Understanding and Minimizing Identity Exposure in Ubiquitous Computing Environments

Feng Zhu, *Member, IEEE,* Sandra Carpenter, Ajinkya Kulkarni, Chockalingam Chidambaram, and Shruti Pathak

***Abstract* — Various miniaturized computing devices that store our identities are emerging rapidly. They allow our identity information to be easily exposed and accessed via wireless networks. When identity information is associated with our personal and context information that is gathered by ubiquitous computing devices, personal privacy might be unprecedentedly sacrificed. People, however, have different privacy protection skills, awareness, and privacy preferences. Individuals can be uniquely identified on the basis of only a few identity elements used in combination. To the best of our knowledge, this is the first study to understand the following issues and their relations: a) what identity elements people think are important; b) what actions people claim to take to protect their identities and privacy; c) privacy concerns; d) how people expose their identities in ubiquitous computing environments; and e) how our rational identity exposure model can help to minimize identity exposure. We build a simulated ubiquitous computing shopping system, called *InfoSource*. It consists of two applications and our rational identity exposure model. We present our experiments and statistical analysis results. Our data show that exposure decisions depend on participants' attitudes about maintaining privacy, but they do not depend on participants' concerns and claimed actions related to identity exposure. Our RationalExposure model helped participants to minimize unnecessary exposures.**

***Index Terms*— game theory, identity management, privacy, ubiqitous computing**

I. Introduction

Personal privacy is a critical factor for the success of ubiquitous computing, yet it faces serious challenges. Various identities are expressed in digital forms, which enable people to easily expose identity information via wireless networks. Several states in the United States issue driver's license cards with embedded RFID tags or smartcards. Starting in 2006, the new version of U.S. passports have RFID chips integrated [1]. Credit cards, health insurance cards, student IDs, and other digital identity tokens with embedded computer chips are emerging rapidly. In ubiquitous computing environments, our identity information might be easily associated with our location and context information read from sensors. Without appropriate control over such exposure, ubiquitous computing environments could become a pervasive surveillance system [2].

We face a wide range of identity exposure threats. Malicious attackers try to acquire identity information for identity theft. Service providers tend to collect as much identity information as possible. Some service providers may collect as many as 100 identity elements from a user [3]. Some of them might use the information for price discrimination, while others might sell it for profit. A study shows that the combination of zip code, birth date, and gender can uniquely identify 87% individuals in the United States [4]. Unfortunately, identity exposure is usually left to individual users to make their decisions.

People are highly concerned about privacy in general [5, 6]. A survey on privacy in E-Commerce on the Internet suggests that people have different levels of willingness to expose information [5]. A further study via an online shopping website shows that people's privacy exposure behaviors do not match their privacy preferences [7]. These studies intended to influence the design of Platform for Privacy Preferences Project (P3P), which is a standard for websites and users' Internet browsers to communicate with each other about privacy preferences [8]. Few surveys and studies have been conducted for ubiquitous computing environments. One survey, however, reveals that people are not concerned about tracking and recording technologies [6]. Unfortunately, there is lack of detailed research on identity exposure. It is unclear whether people are aware of the importance and sensitivity of identity elements, and how their specific concerns about identity exposure and privacy are related to actions that they have taken. In this paper, we discuss our research results on these issues.

One might think that laws and regulations protect privacy, but they only provide protection on data usage, and they have not yet addressed privacy issues in ubiquitous computing environments [9]. Several research projects use policy-based approaches to protect privacy and prevent unnecessary identity exposure in ubiquitous computing environments [10-13]. However, complex policies may require users to have special skills to specify policies, and thus they suffer from usability issues [14]. We propose a game theoretic approach, called RationalExposure, to automatically suggest to users rational exposure decisions in ubiquitous computing environments

Manuscript received March 27, 2009. The research is supported in part by the UAHuntsville Research Mini-Grant 2009.

Feng Zhu, Ajinkya Kulkarni, Chockalingam Chidambaram, and Shruti Pathak are with the Department of Computer Science at The University of Alabama in Huntsville, Huntsville, AL 35899 (phone: 256-824-6255, fax: 256-824-6239, email: {fzhu, ask0004, spathak, cc0006}@cs.uah.edu)

Sandra Carpenter is with the Department of Psychology at The University of Alabama in Huntsville, Huntsville, AL 35899. (email: Sandra.Carpenter@uah.edu)

[15]. In this paper, we study the model's effectiveness and its usage as a factor in our experiments.

The main contributions of the paper are that we provide a thorough analysis that includes users' concerns, the actions they claim to have taken for privacy protection, the identity elements that they think are important, their exposure behavior in ubiquitous computing environments, and the effectiveness of our RationalExposure model. To the best of our knowledge, this is the first paper that provides detailed understanding of identity exposure.

We conducted two stages of research. In the first stage, we used an online survey to ask participants about their perception of identity exposure: the importance of identity elements, their privacy concerns, and actions that they had been taking to protect privacy. There were 229 participants who completed our survey. In the second stage, we conducted in-lab experiments and surveys. We implemented two applications that simulate in-store CD shopping and checkout processes. There were 100 participants who completed the second stage.

Our statistical analysis of the data shows that participants were highly concerned about privacy in general. According to what they said in the surveys, they took proper actions to protect their privacy. Most of them had clear and appropriate understanding of the importance of various identity elements. However, few of them protected their identity information in ubiquitous computing environments by themselves. Our RationalExposure model suggestions helped most participants to minimize their identity exposure. Furthermore, participants' exposure decisions depended on whether they thought identity elements were important. Nevertheless, their exposure behavior was independent of their privacy concerns and the actions that they claimed to have taken.

The rest of the paper is structured as follows. We discuss related work in Section II. In Section III, we describe our experiment design, method, our software, and participants. In Section IV, we illustrate detailed statistical analysis and our key findings. Last, in Section V, we outline our future work and conclude by describing our contributions.

II. Related Work

Ackerman, Cranor, and Regle surveyed 381 Internet users from the United States [5]. With the goal to inform the development of P3P, they designed survey questions to investigate three privacy issues: participants' responses to situations when their personal information was requested; the sensitivity of privacy practices; and their general privacy attitudes. Participants were asked whether they were comfortable providing twelve identity elements on the Internet for themselves and for children. The information ranged from social security number to favorite TV shows. Unlike their questions, we asked participants to rate the importance of keeping 26 identity elements private. We analyzed which identity elements participants considered to be similarly important.

