The Inference Problem and Pervasive Computing

Catherine Dwyer
Pace University
cdwyer@pace.edu

ABSTRACT

The promise of pervasive computing systems, which flood an environment with sensors and smart devices, has been diminished by an inability to adequately address privacy concerns. Measurements of privacy demonstrate a high level of public concern, yet designers of pervasive computing systems do not find these measurements relevant to the use of their systems. This has led designers to distrust reports of privacy unease. A potential solution to this impasse may be found in a privacy condition from the data mining literature known as the "inference problem." The inference problem occurs in data mining when confidential information can be derived by released data. Privacy issues within pervasive computing systems are most acute when they involve a variation of the inference problem. This paper will discuss methods used in data mining to limit the inference problem, and discuss their application to pervasive computing. Finally, a research agenda will suggest how to identify instances of the inference problem within pervasive computing systems.

Keywords

Privacy, Pervasive Computing, Data Mining, Inference Problem

INTRODUCTION

In the futuristic vision of pervasive computing proposed by Mark Weiser (Weiser and Brown, 1998; Weiser, Gold and Brown, 1999), technology disappears as it becomes a built-in part of walls, tables, furniture, and clothing. Pervasive is defined as "spreading or spread throughout," (Mifflin, 2004). The word ubiquitous, meaning "omnipresent, being present everywhere at once" (Mifflin, 2004), is also used to describe these systems.

In pervasive computing, as technology disappears, it simultaneously becomes more powerful, anticipatory, and invaluable. The significance of disappearing technology is that the person no longer has to adapt themselves to technology, now the technology adapts itself to the person. Weiser's idea for pervasive computing is that people shouldn't have to adjust their work habits to the machine, the machine should adjust to people. For example, rather than keeping track of where a computing resource may be, such as a file, server, scanning device, or laptop, instead your need is communicated to a pervasive computing system, and the system obtains the needed resources and provides them promptly and appropriately. Ideally, through continuous and anticipatory monitoring, pervasive computing systems can guess what you might need before you even consider it yourself (Weiser and Brown, 1998). In this vision, pervasive computing systems are extensions of the human mind (Weiser and Brown, 1998).

So like a top flight concierge, pervasive computing systems are intended to support and optimize human intelligence, anticipating and fulfilling needs in real time. The architecture of pervasive computing accomplishes this by blanketing an environment with sensors, context aware notifications, and opportunities for computer interaction with smart devices that are constantly with us or are embedded into the environment (Crang and Graham, 2007).

The promise of these systems remains in the realm of science fiction, in large part due to privacy concerns that remain thorny challenges. In this paper, the nature of these concerns will be presented by summarizing research that documents privacy problems within pervasive computing environments. This paper will argue these problems are not incurable, but are caused by an operationalization that defines online privacy as an individual's control of information unique to themselves. Simply put, it is not technically feasible to give an individual veto power over millions of bits of sensor data collected about them, nor do privacy implementations of this level of control reassure the individual that their privacy is being protected.

This paper argues that a new operationalization of privacy is needed for pervasive computing systems. The paper then introduces a condition from the data mining literature known as the "inference problem," and argues for its relevance in managing privacy problems within pervasive computing systems. The inference problem occurs in data mining when confidential information can be derived by released data. This paper will discuss methods used in data mining to manage the inference problem, and their application to pervasive computing systems. Finally, a research agenda will be proposed to explore the relevance of the inference problem to privacy problems within complex pervasive computing environments.

ARCHITECTURE OF PERVASIVE COMPUTING

The philosophy of pervasive computing flows from a conception of the computer as a tool to support the mind. The earliest expression of this idea is the famous essay by Vannevar Bush, "As We May Think" (Bush, 1945). The idea of computing as a tool for the extension of human intelligence was brought to the general public through the pioneering work of J.C.R. Licklidder and Douglas Englebart, who created the mouse and the graphical user interface (Waldrop, 2000).

Pervasive computing was advanced through the work of Mark Weiser and others at the Xerox Palo Alto Research Center in the late 1980s. Weiser introduced the concept of the disappearing computer (Weiser et al., 1999). It implies a move away from the desktop computer, to a scenario where computing resources are embedded everywhere in the environment, in the form of sensors, context aware notifications, and opportunities for computer supported display and interaction. The computer disappears, but continuously senses the user's presence and needs, and acts accordingly.

A central functionality of pervasive computing systems is the monitoring of location and contextual awareness (Dwyer, Hiltz and Jones, 2006). For example, in a large office complex, you would like to know where your colleagues are located. If they are not in their office, it may be impossible to find them without technology. You can call them on their cell phone and ask "where are you?" Or a pervasive system can give you their exact location.

Consider the difficulties that can arise when individuals collaborate on projects. Managing a collaborative project can be an enormous headache if everyone keeps their files separately on their own machines. A pervasive system can offer a solution to this management problem. When a meeting is held in a conference room, the pervasive computing systems can determine who is attending a meeting, and enable their digital files to be displayed for the group. Pervasive systems can also assist in managing changes made to documents, and enable synchronization with each individual's work.

