce.
23. Pew Internet & American Life, Trust and Privacy Online: Why Americans Want to Rewrite the Rules, 2000. http://www.pewinternet.org/reports/pdfs/PIP_Privacy_Questionnaire.pdf
24. Priyantha, N.B., A. Chakraborty, and H. Balakrishnan. *The Cricket Location-Support System*. In *MobiCom 2000: The Sixth Annual International Conference on Mobile Computing and Networking*. 2000. Boston, Massachusetts: ACM Press.
25. Rhodes, B.J., N. Minar, and J. Weaver. *Wearable Computing Meets Ubiquitous Computing: Reaping the best of both worlds*. In *The Third International Symposium on Wearable Computers (ISWC '99)*. 1999. San Francisco, CA.
26. Samuelson, P., *Privacy As Intellectual Property?* 52 Stanford Law Review 1125, 2000.
27. Sandhu, R. *Engineering Authority and Trust in Cyberspace: the OM-AM and RBAC way*. In *5th ACM Workshop on RBAC*. 2000. Berlin.
28. TRUSTe, http://www.truste.org
29. Varian, H.R., *Economic Aspects of Personal Privacy*, in *Privacy and Self-Regulation in the Information Age*. 1997, US Department of Commerce.
30. Weiser, M., *Some Computer Science Problems in Ubiquitous Computing*, in *Communications of the ACM*. 1993. p. 75-84.
31. Westin, A.F., *Privacy and Freedom*. 1967, New York NY: Atheneum.