Spiekermann, Grossklags, and Berendt extended the study of users' privacy exposure behavior on the Internet [7]. Participants used an anthropomorphic 3-D shopping robot to buy a coat or a camera. Their study revealed that participants were willing to reveal their personal information in spite of individual differences in privacy attitudes; participants' behavior showed sharp contrast to their claimed privacy attitudes. Similar to this research, we simulated the shopping experience and evaluated participants' behavior and their claims. In addition, we used our survey to understand whether users were aware of the importance of protecting their identity elements, and we tested our game theoretic approach to minimize identity exposure.

Lederer, Dey, and Landay's "five privacy design pitfalls" inspired our RationalExposure software design [16]. Specifically, we adopted their suggestions to help users understand the privacy information flow. Via the screen on a handheld device, the requested identity information, risk levels, the suggestion from the rational model, and the negotiation results with service providers were presented to participants.

Marx identified seven types of identity [17]: a person's legal name, address, unique symbols (alphabetic or numerical) to identify a person, pseudonyms that cannot be linked back to a person, a person's distinctive appearance or behavior patterns, social categorization (such as gender, ethnicity, religion, etc.), and possession of knowledge (such as password and secret codes). In our study of identity exposure, we included all types of the identity elements except the last type.

Acquisti and Grossklags pointed out that people often lack adequate information to protect their privacy [18]. Even with enough information, people often trade their privacy for short-term benefits. Acquisti and Grossklags also studied bounded rationality. Less than 10% of the participants who played a beauty contest game followed the perfectly rational strategy. Rather, participants used simplified mental models for privacy decision making. Unlike their approach that directly tested rationality of participants, we provided rational exposure suggestions. In addition, we strived to provide concise and informative exposure information. Then, we observed whether participants adopted our suggestions.

Policy-based privacy protection mechanisms have been adopted in multiple ubiquitous computing projects. The Privacy AWareness System (*pawS*) [12] was based on the P3P. Cranor and Reagle used a "buckets" approach [19]. Leonhardt and Magee adapted Lampson's access matrix and the Bell and LaPadula's (BLP) security labels for ubiquitous computing environments [11]. These approaches suffered from usability issues, and they were considered to be too complex for average users [14]. Hong and Landay designed a toolkit, Confab, for application developers to enforce policies, to send privacy notifications, and to manipulate private data [13]. Users controlled their privacy information in three interaction patterns. While policy-based approaches were too complex for users, Confab and another framework [2] did not provide sophisticated methods to help users make decisions. Our approach complemented their approaches and helped users to make rational decisions.

We propose an identity exposure model, called RationalExposure, for ubiquitous computing environments [15]. Using this model, a person's identity is stored in a hierarchical tree structure. The structure represents the identity elements from general to precise. An identity element is more general if a lager number of people have the same identity element. During the interactions between users and service providers, our model exposes the most general identity elements that service providers accept. We model the interactions as extensive games. (Extensive games are mathematical models, often used in economics, to model two parties' behavioral choices, in turn, as they make decisions and take actions.) By finding rational solutions in a game for both parties, an exposure is optimal for a user and a service provider in our case. In this paper, we evaluated users' acceptance of our approach.

With the advancement in location sensing and tracking technology, location privacy becomes a new challenge in ubiquitous computing environments. Instead of reporting a user's precise location, Snekkenes used lattices structure to express the user's location [10]. A location coordinate *(x, y, z)* is replaced with a volume from *(x, y, z)* to *(x', y', z')*. Another approach, Mix Zone, protects location information via anonymity [20]. A user is identified as an anonymous identity within a geographical area. As users enter and leave the area, new anonymous identities are generated and used. To better control the size of the anonymous set, Gruteser and Grunwald proposed *k*-anonymous location protection method [21], in which a user explicitly specifies the anonymous set size. Therefore, an anonymity server determines the region to expose.

Recently, researchers started to survey and conduct experiments to understand location privacy issues in ubiquitous computing environments. Nguyen, Kobsa, and Hayes asked 54 participants about their privacy concerns in general and their concerns about everyday tracking and recording via RFID, cameras, credit cards, and store VIP cards [6]. While participants were very concerned about privacy, they did not worry much about being tracked and recorded. Consolvo, Smith, Matthews et al. conducted three-stage experiments to learn why, when, and what participants want to share with respect to their location information to their friends, family members, and colleagues [22]. Their study involved both in-lab and *in situ* experiments. They found that participants' privacy attitudes were not a good predictor of their responses to location requests.

III. EXPERIMENTAL DESIGN

In the experiments, we wanted to understand five issues of identity exposure and their relations:

- What identity elements do people think are important? We wanted to gain an understanding of participants' attitudes about various identity elements. We looked for identity elements that participants considered as equally important to protect.
- Are people very concerned about identity exposure and privacy? In addition, we asked participants their concerns about information security and privacy in general.
- What actions have people been taking to protect their identities? Moreover, we wanted to know the actions that participants had been taking to protect their privacy and information security in general.
- What are people's behaviors when they need to expose their identity in ubiquitous computing environments? Is their behavior consistent with their concerns, their attitudes toward protecting the identity elements, and the protection actions that they claim to take?
- Can our RationalExposure model help people to make rational decisions?

To study the issues, we conducted two stages of experiments. In the first stage, we asked participants to complete an online survey. The survey focused on the first three issues. In second stage, we focused on the last two issues. Participants were asked to come to our lab. They used our software (called *InfoSource*), which provided a rich CD shopping experience and the RationalExposure model for the checkout process.

A. Participants

In the spring semester of 2009, we conducted the study. The participants were college students. Their age ranged from 17 to 33. About 90% of the participants were 23 or younger. In the first stage, 229 participants completed an online survey. In the second stage, 100 participants completed experiments and surveys in our lab.

B. Methodology

1) Stage one

Participants finished a survey online. Their sessions lasted for about 30 minutes. We asked three sets of questions: their music preferences and the extent of their online music purchasing experience, their demographic data, and their attitudes and concerns about privacy and security. We asked specific questions related to identity and privacy and more general security and privacy questions. For example:

- *Collection of privacy information about you*
- *Identity theft*
- *Transfer of your private information to other businesses*
- *Profiling and price discrimination*

We asked participants 18 questions related to the actions that they took to protect their identity, privacy and security. For instance:

- *Falsifying information about yourself on a website*
- *Deleting cookies from your computer*
- *Responding to unsolicited emails*
- *Carefully reading privacy policies on websites*

Last, participants rated 26 identity elements on how important it is to keep the identity elements private. The identity elements ranged from social security number, to zip code, to their favorite TV programs.