PRIVACY CHALLENGES IN PERVASIVE COMPUTING

In the collaboration scenario described above, location aware and context aware functionality support an increase in the overall productivity of personal interaction and coordination. Because this is accomplished through location tracking, this functionality can trigger severe privacy anxiety. As an indication of this anxiety, consider the ongoing debate regarding the use of location information in cell phones. Many cell phones and PDAs can be easily configured for location awareness applications, often with the installation of a simple patch (see http://www.accutracking.com). The pieces seem to be all in place, and yet mobile location awareness initiatives such as E911 have triggered resistance and obstacles (Eng, 2005).

A recent Google search with the keywords "cell phone" and "big brother" generated 856,000 hits, with titles like "Will Big Brother Track You By Cell Phone?" and "Privacy Advocates Attack Cell Phone Surveillance." The tension between location awareness and privacy has been noted in pervasive computing research (Consolvo, Smith, Matthews, LaMarca, Tabert and Powledge, 2005; Hudson and Smith, 1996; Smith, 2005).

Projects that deliver "ambient intelligence" have been developed at universities around the country (Sorensen, 2003). While the systems work, the inability to reduce privacy concerns will likely delay further expansion of these systems. While surveys indicate privacy is a substantial public concern, the "big brother" reaction to pervasive systems has not been taken seriously by the research community (Iachello, Smith, Consolvo, Chen and Abowd, 2005). For both the privacy and security of its users, pervasive computing systems must protect the integrity of personal information.

THE OPERATIONALIZATION OF PRIVACY IN COMPUTING SYSTEMS

Why has it been so hard to offer privacy assurances within pervasive computing systems? One reason is the difficulty defining what privacy actually means (Solove, 2008). Privacy has social and cultural contexts that have developed over thousands of years (Lawler and Molluzzo, 2005). Privacy is often describes as a collection or a family of concerns. For example, Tavani describes three types of privacy, which are 1) Accessibility privacy – freedom from intrusion, 2) Decisional privacy – freedom from interference in your personal choices, for example reproductive rights, and 3) Informational privacy – person's ability to manage the sharing and exchange of their personal information (Tavani, 2000).

Westin's information privacy model has been the most influential operationalization of privacy for information systems designers, software engineers, and computer scientists (Steeves, 2009). Westin defines privacy as "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extend information about them is communicated to others," (Westin, 1967). Here, privacy equals control over information, and private information "belongs" to an individual, as a type of property right. Like other property, an individual can keep (conceal) or dispose of privacy (disclose or make public). The conceptualization of privacy as a property right has been quite influential within e-commerce (Solove, 2004).

The operationalization of privacy as control over information is influential because so much of our personal information is managed by computers. In addition, software designers can write fairly simple, straight forward code to support this operationalization, by designating privacy as a property or characteristic of a piece of data. The social context of that information, whether it relates to other people or organizations is not referenced when considering the privacy state of information.

Designs that treat information as discreet, independent data elements to be managed one by one place a significant burden on users. To simplify this privacy management burden, e-commerce systems divide information into two categories, 'personally identifiable information' (PII) and not personally identifiable (non-PII). Categories of PII include name, email address, and social security number. E-commerce systems designers only address issues of informed consent, adequate data security, and rules for data sharing for PII. Everything else collected from your online browsing, including sensitive items such as location (obtained from your IP address), what online articles you read, and what keywords you enter into a search engine are excluded from official considerations of privacy. By defining privacy to mean control over information, you only need to control your information if you can be identified. As long as you are treated anonymously, then you have no privacy concerns (Dwyer, 2009).

Many studies of privacy issues within pervasive computing systems begin with a definition of privacy grounded in the informational privacy model. For example, Minch (2004) defines privacy using the informational privacy model developed by Westin (1996). "The present research will build upon the definition of Westin and employ a working definition of privacy as essentially an information and communication-based construct—namely the manner and extent to which **persons can control how information about them is: (1) collected; (2) retained and/or maintained; (3) used; and (4) communicated, disclosed or shared** [emphasis added]. Location privacy may then be defined as privacy relating to location-specific information."

A more nuanced understanding of privacy is offered by Nissenbaum, who holds that information flow is governed by social norms that are highly dependent on context (Nissenbaum, 2004). So for example, information flow can differ within a health context, among friends, or in a public setting. Nissenbaum argues that privacy should be conceptualized as a relational and social property, meaning "privacy is a societal license that exempts a category of acts (including thoughts and emotions) from communal, public, and governmental scrutiny," (Solove, 2007). When software designers conceptualize privacy as an individual right, they lose sight of how to implement it within an online social environment.

The Complexities of Privacy Management

Privacy is a culturally and socially constructed property of relationships and groups (Petronio, 2002). People evolve privacy management strategies during years of socialization. The recognition of boundaries between public and private, for example when a child first comprehends she is naked, is marked as a cognitive milestone. While the social structures that people draw on for privacy management are quite powerful, they are nevertheless difficult to articulate and describe. They take the form of tacit knowledge (Giddens, 1984).

