Game Theoretic Approach to Location Sharing with Privacy in a Community-based Mobile Safety Application*

Hua Liu
Computer Science Department
University of Southern California
Los Angeles, CA
hual@usc.edu

Bhaskar Krishnamachari
Ming Hsieh Department of Electrical Engineering
University of Southern California
Los Angeles, CA
bkrishna@usc.edu

Murali Annavaram
Ming Hsieh Department of Electrical Engineering
University of Southern California
Los Angeles, CA
annavara@usc.edu

ABSTRACT

A new generation of community-based social networking mobile applications is emerging. In these applications, there is often a fundamental tension between users' desire for preserving the privacy of their own data and their need for fine-grained information about others. Our work is motivated by a community-based mobile application called Aegis, a personal safety enhancement service based on sharing location information with trusted nearby friends. We model the privacy-participation tradeoffs in this application using a game theoretic formulation. Users in this game are assumed to be self-interested. They prefer to obtain more fine-grained knowledge from others while limiting their own privacy leak (i.e. their own contributions to the game) as much as possible. We design a tit-for-tat mechanism to give user incentives to contribute to the application. We investigate the convergence of two best response dynamics to achieve a non-trivial Nash equilibrium for this game. Further, we propose an algorithm that yields a Pareto optimal Nash equilibrium. We show that this algorithm guarantees polynomial time convergence and can be executed in a distributed manner.

Categories and Subject Descriptors

I.6.8 [**Computing Methodologies**]: Simulation And ModelingModel Development; F.2.0 [**Theory of Computation**]: Analysis Of Algorithms And Problem Complexity—*General*

General Terms

Algorithms, Design, Human Factors, Theory

*We'd like to acknowledge partial support for this work from Nokia Inc and National Science Foundation, numbered NSF CNS-0831545. This work has been also supported in part by a CAREER grant from the National Science Foundation, numbered CNS-0347621.

MSWiM'08, October 27–31, 2008, Vancouver, BC, Canada.

1. INTRODUCTION

In today's society the penetration of hand-held mobile devices is substantially higher than any other compute or communication device. There are roughly 3.3 billion devices being used in the world as of November 2007 [1]. Till date most mobile applications are based on a simple client-server model, where a mobile device requests for a service from a single service provider. As mobile devices enter a new era with high speed connectivity and increasing compute capabilities, a new class of *community-based social networking mobile applications* are being showcased as the next revolution in mobile computing. In this class of applications each user in a social group contributes their knowledge about their surrounding environments and the collective knowledge can then be exploited by the community members for a personal or social benefit. One example of such an application has been developed to obtain real time information about traffic congestion on roads [3].

The primary difference between these new mobile social networking applications and prior mobile applications is that the information provided by the service provider is an aggregation of the data provided by multiple users. In their quest to increase relevancy of information to a specific user in a social networking scenario, however, mobile applications are beginning to aggressively collect information pertaining to a user. As the popularity of mobile social networks increases there is a growing realization that information collected about an individual user can compromise one's privacy and potentially security [9] [12] [17]. The information collected from a user include location and contact logs and hence when privacy is compromised it may lead to serious security concerns.

There is therefore a need for technological solutions for providing privacy in mobile social networking applications. In these applications there is a fundamental tension between a user's desire to protect privacy and their desire to take advantage of the community knowledge. On the one hand if everyone shares their information freely, the community as a whole will get a better experience; on the other hand, users prefer not to reveal too much personal information to protect their privacy.

We take a new perspective on this problem that is based on game theory [15]. Originally developed by economists to model strategic interactions between rational agents in market settings, game theory has been applied to many distributed network settings where users must interact while pursuing their self-interest. It is in many ways a natural fit for this domain of community-based mobile applications. Specifically, this framework will allow one to identify the Nash Equilibria for particular mechanisms - strategy profiles where users have no incentive to deviate unilaterally. It further enables the design of new mechanisms where the equilibria satisfied desired global performance for the community of users while allowing users to effect the privacy tradeoffs that they desire. Finally, it motivates the design of iterative algorithms that ensure that users can converge to the desired equilibrium in a distributed manner and maintain stable performance in the face of dynamics.