2) Stage two

Participants came to our lab for the second stage of the experiment. Each session took about 30 minutes, and up to 4 participants were in a session. Each participant was given a PDA, a pair of earphones, and a computer. In addition, they were given a brochure on how to use the *InfoSource* software. The PDAs were used to store their identity information. They used the computers to complete a survey afterwards.

We told participants that the experiment was to study their music preferences and simulate a future music shopping experience with handheld devices. First, a participant entered three pieces of his or her identity information on a PDA: a phone number, a credit card number, and driver's license information as shown in Figure 1 (a). Then, he or she supplied a password to encrypt all identity information. We asked them to pretend that the PDA was their personal cell phone, into which they entered information once and could use many times.

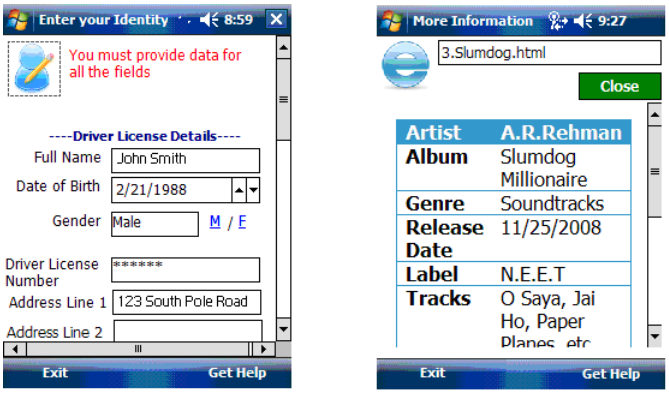

Figure 1. *InfoSource* screen shots. (a) A participant enters identity information. (b) More detailed information about a CD and a 30 seconds sample are accessible from the PDA.

Devices such as PDAs in our experiment may potentially serve as both one's cell phone and a digital wallet. It provides a much richer user interface (a touch screen, a microphone, and a speaker) for its owner to use and manage identity information than to manage digital identity cards such as one's driver's license and credit cards with RFID tags.

To protect participants' identity information, we recorded only whether they provided a certain piece of information. We did not record the actual information. When participants entered their identity information, they may have entered fake information, but we asked them to treat the identity information as though it was real during the experiment. No matter what they entered, the driver's license number and the first 12 digits of the credit card number were replaced with * during the process for purposes of participants' security. (The participants did not know that we replaced the information.) Each participant removed all his or her information on the PDA before the end of the experiment. In addition, the lab was configured in such as way that wireless communication was encrypted and all PDAs and computers were not connected to the Internet or any other computer that were not used for this study.

During the shopping simulation, participants were asked to look at CDs as if they were shopping in a store. They read additional information about CDs on the PDAs as shown in Figure 1 (b) and listened to sample songs. Imagine that in the near future products such as CDs may be tagged with RFID tags. Based on the tag IDs, additional information about the CDs may be acquired from a server in the store.

After participants selected CDs, they went through the checkout process. They used the PDAs to provide their credit card numbers and other information. The interactions between their PDAs and our server were over a wireless network. We asked following information:

- *Credit card information to pay for the CDs*
- *Phone number*
- *Driver's license information to verify the buyer's name on the credit card*
- *An offer to become a VIP member, requiring additional information*

Forty-two participants used the software without the RationalExposure model. They were asked for identity information and needed to make their own decisions as shown in Figure 2 (a). Fifty-five participants used *InfoSource* software with our RationalExposure model. The rational model suggested actions and let users to make final decisions as shown in Figure 2 (b). The RationalExposure model in the experiment used the extensive games with complete information when phone numbers and driver's license information were requested. In each session, all participants used the same version.

Participants needed to provide their credit card numbers to checkout or, alternatively, could quit the checkout process. Similarly, a driver's license was mandatory to finish the transaction. We let participants send the information via the PDAs. This is analogous to letting a store read a customer's information from his or her driver's license with an RFID tag. For the RationalExposure model, the PDA software negotiated with the server and found an outcome to provide a participant's name only.

Participants could finish the checkout process without giving their phone number or becoming a VIP member. Most identity elements requested for the VIP membership could be acquired from one's driver's license automatically. If participants wanted to become a VIP member but not give their real information, they could manually edit the fields.

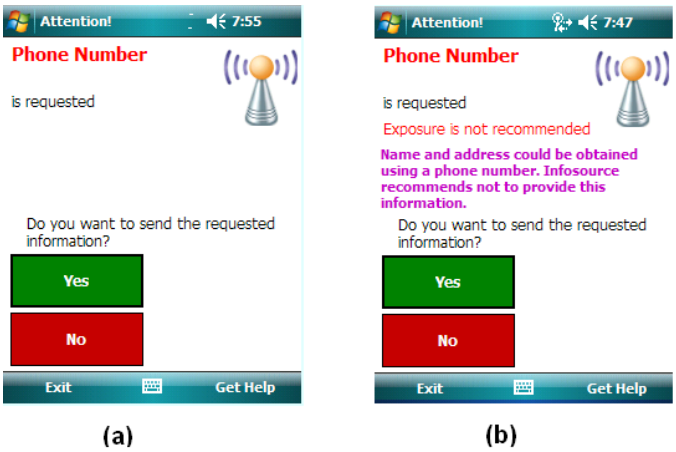

Figure 2. (a) Checkout process without the RationalExposure model. (b) The RationalExposure model provides users suggestions.

Last, participants filled out a survey. We asked them to evaluate their experience of using the *InfoSource* software, to rate the importance of eight identity elements, their identity exposure and privacy concerns (6 questions), and the actions that they took to protect their information privacy (10 questions). To keep the experiment time reasonable (within 30 minutes), we selected a subset of questions that we used in the first stage survey data. The questions are representative and the selection was based on the analysis of the first stage survey data. The detailed analysis is given in the next section.

IV. Results and Key Findings

The extensive survey data that we acquired in the first stage helped us to have a clearer understanding of what participants' attitudes, concerns, and actions that they claimed to have been taking with respect to security and privacy. Our findings in the data guided us in the second stage of the experiment.

A. *Importance of Keeping Identity Information Private*

We asked participants to rate 26 identity elements from "not at all important," "somewhat important," "substantially important," to "extremely important." Figure 3 shows the histograms of their ratings. Each bin in a histogram represents the percent of the users giving the rating at that level. We arranged the identity elements in order from the least important element to keep private to the most important element to protect. Overall, their ratings ranged from unanimous understanding of whether an element was important to totally different perceptions of the importance for some identity elements.