Because privacy management draws on tacit knowledge, it can at first glance seem to be a simple part of everyday life. In actuality it is quite complex, and for each person it is a moving target. What might be private between two people one day can become public the next. With the explosive growth of online social spaces, the complexities of privacy management have increased exponentially (Dwyer, 2008).

In any social setting, privacy management involves real time information boundary management decisions (Petronio, 2002). Studies in the area of cognitive psychology have found that privacy management decisions are constrained by cognitive limitations due to the need to make judgments quickly (Carey and Burkell, 2009). We can think of people applying the same "bounded rationality" (Simon, 1955) to privacy management decisions that are applied to other types of decisions.

Research from psychologists and social scientists show that people do not make privacy decisions based on a comprehensive consideration of the consequences of disclosure. Instead, they operate in what is called the "experiential" mode of thinking, which is "intuitive, fast, mostly automatic, and not very accessible to conscious awareness," (Carey and Burkell, 2009, p. 75). Privacy decisions are guided by heuristics. A heuristic is defined as "a speculative formulation serving as a guide in the investigation or solution of a problem," (Mifflin, 2004). The affect heuristic, the availability heuristic, and the representativeness heuristic as all examples of approaches people take in reaching decisions on issues with social consequences.

The affect heuristic refers to a feeling of goodness or badness related to perceptions of a situation. Feelings of goodness lower a person's concern about the risk of a certain action. The availability heuristic helps the decision maker by mapping the current situation to an experience in the past that can be readily called to mind. This means that categories of events that are

easier to remember are therefore considered more likely to occur. The representativeness heuristic aids the decision maker by relying on mental models or stereotypes, rather than a more complex analysis of a situation (Carey and Burkell, 2009).

The Incompatibility of Informational Privacy Models With Pervasive Computing

The operationalization of privacy as individual control over information is not practical or relevant to pervasive computing systems, because it implies that pervasive computing systems could never achieve any measure of privacy. Pervasive computing depends on the collection of millions of data points from complex sensor and computing structures. Giving individuals veto power over bits of sensor data is impractical and destructive to the goals of pervasive computing. In addition, this perspective concentrates on protecting certain types of data without establishing what people regard as private information (Adams and Sasse, 2001). What people consider to be an invasion of privacy is not the data itself but the implications they can provoke.

For these reasons, the information privacy model is inadequate and does not provide guidance for the design of pervasive computing systems. Therefore, a new operationalization of digital privacy must be developed, based on expositions of relationships between data, and what privacy revelations may occur through inference. Although it is true that pervasive systems use computers and collect information, the management of the privacy concerns they trigger are not adequately addressed by the informational privacy models described above. The next section introduces the notion of inference problems, describes their privacy components, and discusses their relevance to pervasive computing.

WHAT IS "THE INFERENCE PROBLEM?"

Inference is defined as "the process of arriving at some conclusion that, though it is not logically derivable from the assumed premises, possesses some degree of probability relative to the premises," (Mifflin, 2004). Inference is a fact of everyday life. An inference problem occurs when someone can combine clues and pieces of information to deduce confidential information. For example, calling someone out of class to the principal's office allows other students to infer that student is in trouble. The combination of the public action with contextual knowledge of the significance of a trip to the principal results in a breach of privacy for the student in question. Inference problems cause revelations of private information.

Preventing inference problems requires great sensitivity within a social setting. Adding technology compounds this difficult task. Technology based systems do not allow sufficient nuance, are not socially flexible, and do not allow for sufficient ambiguity (Ackerman, 2000). They also lack an important component of social grace, namely tact (Phillips, 2009). Tact suggests a willingness to be flexible, to look the other way and pretend not to notice a messy situation. Since controlling inference problems is not easily managed in everyday social situations, it is reasonable to expect this is quite difficult to tackle in pervasive computing environments.

The central issue is who is in control of inference problem strategies, the person or the system? As Adams and Sasse explain, "most invasions of privacy are not intentional but due to designers inability to anticipate how this data could be used, by whom, and how this might affect users," (2001). The worst possible outcome is for a system to sabotage a person's inference problem strategies, such as revealing their location to their spouse as they lie about their location in conversation on the cell phone. This could result in rejection of a pervasive computing system.

Inference Problem Examples

Most predictions 10 or 20 years out foresee an expansion of services offered through pervasive computing infrastructure. Already some of these ingredients are in place. The EZPass toll system is used by millions of drivers to pay tolls through an RFID tag that identifies their car and deducts the toll from their EZPass account. Drivers quickly signed up for the service once they realized their commuting times were reduced. While EZPass has been widely accepted, it has evoked privacy concerns. For example, can it be used to track whether someone is speeding? Or can it track someone's location when they would prefer it to be kept private?

The increasing communications infrastructure created by GPS cell phones and low cost broadband access will enable pervasive computing technologies to provide highly personalized services. The mass collection of transactional information on a global scale has consequences that are not well understood. While there are many doomsday scenarios, there are legitimate concerns that may degrade individual rights within a democracy.