To ground our work, we describe a community-based mobile social networking application called *Aegis* that can be envisioned for personal safety enhancement particularly in high-crime urban areas. The basic idea of the Aegis system is that users share their locations with trusted others, and each can in turn view the locations of near-by individuals within their trusted circle, to enhance their sense of personal safety.

To our knowledge, this is the first work to quantify the privacy-service tradeoff central to these emerging mobile social network applications in a game theoretic setting. The paper is organized as follows. First, we describe related work in section 2. After a further description of Aegis in section 3, in section 4, we formulate a game played by mobile users that are interested in getting fine-grained information about each others locations while wanting to provide only coarse-grained information about themselves. We design a system that enforces a tit-for-tat information trade, with each mobile user getting location information about other nearby users at a granularity that is no higher than the information they are willing to share with others about themselves. The game that results turns out to have multiple Nash Equilibria, including trivial solutions where subsets of users choose not to share any information at all. In section 5.5 we show how the selfish best response can be calculated in a distributed manner by each user and discuss two simple iterative best-response algorithms. We consider the most simple simultaneous best response first in section 5.2 and show that it can sometimes fail to converge. However, in section 5.3 we consider a minor variation, sequential best response, that provides better performance. But going beyond these simple best-response heuristics, we show in section 5.4 that it is possible to solve for a Pareto-Optimal Nash Equilibrium of this game using a Pareto-improvement algorithm that converges in a polynomial number of steps. We present numerical evaluations comparing the performance of the sequential best-response with this Pareto-improvement algorithm in section 6 before our concluding comments in section 7.

2. RELATED WORK

Privacy and security concerns already pervade most of the internet application domain. The dramatic rise of social networking sites, such as Facebook and Bebo, has already ignited the debate on the effectiveness of legislation and privacy policies that are either ineffective or prone to human errors. A recent study by Pew Internet Research Project [2] shows that one third of US teenagers are subjected to cyberbullying due to privacy compromises. However, bringing the concept of mobility to social networking magnifies these concerns immensely as compromising location privacy may may lead to serious security concerns. There is a large body of research in the area of privacy preservation in traditional internet based social networking applications [4, 12, 31, 24, 10, 7, 35]. However, in traditional internet based application user's precise location is not revealed either to application provider or to other users, unless explicitly disclosed by the users themselves, such as the city name or zipcode where they are located. These social networking applications do not make use of precise location in providing location relevant information. Hence location privacy is primarily relevant in *mobile* social networking application and as such this paper focuses on this issue by using game theory to trade off privacy with user utility function.

Game theory [15] has been recently applied to wireless networks [27, 25]. In [38], Voorneveld *et al.* applied game theory to non cooperative games, also called anonymous crowding problem, where a user's value in visiting a location is inversely proportional to the number of people who have been already there which is the opposite goal of our safety application Aegis. Patwardhan *et al.* [33, 32] introduced the notion of packs to create a framework for providing privacy, security in mobile ad-hoc networks. A pack is dynamic set of individuals that collaborate for achieving a collective goal. Agah *et al.* [5] proposed game theory approach to security in sensor networks where each node in a network achieves better payoff when the node cooperates and its payoff is decreased when misbehavior is detected. To our knowledge, however, there has been no prior attempt to systematically apply game theory to the problem of privacy preservation in community-based mobile applications.