Most participants thought that their favorite TV program, favorite hobby, frequency of tobacco and alcohol usage, and frequency of Internet usage were not sensitive information to keep private. On the other end of the spectrum, most of them agreed that credit card numbers, the driver's license number, and the social security number were highly sensitive information to keep private. Note that more participants believed that credit card numbers were more important to protect than driver's license numbers, even though credit card numbers are usually easier to invalidate and change.

Their ratings for number of credit cards, monthly income, first and last name, IP address, phone number, and their location were widely spread out. That is, the standard deviations of the rating for these elements were large. We believe that participants were unsure or were not aware of the sensitivity of these elements. On the other hand, first and last name, location information, and phone number may be very sensitive. Nguyen, Kobsa, and Hayes' survey on location

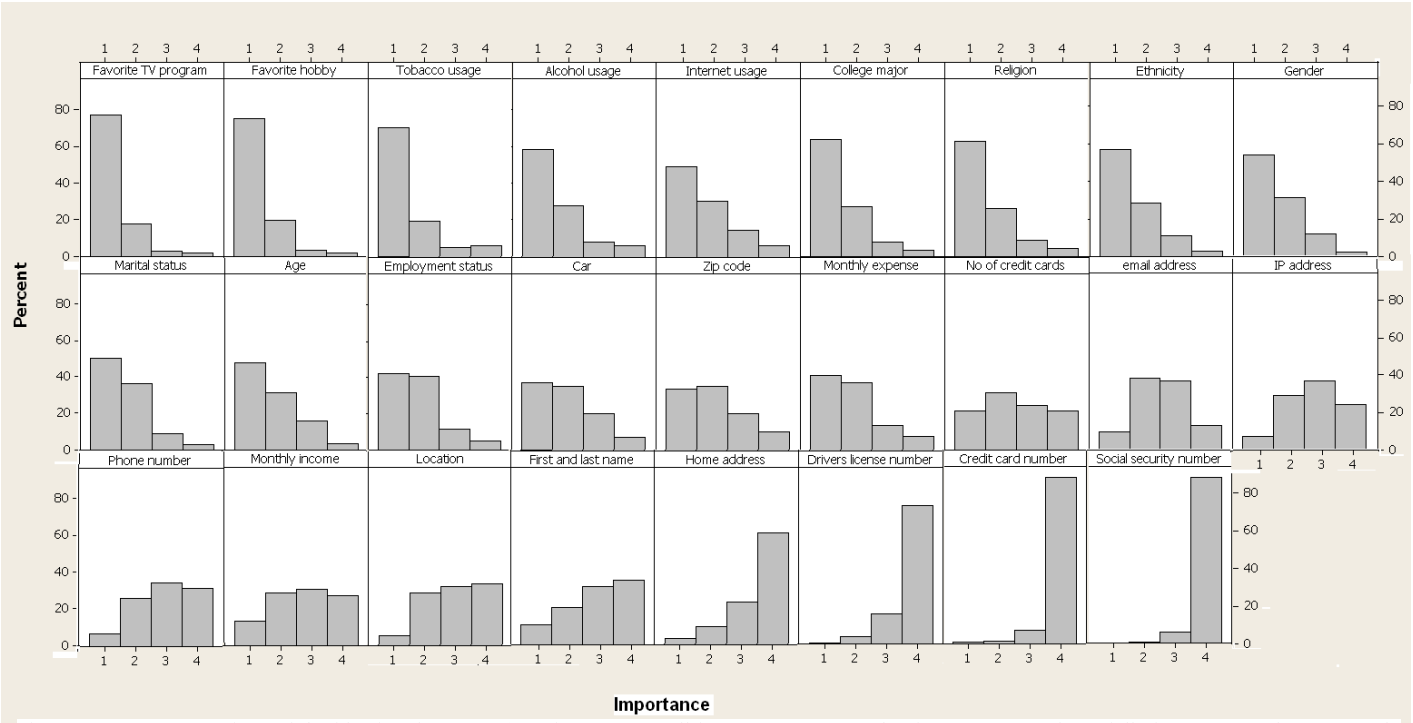

Figure 3. Importance ratings of the identity elements. (X axis -1. Not at all important, 2. Somewhat important, 3. Substantially important, and 4. Extremely important; Y axis – Percent of participants.)

sensing and tracking technology reached a similar conclusion, that people are uncertain about their location privacy [6].

We analyzed the similarity of the participants' importance ratings and clustered the identity elements to determine which identity elements they considered were similarly important to protect. We found three clusters, as shown in the dendrogram in Figure 4. We used average linkage method to measure the distance. Driver's license number, social security number, and credit card number are in one cluster. This was the group of identity elements that they thought the most important to protect. They rated first and last name, phone number, email address, and location information similarly. IP address was rated quite differently from other identity elements; it might be a unique digital identity, if one connects to Internet directly via an ISP. Or, at least, it is a unique identity within the first hop of the network. Participants may not be aware of the technical details and its true representation. The rest of the identity elements were in another cluster.

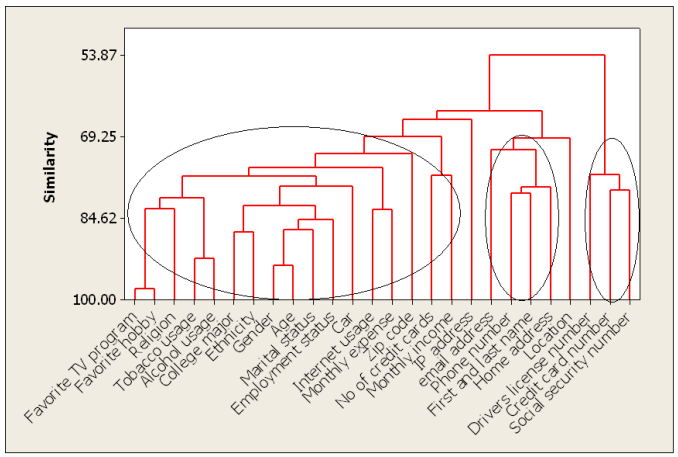

Figure 4. Dendrogram of the importance rating for the identity elements.

B. *Participants' Privacy Concerns*

Participants rated privacy concerns from not at all concerned, a little concerned, somewhat concerned, to very concerned. (Higher values indicate higher concerns.) Figure 5 shows their ratings in histograms.