Concerns regarding the future impact of these technologies resulted in the creation of the SWAMI (Safeguards in a World of Ambient Intelligence) project (Punie, Delaitre, Maghiros and Wright, 2005). The SWAMI Project is funded by the European Commission under the Information Society Programme (IST) of the Sixth Framework Programme (FP6). Its purpose is to

analyze the potential and challenges of a society with fully integrated ambient intelligence. Ambient intelligence is the European term for pervasive computing.

SWAMI is a multidisciplinary research project investigating the challenges with respect to privacy, security, identity, and trust. The SWAMI project's objective is to define potential threats and vulnerabilities, and to outline appropriate safeguards. The initial method for carrying out this objective has been the development and analysis of future scenarios, more precisely "dark scenarios." These scenarios aid the analysis and understanding of future risks and vulnerabilities related to ambient intelligence. Many of these scenarios deal with privacy and loss of control, and upon further analysis, are examples of inference problems.

Here is an example from the SWAMI report (Punie et al., 2005):

> We can imagine, in the future, everyone will have a 'friend-locater' function on, for example, his mobile phone. Now imagine the following situation which could occur to any fictive person:
>
> "In Munich, I experienced an awkward situation after I located a former colleague of mine using the 'friend-locater' function (LBS) of my PWC.33 I just wanted to say hi, but when I walked up to him, I was surprised to see that he had a good-looking, younger woman with him who obviously was not his wife. He blushed, mumbled a few words and disappeared in the crowd."

This scenario describes an inference problem. Running into a colleague in a different context without his wife leads to the inference of potential infidelity. While this scenario could occur without location aware technology, the suggestion is that it is more likely to occur with technology than without.

Social component of the inference problem

What pervasive computing calls ambient intelligence could easily be labeled spying in another setting. The thought of being spied on is an uncomfortable feeling. What makes it so uncomfortable is that we are unaware of what impression we are projecting. Erving Goffman describes impression management as "the way in which the individual in ordinary work situations presents himself and his activity to others, the ways in which he guides and controls the impression they form of him, and the kind of things he may or may not do while sustaining his performance before them," (Goffman, 1959).

People desire privacy not in order to control information, but to control the impressions it engenders, and disrupt potentially embarrassing inferences. What makes this especially difficult in an information technology application is that the audience is unknown and/or the feedback loop required for management of inference problems is missing.

The work of the sociologist Robert K. Merton (1949) describes the concept of a reference group, where an individual takes the values or standards of other individuals and groups as a comparative frame of reference. These should include normative as well as comparative frames of reference.

Merton described social visibility as the degree to which the norms and role performances within a group are readily open to observation by others, allowing for a normative impact. Merton noted in 1949 that full visibility of conduct and unrestrained enforcement of the letter of normative standards would convert society into a jungle. Merton writes (p. 399) "it is this central idea which is contained in the concept that **some limits on full visibility of behavior are functionally required** [emphasis added] for the effective operation of society." In a pervasive computing system, full visibility is the goal. Does that mean pervasive computing systems will fail by disrupting social relationships?

THE INFERENCE PROBLEM IN DATA MINING

The inference problem has been recognized as a security and privacy issue for databases since the 1970s (Denning, 1982; Denning and Denning, 1979). The earliest analysis of inference problems came for security issues within statistical databases (Farkas and Jajodia, 2002).

Statistical databases provide information about groups while protecting the confidentiality of individual entities. An example of such a database would be data about SAT scores and ethnicity. Although these databases were designed to prevent direct access to identifying information about entities, it became clear that confidential information was vulnerable to indirect queries.

A query is a request to return information meeting specific criteria, i.e., males between the ages of 15 and 18 who smoke. Researchers found that if queries were constructed to return a small query set (i.e., few records matched that query), then the use of other statistics could uncover confidential information.

Database designers introduced a number of inference control mechanisms, including query size and query overlap control, data swapping, and multidimensional transformation. These efforts were not adequate, as simple inference control mechanisms were easily subverted (Farkas and Jajodia, 2002).

Database designers then began to consider inference channels, which are vulnerabilities within the design of the database that allow higher level information to be inferred from lower level information. An example of an inference channel would be a dependency relationship between salary and rank. Someone could look up the records of an employee, but not be able to see their salary. However, if their rank was displayed, this could be used to discover salary.

Even if the relationship between employee and salary is restricted, other inference channels are possible. If the system maintains the data order in the database for all records, then separately queries on employee names and then the employee salaries enables the user to uncover confidential data.

Inferences can also result from known constraints. Assume that attribute 'A' is public but attribute 'B' is restricted with the constraint that A + B <= 20. In a query, because of this constraint, a query may only return certain values of A, thus creating an inference channel as to the value of B.