Several software solutions [16, 8, 21, 22, 13, 28] have also been proposed to protect privacy. Hong *et al.* [21, 22] proposed Confab, a toolkit for developing mobile application that allow developers and end users to support a broad spectrum of privacy needs. Desmet *et al.* [13] implemented a software architecture to allow the secure execution of third party applications on a Windows Mobile device. In [8] Capra *et al.* proposed a middleware architecture for providing privacy in mobile environments. Tang *et al.* [36] proposed a distributed method for storing personal information in mobile devices where personal information is split between mobile device and a trusted central server. Several experimental systems [23, 37, 18] also built location based services where the location of a mobile device is hidden from the service provider for protecting privacy. The primary focus of these researchers is the first generation mobile applications where a central authority can be trusted to provide accurate information. Hence the goal of a mobile device is to protect its privacy from this central authority. Our proposal focuses on mobile social networking applications where the data provided by multiple users with no central authority.

It is only recently that mobile social networking applications have come to the main stream of mobile computing [34, 19, 30, 14, 11, 6]. Reddy *et al.* [34] developed Campaignr framework for creating urban participatory sensing using mobile devices. Hoh *et al.* [19] explored temporal and spatial distortion of location data to protect privacy. Annavaram *et al.* [6] developed HangOut a social networking application that uses a combination of anonymous data aggregation and encryption to show where people with similar interests are likely to congregate. In Hangout the mobile device decides on the granularity of its location update based on how many other users are already seen by the server in a given area. Furthermore the device identification and location update packets are encrypted differently such that the data link provider can only identify the device but not the content and the application service provider can only identify the content but not the device. Hoh *et al.* [20] proposed a social network based traffic sensing application using the concept of spatial sampling with virtual trip lines. Using a combination of spatial, temporal and speed distortions they showed how real time traffic can be estimated without loss of privacy. In these previous studies the focus is primarily on absolute user privacy rather than trading privacy with utility. This paper specifically focuses on relative privacy where multiple users can trade their privacy with utility value derived from a community application.

3. AEGIS: A COMMUNITY MOBILE APPLICATION FOR PERSONAL SAFETY

Crime is a serious social malice that has received significant attention in social studies. In a recent survey [29] over 80% of people believe that the notion of *perceived* crime is an important factor in determining where people will stay and what places they will visit. Studies like this have also showed that a person's perceived notion of safety increases when they carry a mobile phone since the device provides a way for instant communication with their friends and family, and if needed with law enforcement agencies. While instant communication is an obvious benefit of mobile devices, we believe that more comprehensive approaches to personal safety can be achieved by exploiting the rich set of sensors on mobile devices. In order to explore these rich dimensions to personal safety, we envision the development of new personal safety applications on mobile devices that are based on the notion that a person's sense of security can be closely correlated to how many people that person can trust within his surroundings. While personal trust is subjective, it is generally believed that if there are more people around a user that he/she has some trusted relationship with (either directly or indirectly through a social network) then that user's sense of security is enhanced. The Aegis system is based on this idea, displaying the locations of near-by trusted individuals to a user to enhance their sense of safety.

While practical full-scale implementations of the Aegis application are likely to be quite sophisticated (for instance, taking into account a rich combination of information from call-logs to determine each individual's circle of trust), we treat a bare-bones version of this application in this study.

In this simplified version of Aegis, we assume that all users belong in each others' circle of trust. Each device registered with the system provides the system with its location. All users with mobile devices within some neighborhood (defined by some physical distance range) of this device can potentially be notified of its location by the system.

The fundamental tradeoff that we explore in this paper pertains to the granularity of location provided by and to the users. On the one hand they all desire to know the locations of the other users with high accuracy; on the other hand, they each prefer not to reveal their own location with accuracy. We try to resolve this conflict by treating each user as an self-interested entity playing a game.

An important part of defining such a game is modeling the utilities for each user. Modeling safety perception by humans realistically is a very challenging task (perhaps best left to sociologists). Our approach in this work is to pick a simple, tractable, almost-linear utility model for each user that has some intuitive features. The utility model has two components: the gain from knowledge of others' locations, and the loss from the revelation of ones' own location. In the model we adopt, location accuracy is treated as a tunable term — it may be varied in practice by adding zero mean noise to the true location with different variance, or by selecting different zoom levels of locations). The gain term captures the essence that each user is generally more happy when more other users provide location information, that each user is generally more happy when each other user provides more accurate location information, but that there is a point of saturation beyond which the user can be made no happier. The loss term is treated to be linear in the accuracy of the information provided by the user.