They were concerned about their privacy information being collected in general. They were also concerned about their private information being transferred to others, being hacked, or being stolen. However, not many participants worried about law enforcement acquiring their private information. When we asked them more detailed questions, they showed various levels of concerns. They showed most concerns about their current location being known and infiltration of their computers, while they showed least concerns if people knew their hobbies, their clothes' and shoes' brands. For health condition, financial situation, purchases, visited websites, and emails being read, they had different opinions. Some showed great concerns, some did not worry at all, and some had a little concerns. The standard deviations of these ratings were very large.

When we asked them whether they worried about their private information being used against them in general (last chart in Figure 5), most of them were concerned. They expressed great concerns for identity theft and for being harassed. Surprisingly, many participants did not worry about price discrimination and profiling.

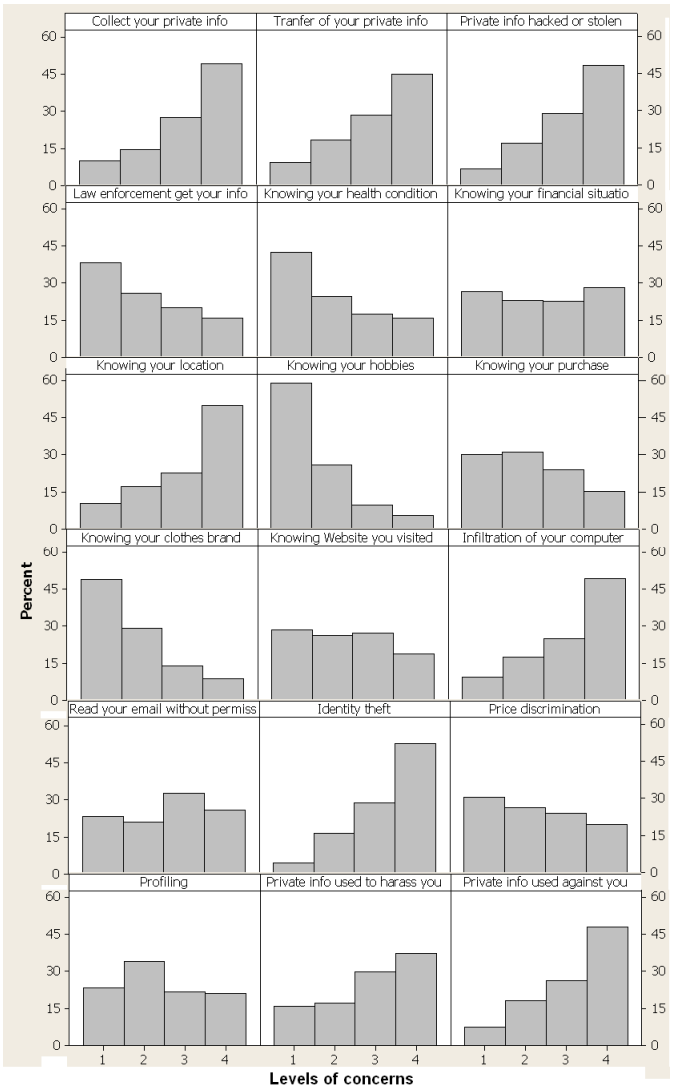

Figure 5. Participants' privacy concerns. (X axis -1. Not at all concerned, 2. A little concerned, 3. Somewhat concerned, and 4. Very concerned; Y axis – Percent of the participants.)

After comparing their ratings, we found interesting similarities among the privacy concerns as shown in the dendrogram in Figure 6. We separated the concerns into five groups. Participants considered price discrimination, profiling, their health condition, their financial situation, and their purchases very similarly. We believe that they considered that collection of their private information would be used against them negatively, because they rated seven privacy concerns in the left portion of the dendrogram most similarly. In addition, information about their current physical location did not cluster with other concerns.

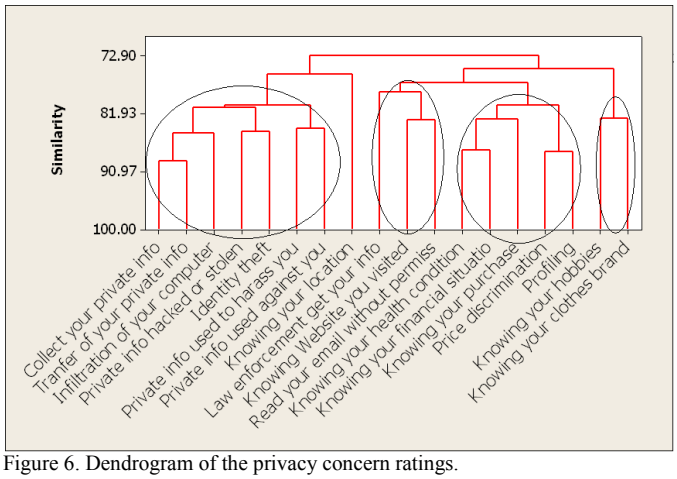

Figure 6. Dendrogram of the privacy concern ratings.

C. *Privacy Protection Actions that Participants Claimed to Have Taken*

Last in the online survey, participants gave ratings on a scale of 1 to 5 for the frequency of eighteen actions that they took to protect their information privacy and security. The higher number represents higher frequency.

They expressed prudence when they interacted with unfamiliar parties and did not actively provide information. As shown in the first row of Figure 7, most of them did not respond to telemarketing calls, unsolicited emails, and unknown instant messenger chat requests.

They claimed to actively protect their privacy information. About 70% of the participants never or almost never gave their information for better prices and services. More than 47% of them used more than one email address for privacy reason. Approximately 50% of them falsified their personal information on the Internet, at least sometimes, to protect their identity information and privacy. However, it seems that only 20% of the participants paid to not be listed in phone directories.

The participants were familiar with computers. They had taken actions to secure their computers and protected their digital identities. Most of them used antivirus software, firewall, and download security patches. About 70% of them also delete cookies at least sometimes. In their daily life, the majority participants protected their identity and financial information by shredding credit card receipts (70%) and checking credit card statements (79%), but over 67% of the participants did not order and check their credit reports.

About 18% of the participants claimed that they were frequently interested in finding out how their personal information was used by companies. Similarly, 18% of the participants frequently read privacy policies carefully. The dendrogram in Figure 8 suggests that those who cared about their information being used also carefully read privacy policies. About 40% of the participants claimed that they used encryption at least sometimes to protect their email messages. This percentage is higher than we expected.