Control mechanisms for the database inference problem

Management of inference channels is a serious problem for high security databases, such as those maintained by the military. Research into eliminating these channels has gone in several directions. One is the idea of Lock Data Views (LDV). This technique uses classification constraints to prevent inference problems. In the LDV model, classification constraints are defined on sets of data according to the level of information that can be inferred from the data, not the level of the target of the query. Given a query, the result is upgraded to the appropriate level according to the classification constraints (Stachour and Thuraisingham, 1990).

Another approach to preventing the inference problem is to maintain a history of past data accesses. This additional information can assist in evaluating the security level of a query. This approach would prevent the inference channel described above, created by the back to back queries of employee name and employee salary. The inference succeeds because the user knows the data order is consistent for both sets. A check of query history, and an analysis of its inference potential, would prevent the execution of the second query (Farkas and Jajodia, 2002).

Automated inference channel detection

Research in the data mining and database security area is exploring the idea of automatic inference detection engines (Hinke, Delugach and Wolf, 1997). For pervasive computing systems, the development of automated methods of inference control holds the most promise for developing privacy protection. This is because the huge volume of data collected by a pervasive computing system can only be realistically managed through an automated system.

Data mining and data warehousing methods increase the complexity of eliminating inference channels. Current methods to control inference channels only work within one database. There is no method that can prevent inferences from one database being used to reveal confidential information in a second database (Farkas and Jajodia, 2002).

Y. Chen and W.W. Chu, researchers at UCLA, have an NSF funded project entitled "An Inference-based Approach to Data Access Violation Detection and Privacy Protection," (see http://cobase-www.cs.ucla.edu/projects/isp/). Their approach is the development of a semantic inference model (SIM). This is constructed using the data dependency, database schema and semantic relationship among data. The SIM is then mapped to a Bayesian network in order to evaluate the inference probability. So before any query is executed, the SIM is used to determine if any sensitive attributes can be inferred with a probability higher than their pre-specified thresholds. If a query fails this test, it will not be executed. The advantage of this method is that is a dynamic method for uncovering inference channels, rather than a static solution built into database design. Dynamic methods have the potential to be more responsive to new attempts to exploit inference channels (Chen and Chu, 2006).

THE INFERENCE PROBLEM AND IN PERVASIVE COMPUTING

Case studies of pervasive computing systems suggest that many privacy concerns are examples of inference problems. Consolvo et al. conducted an in-depth study of privacy issues with respect to the disclosure of location information (Consolvo et al., 2005). This study collected data using the experience sampling method (ESM). ESM uses monitoring devices to assess the situation of interest, by triggering events that can be measured as they occur while people are in natural

settings. Subjects in ESM studies carry with them a device, such as a smart phone, and are asked a few short questions from time to time in the course of their normal activities.

The study by Consolvo et al. examined the location disclosure privacy tradeoff. Disclosing one's location could be extremely valuable or extremely dangerous. The key question to answer is why, when, to whom are people willing to share their location information?

The researchers collected demographic information about their subject, as well as general privacy attitudes, and a description of their social network (i.e. their buddy list). Through ESM, participants received 10 randomly timed questionnaires every day for two weeks. Each ESM event was a hypothetical request to reveal location information to a person selected from the buddy list. The timing of the ESM requests was bounded by 9 am to 9 pm on weekdays and 10 am to 10 pm on weekends. Each questionnaire was made up of several questions and took about 2-3 minutes to finish.

In this study, the ESM data revealed that subjects made decisions about whether to disclose their location by determining who was making the request, why the requester wanted the participant's location, and what detail would be most useful to the requester. Another interesting finding is that participants typically disclosed what they perceived to be the most useful detail about their location (which is not necessarily the most detailed), or did not disclose their location at all.

These findings show that location disclosure is a social process. When given a request for location disclosure, who is asking and what they want are key components of the decision process. In the ESM scenarios, subjects could guess with a degree of certainty why their location was being requested. However, if the subject could not comfortably determine what the requestor's motivation was, they were more likely to reject the request.

It is interesting to see how subjects apply inference problem management strategies as they explain their decisions on location disclosure. For example, several subjects indicated they did not want to disclose their location if they were out running errands. "I wouldn't want people to ask me to pick stuff up." "If someone else knows you're there – you might have to do something for them, which will take more time." Another subject chose to hide her location in a particular shopping location: "When I go to BabyGap, I don't want my husband to know [as he thought she spent too much money there]." (Consolvo et al., 2005, p. 87)

Another ESM study by Iachello et. al. (2005) collected data from a location awareness system called Reno. Eight families with teenage children were the subject of this study. Although teenagers are famous for their fierce sense of privacy from intrusion by their parents, the results of this study found no evidence the teenagers used any deceptive methods when their parents requested their location information.

When the researchers probed this surprising result, they found a very practical reason behind it. Teenage children are often completely dependent on their parents for transportation. They would not have a social life without their parent's assistance. Using this device to enable their transportation trumped privacy concerns. So the decision process of determining who wants your location (i.e. your parent) and for what purpose (i.e. to drive you to another location) explains the location disclosure pattern much more clearly than an analysis of privacy attitudes inherent in a parent child relationship.

Another study conducted by Richard Beckwith (2003) sheds insight on how people's attitudes regarding privacy and technology are quite naïve, and how privacy concerns can arise as a result of inference problems. Beckwith conducted an ethnographic study of the first US eldercare facility to implement a sensor rich environment. In this facility, all residents and staff wore badges that indicated their identity and location to the system. Interviews with the staff, residents, and family members, revealed that the implications of such a system were not well understood.

Staff members did understand that their locations would be tracked, but they believed they could dis-arm this ability by leaving their badges behind if they went outside to take an extended cigarette break. However, other sensors would detect that someone leaving the building. Since many members of eldercare facilities suffer from Alzheimer's syndrome, these facilities must take extra care to prevent patients from wandering off. A staff member trying to sneak an extra break may find himself caught very quickly when sensors detect an unknown person leaving the building and trigger an emergency response to capture a wandering patient. This seems like a case of an inference problem.

Another clear example of an inference problem illustrated in this study arose from the use of load sensors on patient beds. A load sensor measures the weight of the person in bed. The weight of patients within these facilities is carefully monitored, because a rapid loss of weight is an indication of serious illness. Beckwith describes a scenario he describes as data fusion, where data from multiple sensors lead to a second order conclusion (i.e. an inference problem), (Beckwith, 2003).

> Data from various sensors can be merged to yield second order data, such as what time a resident entered his room, who entered with him, and what movements (and, to some extent, activities) occurred thereafter. For residents

> involved in campus romances, for example, load cell data could prove embarrassing. Data fusion is a general problem. It's difficult to imagine various uses for fused data when you don't even consider that a fusion could take place. (p. 43)

The situation Beckwith describes in the passage above has been labeled as the "unintended consequences" of pervasive computing. These unintended consequences share many of the characteristics of inference problems – revelations occur when unrelated pieces of a digital puzzle fall into place.

Beckwith research found that users expect the system to protect them from the impact of unintended consequences. Is this a reasonable expectation? It may not be reasonable, but we can think of it going forward as a non-functional requirement for pervasive computing systems. In software design, a non-functional requirement defines a system property or constraint. Non-functional requirements are critical to user acceptance, and systems that do not meet these requirements are rendered useless (Sommerville, 2001).

Pervasive computing will not be able to comply with the non-functional requirement of protection from inference problems until researchers provide a better understanding of the cognitive, social, and emotional processes people follow in their attempts to manage inference problems.

DISCOVERING INFERENCE CHANNELS WITHIN PERVASIVE COMPUTING SYSTEMS

The analysis of privacy within computing systems often begins with a definition of personal information. To what extent information is "personal" therefore determines its privacy component. Pervasive computing systems are based on the widespread distribution of smart devices and the continuous collection of environmental data. These data points and sources can represent billions of examples of potentially "personal" information. The computational complexity of determining the privacy level of this data (N data points by N sensors) is recognized to be an intractable computing problem, i.e. cannot be solved in practice (Aho, Hopcroft and Ullman, 1983).

Adams and Sasse (2001) argue that information sensitivity is determined by users' perception of the data being transmitted as well as how information is interpreted by the receiver. Their process oriented model of privacy is more relevant to the design of pervasive computing systems. It stands in contrast to the informational privacy approach by explaining how users make judgments with regard to information sensitivity. It appears that users will judge information sensitivity with a flexible scale rather than a private vs. not private distinction.

Adams and Sasse found that many perceptions of privacy invasions occur when users fail to appreciate that the data in question can reveal more than primary level information. Or in other words, when users realize that information can trigger an inference problem. Consider the case of sales staff that discovers the security cameras in their store are also used to evaluate their performance. When this occurs, and they discover information has implications they did not anticipate, they feel that their privacy has been invaded.

RESEARCH AGENDA FOR STUDYING THE INFERENCE PROBLEM WITHIN PERVASIVE COMPUTING

Does the use of pervasive computing systems increase the likelihood of inference problems? Can inference channels be identified and controlled within pervasive computing systems? Adams and Sasse argue that "it is the increased potential for ubiquitous technology to vary these factors [that lead to inference problems] without the user's full awareness of the repercussions, which increases the likelihood of unacceptable privacy risks," (Adams and Sasse, 2001).

Privacy problems within pervasive computing occur when systems are designed from an informational privacy perspective. Ackerman argues that functional systems based on this perspective can never be implemented in code (Ackerman, 2000).

> People have very nuanced behavior concerning how and with whom they wish to share information. People are concerned about whether to release this piece of information to that person at this time, and they have very complex understandings of people's views of themselves, the current situation, and the effects of disclosure. Yet, access control systems often have very simple models.

Following the informational privacy model, Ackerman goes on to describe the functional requirements of a system using this model to implement privacy [emphasis added].

> Even a cursory examination shows a wicked problem (in the computer science sense of "wicked", meaning an ill-formed, intractable problem). **A user would wish to control** the release of his private information on an ongoing basis to the various individuals and institutions within the environment. Roughly, this translates to **allowing the**

user to customize information transfer in two dimensions. Users must be able to handle essentially an infinite information space.