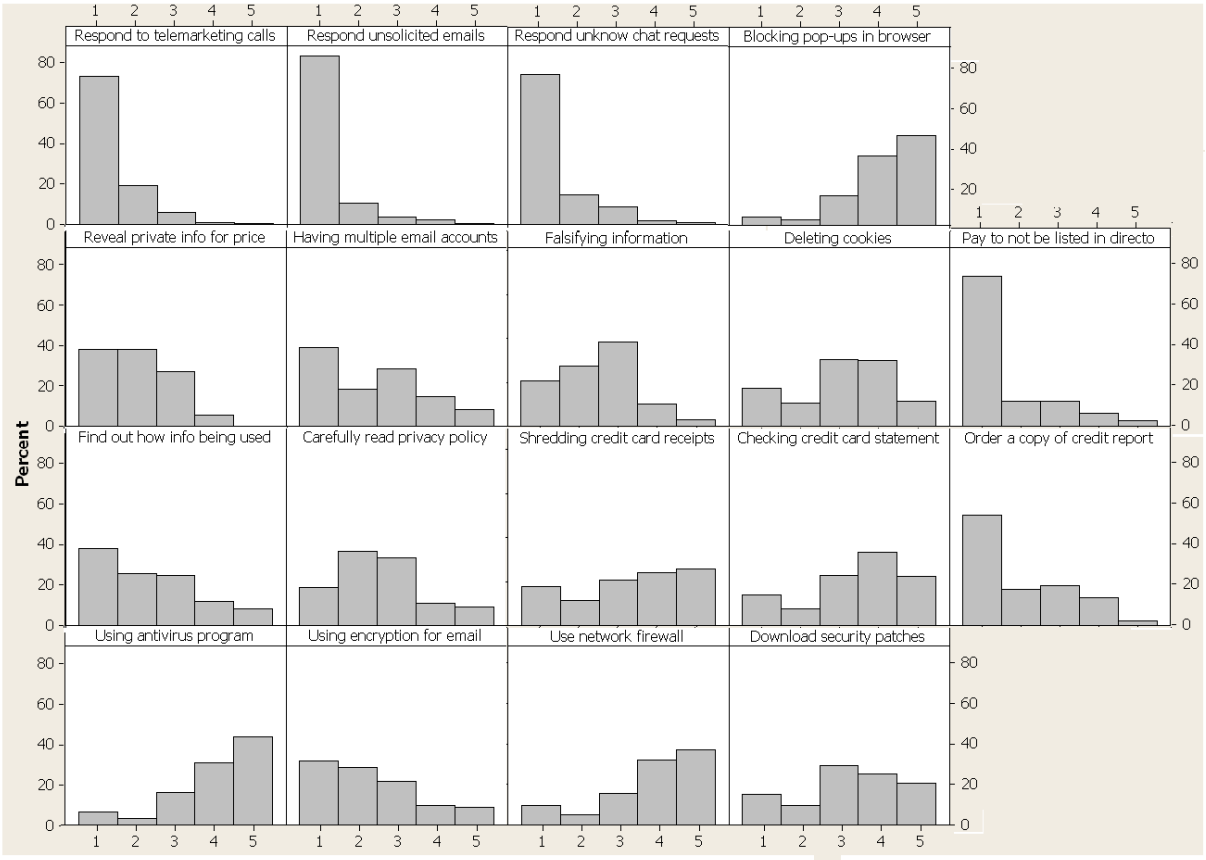

Figure 7. Actions that participants claimed taken. (X axis -1. Never, 2. Almost never, 3. Sometimes, 4. Frequently, 5. Very often; Y axis – Percent of the participants.)

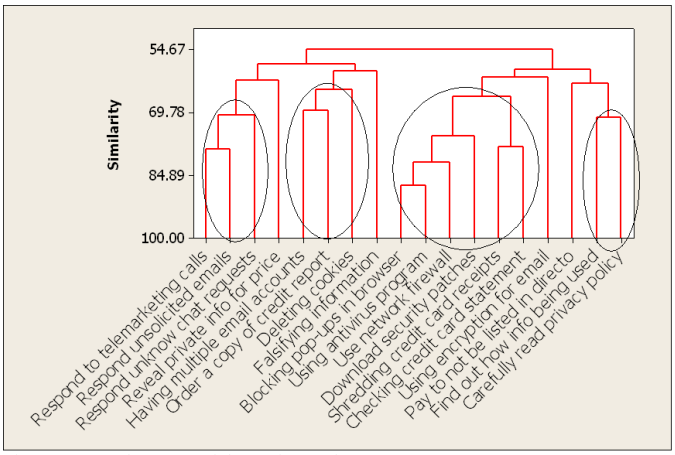

Figure 8. Dendrogram of the actions taken.

In summary, people indicated that they had the (a) strongest privacy attitudes about their address, driver's license, credit card number and social security number, (b) strongest concerns about their computers being infiltrated, identity theft, and having their private information collected, transferred, or stolen, and (c) most frequent security behaviors being blocking pop-ups, checking their credit card statements, and using anti-virus software. Thus, peoples' attitudes, concerns, and behaviors seem to be congruent with each other. The question remains, however, as to whether people actually take appropriate actions based on their attitudes and concerns. That is, do people truly act in ways that best protect their privacy and security? If so, why do so many people fall prey to identity theft and email scams? Anderson indicated that real attacks exploit psychology at least as much as technology [23]. The second portion of our research project focused on how people behave in a shopping situation in which some personal information (i.e., name, phone number, and driver's license information) is requested. This shopping experience was simulated using PDAs (as previously described in this paper). Some of the "shoppers" were given no help in maintaining privacy, whereas others used software that provided help – in the form of warnings not to expose private information – in maintaining privacy.

D. Participants' Identity Exposure Behavior

There were two participants who withdrew from the experiment. They felt that it was unsafe to give their credit card numbers in our experiments. (They were not aware that we replaced their card numbers while they were inputting). Another participant did not have a credit card. We acquired complete experimental data from 97 participants. There were 55 participants who used the software that had RationalExposure model, and 42 participants used software without it.

Among the 42 participants who needed to make their own decisions, 36 of them (86%) provided their phone numbers as shown in Table I. Note that the phone number was not mandatory to finish the transaction. For the 55 participants who used the RationalExposure model, 21 of them (38%) provided their phone numbers. This was very close to the percentage (35%) of the participants who rated that phone numbers as not important or somewhat important to protect. Therefore, our RationalExposure model suggestions seem to help participants make decisions that match their attitudes toward privacy and security.

Driver's license information was required to finish the shopping transaction. During the checkout process, participants were shown (with a message) that their driver's licenses were used to verify their names. Table I shows that 37 out of the 42 participants who did not use the RationalExposure model provided their full driver's license information by clicking a "Yes" button. The other 5 participants stopped the transactions. Four of them explained that it was not safe to give the digital driver's license. One said that he or she did not have a driver's license. Thus, 88% of the participants gave the full driver's license information, which included unique information such as their addresses and driver's license numbers, when only name information was required.

TABLE I. PARTICIPANTS' BEHAVIOR BY EXPERIMENT CONDITIONS.

Experiment condition	Without RationalExposure model	With RationalExposure model
Number of participants	42	55
Phone number	36 (86%) provided phone numbers	21 (38%) provided phone nmbers
Name verification	37 (88%) provided driver's license information	49 (89%) provided the name on the driver's license
VIP membership	19 (45%) applied VIP membership	24 (44%) applied VIP membership

For the participants in the RationalExposure condition, the software automatically negotiated with the checkout server. First, participants saw a message that their driver's licenses were requested. Then, the systems started negotiations with the server. Last, participants were prompted that only their names were required. Six participants felt uncomfortable providing their names, whereas 49 participants (89%) gave their names. Note that in the RationalExposure condition, the same percentage of participants concluded the checkout process, but by only providing name, rather than full driver's license information. Thus, the RationalExposure model suggestions protected participants, by encouraging them to expose minimum identity information. Interested readers might refer to our paper for rational exposure, negotiations, and best outcomes [15].

After the checkout process, all participants were notified by the checkout server that they might become VIP members. VIP membership would always give them the best price and hi-tech shopping carts. All participants made their own decision. (We did not provide rational suggestions.) Overall, 43 participants (44%) chose to provide their information for better prices or services. Both groups had the similar percentage of the participants selected to do so as shown in Table I. To obtain VIP status, all 43 of participants provided

their monthly income, email addresses, home address, and date of birth. Their PDA automatically read the information about home address and date of birth from their driver's licenses. Seven of those who wanted to acquire VIP cards clicked the edit button to modify the information before they sent it. (Four of the seven participants were in the RationalExposure condition and three were in the other condition.) Thus, only 7% (7/97) participants were not willing to provide the additional private information required, but still indicated a desire to acquire VIP cards. Recall that all participants made decisions without RationalExposure suggestions at this point. Thus, it seems that participants who had prior experience with the RationalExposure suggestions did not learn from those experiences to be more prudent with their private information.

E. The Relationship among Attitudes, Concerns, Claimed Actions, Actual Behavior, and our Rational Model

We used a logistic regression to analyze the relationship among participants' attitudes, their privacy concerns, their claimed actions to protect their privacy, and their actual behavior with and without rational suggestions, as shown in equation 1 below. Since the exposure of driver's license information was the different for the two groups (one group needed to expose the full driver's license information, whereas the other group only needed to expose names), in the discussion below we focus on their exposure of the phone numbers, which differed as a function of experimental condition.

$$\textit{Phone no. exposure} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \qquad (1)$$

where x_1 = *"Used RationalExposure model"*
x_2 = *"Thoughts"*
x_3 = *"Concerns"*
x_4 = *"Claimed actions"*

Their attitudes, concerns, and claimed actions were acquired in the survey right after they finished the simulated shopping portion of the study. First, we selected an item in each cluster that related to the identity exposure in the three dendrograms, respectively. We selected ratings of three identity elements (zip code, home address, and credit card number) in the three clusters as shown in Figure 4. Likewise, we selected two concerns (collect your private information and price discrimination) in the two clusters as shown in Figure 6. We believe that the other two clusters in the dendrogram in Figure 6 were not much related to the identity exposure. Similarly, four actions were chosen from the four clusters (carefully read privacy policy, reveal private information for better price and service, download security patch, and having multiple email accounts) as shown in Figure 8. Second, we calculated the averages of the ratings within each of the three clusters and rounded results to the nearest integers. These averages were used in the logistical regression analysis as predictors of exposure behavior.

Logistic Regression Table

Predictor	Coef	Z	P	Odds Ratio
Constant	7.07924	2.96	0.003	
Used RationalExposure model	-2.35530	-4.21	0.000	0.09
Thoughts	-1.02465	-2.61	0.009	0.36
Concerns	-0.0171817	-0.05	0.959	0.98
Claimed actions	-0.397237	-0.97	0.330	0.67

Figure 9. Logistic regression results for the model.

We subsequently tested a model that included only the significant predictors obtained in the first analysis – their attitudes and whether they used our rational model.

$$\textit{Phone no. exposure} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \qquad (2)$$

where x_1 = *"Used RationalExposure model"*
x_2 = *"Thoughts"*

The statistical results in Figure 9 suggest that the coefficients for the participants' attitudes and whether they used the RationalExposure model are not zero and the p-values are significant. Thus participants' attitudes toward privacy and the experimental manipulation predicted their disclosure behavior. However, participants' concerns and their claimed actions do not seem to be related to their actual behavior.

In Figure 10, the negative coefficient for participants' thoughts suggests that the more important the participants considered the identity elements the less likely they exposed their phone numbers. Similarly, the negative coefficient for the RationalExposure model indicates that if the model was provided it was less likely that participants would expose their phone number. The odds ratio further suggests that given the same rating for the importance of the identity elements, those who were provided the RationalExposure model were much less likely (0.09) to expose their phone numbers. In addition, the Pearson, Deviance, and Hosmer-Lemeshow goodness-of-Fit tests show that there is no evidence that the model does not fit the data adequately. In the measures of association section, the summary measures (Somer's D, Goodman-Kruskal Gamma, and Kendall's Tau-a) indicates that the model provides 30% to 74% of the predictive ability.

Logistic Regression Table

Predictor	Coef	Z	P	Odds Ratio	95% CI Lower
Constant	5.79306	3.53	0.000		
Used RationalExposure model	-2.43147	-4.38	0.000	0.09	0.03
Thoughts	-1.00426	-2.63	0.008	0.37	0.17

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	1.74945	5	0.883
Deviance	2.23248	5	0.816
Hosmer-Lemeshow	0.64930	4	0.957

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	1662	72.9	Somers' D	0.62
Discordant	250	11.0	Goodman-Kruskal Gamma	0.74
Ties	368	16.1	Kendall's Tau-a	0.30
Total	2280	100.0		

Figure 10. Logistic regression results for the third model.