Ackerman analysis is grounded in the operationalization of informational privacy discussed above. Notice privacy is defined as elaborate control mechanisms for the user. Ackerman successfully argues that an informational privacy model cannot be implemented within the interactive computing systems in use today. Without another way to model privacy, we may as well stop building information systems. We need a new model of privacy that is consistent with being able to function in a digitally enhanced environment.

CONCLUSION

One obstacle to creating secure systems is the misconception of privacy as an individual control issue, which implies that if we can only figure out the right combination of buttons, check boxes, and privacy options then we can build secure systems. This conception permeates the computing and information assurance community. In 2003, the Computing Research Association (CRA) defined one of the grand challenges in computing as improving security and privacy within pervasive systems. The CRA described the follow goal: "For the dynamic, pervasive computing environments of the future, we will give computer end-users security they can understand and **privacy they can control** [emphasis added]," (CRA, 2003).

Progress in this area requires a paradigm shift to a new understanding of digital privacy. This does not mean privacy is not important, it means it must be defined differently (Adams and Sasse, 2001). The existing informational privacy models are inadequate. Therefore, a new definition of digital privacy must be created. Although this is clearly a difficult problem, the following are logical requirements for a new model of digital privacy:

- Digital privacy must be implemented through technology. Humans do not have the capacity to sift through millions of data points and analyze their privacy implications
- Digital privacy cannot be implemented as a sped up version of informational privacy models, where machines do the sifting of bits rather than people. The approach is too computationally complex
- Database security was improved by the identification of inference channels. Research should pursue how to identify inference channels within pervasive computing systems
- A better model will require a more detailed understanding of the cognitive and social processes people carry out as they navigate inference problems in their daily life. This understanding can only come from more intensive ethnographic and ESM research.

In conclusion, this paper has described informational privacy and discussed its application to pervasive computing systems. Definitions of digital privacy based on informational privacy models were shown to be impossible to implement within technology based systems.

A different understanding of privacy issues, the inference problem, was introduced from the database security literature. A discussion of privacy research within pervasive computing systems indicates many privacy issues are variations of inference problems.

The concept of inference channels has been used to improve security within databases. It is suggested that the definition of inference channels within pervasive computing systems be a target of future research. Inference channels can be identified through a combination of social and technical characteristics. It is suggested that research projects to identify inference channels within pervasive computing systems by a combination of ethnographic, ESM, scenario based, as well as grounded theory. Identifying inference channels will be an important first step in enabling the design of trustworthy pervasive computing systems.

REFERENCES

Ackerman, M. "The Intellectual Challenge of CSCW: The Gap between Social Requirements and Technical Feasibility," *Human-Computer Interaction* (15:2/3) 2000, pp 179-203.

Adams, A. and Sasse, M.A. "Privacy in multimedia communications : Protecting users, not just data," *Proceedings of the People and Computers XV - Interaction without frontiers. Joint Proceedings of HCI2001*, Lille, France, 2001.

Aho, A.V., Hopcroft, J.E. and Ullman, J.D. *Data Structures and Algorithms*, Addison-Wesley Longman Publishing Co., Boston, MA, 1983.

Beckwith, R. "Designing for Ubiquity: The Perception of Privacy," *Pervasive Computing* (2:2) 2003, pp 40-46.

Bush, V. "As We May Think," in: *The Atlantic Monthly*, 1945.

Carey, R. and Burkell, J. "A Heuristics Approach to Understanding Privacy-protecting Behaviors in Digital Social Environments," in: *Lessons From the Identity Trail*, I. Kerr, V. Steeves and C. Lucock (eds.), Oxford University Press, New York, 2009.

Chen, Y. and Chu, W.W. "Database Security Protection via Inference Detection," *Proceedings of the IEEE International Conference on Intelligence and Security Informatics*, San Diego, CA, 2006.

Consolvo, S., Smith, I.E., Matthews, T., LaMarca, A., Tabert, J. and Powledge, P. "Location disclosure to social relations: why, when, and what people want to share," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Portland, Oregon, USA, 2005, pp. 81-90.

CRA "Four Grand Challenges in Trustworthy Computing," Computing Research Association, http://www.cra.org/reports/trustworthy.computing.pdf, 2003, accessed on December 17, 2006.

Crang, M. and Graham, S. "Sentient Cities: Ambient intelligence and the politics of urban space," *Information, Communication and Society* (10:6) 2007, pp 789-817.

Denning, D.E. *Cryptography and data security*, Addison-Wesley, Reading, Mass., 1982.

Denning, D.E. and Denning, P.J. "The tracker: a threat to statistical database security," *ACM Trans. Database Syst.* (4:1) 1979, pp 76-96.

Dwyer, C. "Appropriation of Privacy Management Within Social Networking Sites," in: *Information Systems*, New Jersey Institute of Technology, Newark, NJ, 2008, p. 352.