V. Conclusion and Future Work

In this paper, we present our study of five aspects related to identity exposure: identity elements that people think are important to keep private, their identity and privacy concerns, the actions that people take to protect their identities and privacy, their identity exposure behavior, and the effectiveness of our RationalExposure model. We find that identity exposure behavior was related to (a) whether participants were given rational suggestions and (b) their attitudes about the importance of various identity elements. Our data do not provide evidence that their exposure behavior was related to their privacy concerns or to their claimed security actions.

An important finding of this research is that participants followed the suggestions provided by the RationalExposure model for disclosure of phone numbers and of full driver's license information, but did not learn to be more prudent in their disclosures when suggestions were not provided (as in the VIP privacy disclosures). We speculate that people either need more practice in negotiating which identity information to expose and/or need explicit instructions from a RationalExposure model every time they are engaging in exposure behaviors.

The exposure games that we used in the experiments were extensive games with complete information. That is, both the users and service providers know the preferences and payoff values of the other parties. Sometimes, users or services providers may not know the complete information about the other party's preferences and payoff values. We are designing games with incomplete information.

While we implement software for the new games, we are working on several aspects to improve our *InfoSource* software. First, we are revising our UI design for our RationalExposure model. During the experiments, we found that some participants kept clicking the "Yes" buttons without paying much attention to the text on the screen. We believe that the layout of the components on the screen and proper usage of the colors for visual notifications will acquire more attention from users. Second, we are integrating RFID tags in our experiments to further improve the simulated future shopping experience. Third, most of our participants liked the experiments. Based on their feedback, we are enabling more interactions between users and our *InfoSource* programs.

Acknowledgment

The authors are grateful to Dr. Wei Zhu for fruitful discussion and comments on the early versions of the paper.

References

[1] "United States to Require RFID Chips in Passports," in PC World, vol. October, 26, 2005.

[2] R. Campbell, J. Al-Muhtadi, P. Naldurg, G. Sampemane, and M. D. Mickunas, "Towards Security and Privacy for Pervasive Computing," presented at International Symposium on Software Security, Tokyo, Japan, 2002.

[3] L. Sweeney, "k-ANONYMITY: a Model for Protecting Privacy," International Journal on Uncertainty,Fuzziness and Knowledge-based Systems, vol. 10, pp. 557-570, 2002.

[4] L. Sweeney, "Uniqueness of Simple Demographics in the U.S. Population," Carnegie Mellon University, Laboratory for International Data Privacy, Pittsburgh 2000.

[5] M. S. Ackerman, L. F. Cranor, and J. Reagle, "Privacy in E-Commerce: Examining User Scenarios and Privacy Preference," presented at Proceedings of the 1st ACM conference on Electronic commerce, Denver, Colorado, 1999.

[6] D. H. Nguyen, A. Kobsa, and G. R. Hayes, "An Empirical Investigation of Concerns of Everyday Tracking and Recording Technologies," presented at Proceedings of the 10th international conference on Ubiquitous computing, Seoul, Korea, 2008.

[7] S. Spiekermann, J. Grossklags, and B. Berendt, "E-privacy in 2nd Generation E-Commerce: Privacy Preferences versus actual Behavior," presented at Proceedings of the 3rd ACM conference on Electronic Commerce, Tampa, Florida, 2001.

[8] L. Cranor, B. Dobbs, S. Egelman, G. Hogben, J. Humphrey, M. Langheinrich, M. Marchiori, M. Presler-Marshall, J. Reagle, M. Schunter, D. A. Stampley, and R. Wenning, "The Platform for Privacy Preferences 1.1 (P3P1.1) Specification," W3C, November 2006.

[9] M. Langheinrich, "Privacy by Design - Principles of Privacy-Aware Ubiquitous Systems," in Ubicomp 2001 Proceedings, volume 2201 of Lecture Notes in Computer Science, 2001, pp. 273–291.

[10] E. Snekkenes, "Concepts for Personal Location Privacy Policies," presented at 3rd ACM conference on Electronic Commerce, Tampa, Florida, USA, 2001.

[11] U. Leonhardt and J. Magee, "Security Considerations for a Distributed Location Service," Journal of Network and Systems Management, vol. 6, pp. 51-70, 1998.

[12] M. Langheinrich, "A Privacy Awareness System for Ubiquitous Computing Environments," presented at 4th international conference on Ubiquitous Computing, Göteborg, Sweden, 2002.

[13] J. Hong and J. Landay, "An Architecture for Privacy-Sensitive Ubiquitous Computing," presented at 2nd international conference on Mobile systems, applications, and services, Boston, MA, 2004.

[14] A. Soppera and T. Burbridge, "Maintaining privacy in pervasive computing — enabling acceptance of sensor-based services," BT Technology Journal, vol. 22, pp. 106-118, 2004.

[15] F. Zhu and W. Zhu, "RationalExposure: a Game Theoretic Approach to Optimize Identity Exposure in Pervasive Computing Environments," presented at IEEE Annual Conference on Pervasive Computing and Communications (Percom 2009), Galveston, TX, 2009.

[16] S. Lederer, A. Dey, and J. Landay, "Personal privacy through understanding and action: five pitfalls for designers," Personal and Ubiquitous Computing, vol. 8, pp. 440-454, 2004.

[17] G. Marx, "Identity and Anonymity: Some Conceptual Distinctions and Issues for Research," in Documenting Individual Identity: The Development of State Practices in the Modern World, J. Caplan and J. C. Torpey, Eds., 2001.

[18] A. Acquisti and J. Grossklags, "Privacy and Rationality in Individual Decision Marking," IEEE Security & Privacy, pp. 26-33, 2005.

[19] L. F. Cranor and J. Reagle, "Designing a Social Protocol: Lessons Learned from the Platform for Privacy Preferences Project," in Telephony, the Internet, and the Meda, J. K. MacKie-Mason and D. Waterman, Eds.: Mahwah: Lawrence Erlbaum Associates, 1998.

[20] A. Beresford and F. Stajano, "Location Privacy in Pervasive Computing," IEEE Pervasive Computing, vol. January-March, pp. 47-55, 2003.

[21] M. Gruteser and D. Grunwald, "Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking," presented at 1st international conference on Mobile systems, applications and services, New York, NY, 2003.

[22] S. Consolvo, I. E. Smith, T. Matthews, A. LaMarca, J. Tabert, and P. Powledge, "Location Disclosure to Social Relations: Why, When, & What People Want to Share," presented at Proceedings of the SIGCHI conference on Human factors in computing systems, 2005.

[23] A. Ross, Security Engineering: A Guide to Building Dependable Distributed Systems, 2nd Edition: Wiley, 2008.