Dwyer, C. "Behavioral Targeting: A Case Study of Consumer Tracking on Levis.com," in: *America's Conference on Information Systems*, San Francisco, CA, 2009.

Dwyer, C., Hiltz, S.R. and Jones, Q. "Discovering Boundaries for Mobile Awareness: An Analysis of Relevant Design Factors," *Proceedings of the Americas Conference on Information Systems*, Acapulco, Mexico, 2006.

Eng, P. "Will Big Brother Spy on E911 Cell Calls?," ABC News, http://abcnews.go.com/Technology/FutureTech/story?id=97704&page=1, 2005, accessed on 2/27/2006.

Farkas, C. and Jajodia, S. "The inference problem: a survey," *SIGKDD Explorer Newsletter* (4:2) 2002, pp 6-11.

Giddens, A. *The Constitution of Society*, University of California Press, Berkeley, CA, 1984.

Goffman, E. *The Presentation of Self in Everyday Life*, Doubleday and Co., Garden City, NY, 1959.

Hinke, T.H., Delugach, H.S. and Wolf, R.P. "Protecting Databases From Inference Attacks," *Computers and Security* (16:8) 1997, pp 687-708.

Hudson, S.E. and Smith, I. "Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems," *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 1996, pp. 248-257.

Iachello, G., Smith, I., Consolvo, S., Chen, M. and Abowd, G. "Developing Privacy Guidelines for Social Location Disclosure Applications and Services," *Proceedings of the Symposium On Usable Privacy and Security*, Pittsburgh, PA, 2005, pp. 65-76.

Lawler, J. and Molluzzo, J. "A Study of Data Mining and Information Ethics in Information Systems Curricula," *Proceedings of the Information Systems Educators Conference*, Columbus, Ohio, 2005.

Merton, R.K. *Social theory and social structure: Toward the codification of theory and research*, Free Press, New York, NY, 1949.

Mifflin, H. (ed.) *The American Heritage Dictionary of the English Language*. Houghton Mifflin Company, New York, 2004.

Minch, R.P. "Privacy issues in location-aware mobile devices," *Proceedings of the Hawaii International Conference on System Sciences* (37) 2004, pp 2019-2028.

Nissenbaum, H. "Privacy as contextual integrity," *Washington Law Review* (79:1) 2004, pp 101-158.

Petronio, S. *Boundaries of Privacy: Dialectics of Disclosure*, State University of New York Press, Albany, 2002.

Phillips, D.J. "Ubiquitous Computing, Spatiality, and the Construction of Identity: Directions for Policy Response," in: *Lessons From the Identity Trail*, I. Kerr, V. Steeves and C. Lucock (eds.), Oxford University Press, New York, 2009.

Punie, Y., Delaitre, S., Maghiros, I. and Wright, D. "Dark scenarios on ambient intelligence: Highlighting risks and vulnerabilities," http://swami.jrc.es, 2005, accessed on December 15, 2006.

Simon, H. "A Behavioral Model of Rational Choice," *The Quarterly Journal of Economics* (LXIL) 1955, pp 99-118.

Smith, I. "Social-Mobile Applications," *Computer* (38:4), April 2005, pp 84-85.

Solove, D.J. "'I've Got Nothing to Hide' and Other Misunderstandings of Privacy," *San Diego Law Review* (44) 2007, p 745.

Solove, D.J. *Understanding Privacy*, Harvard University Press, Boston, MA, 2008.

Sommerville, I. *Software Engineering*, (Sixth Edition ed.), Addison Wesley, New York, NY, 2001.

Sorensen, M.H. "Assitive ecologies - biomimetic design of ambient intelligence," *Proceedings of the IEEE/WIC International Conference on Intelligent Agent Technology*, Halifax, Canada, 2003, pp. 254-260.

Stachour, P.D. and Thuraisingham, B. "Design of LDV: A Multilevel Secure Relational Database Management," *IEEE Transactions on Knowledge and Data Engineering* (2:2) 1990, pp 190-209.

Steeves, V. "Reclaiming the Social Value of Privacy," in: *Lessons From the Identity Trail*, I. Kerr, V. Steeves and C. Lucock (eds.), Oxford University Press, New York, 2009.

Tavani, H.T. *Ethics and Technology: Ethical Issues in an Age of Information and Communication Technology*, John Wiley and Sons, Hoboken, NJ, 2000.

Waldrop, M.M. "No, This Man Invented The Internet," in: *Forbes*, 2000.

Weiser, M. and Brown, J.S. "The Coming Age of Calm Technology," in: *Beyond Calculation: The Next Fifty Years of Computing*, P.J. Denning and R.M. Metcalfe (eds.), Springer-Verlag, New York, 1998.

Weiser, M., Gold, R. and Brown, J.S. "The Origins of Ubiquitous Computing Research at PARC in the Late 1980s.," *IBM Systems Journal* (38:4) 1999, pp 693-696.

Westin, A.F. *Harris-Equifax Consumer Privacy Survey*, Equifax, Inc., Atlanta, GA, 1996.