4. PROBLEM DEFINITION

Let $\mathcal{N}$ denote the set containing all users in this application. After describing a neighborhood range R, user i can see a set of nearby users $\mathcal{N}(i)$ within distance R on a map on his/her mobile device. Each user is able to specify the granularity with which their location should be made available to others. In sparser areas, for reasons of safety, each user is more interested in knowing the exact location of others than in denser areas. Let $a_i \in [a_{min}, a_{max}]$ be a real value that denotes the granularity of location provided by user i, where higher value of a_i corresponds to more accurate location information. Let us consider a particular concrete model to quantify the utility $U(a_i, a_{-i})$ (a_{-i} denotes the strategy vector for all users except user i) for user i:

$$U(a_i, a_{-i}) = min(K, \sum_{j \in \mathcal{N}(i)} a_j) - ca_i \tag{1}$$

Where K is a pre-defined positive real number to indicate an upper bound on benefits for node i and c is a positive penalty factor.

With this model, the user's benefit function is additive in the information accuracy of its neighbors, but saturates at a certain point. Notice that this utility function doesn't give incentive for nodes to share their location information. A user's benefit comes from the actions of others but the cost depends only on the user's own action. It can be shown that the only Nash Equilibrium point in the game by using this utility function is the trivial outcome: $a_i = a_{min}, \forall i$; i.e., each user always provides minimum accuracy. While this is ideal for each user in terms of maximizing privacy, it results in arbitrarily poor service.

From a game-theoretic point of view, what is missing is a direct incentive for the users to provide high accuracy data to others. A simple tit-for-tat mechanism that can implement such an incentive is to provide information to a user with download granularity commensurate with the user's upload granularity. An authorized system server through which the users interact can be involved as an information filter to implement this mechanism. The perceived accuracy of a neighbor j for i will then be given by $a_j = min(a_j, a_i)$, so that the utility now becomes:

$$U(a_i, a_{-i}) = min(K, \sum_{j \in \mathcal{N}(i)} min(a_i, a_j)) - c \cdot a_i \tag{2}$$

For ease of exposition, we assume the range of a_i is $a_i \in [0, K]$ from now on (this is equivalent to assuming that $a_{max} \geq K$. However, as we will point out, all the results can be extended in a straight forward fashion to the case when $a_{max} < K$). Further, to restrict the utility function to be non-decreasing in a_i before the saturated point, we also assume that the penalty factor c is less than 1 ($0 < c < 1$).[1] We use the word "nodes", "users", "players" interchangeably in the following sections.

The utility function in (2) provides some desired properties for the application. There exists at least one Nash equilibrium for any network topology. The trivial Nash equilibrium is $a_i = a_{min} = 0$. Consider the special case when all $\mathcal{N}$ nodes are within the same vicinity; there exist infinitely many Nash equilibria. All solutions of the form $a_i = \alpha$, $\forall i$ (where $\alpha \in [a_{min}, min(a_{max}, \frac{K}{\mathcal{N}-1})]$) are Nash equilibria. However, there is a unique Pareto-optimal Nash equilibrium given by the solution $a_i = min(a_{max}, \frac{K}{\mathcal{N}-1})$, $\forall i$, which is the best possible solution from a global (social welfare) point of view with respect to the utility. This solution is also intuitively appealing: since the benefit saturates beyond some point, it is best to provide more privacy (less accurate coordinates) when there are more neighbors.

5. ALGORITHMS

In this section, we give three different algorithms to find a non-trivial Nash equilibrium in the game we defined in the previous section, named as synchronized best response dynamic (SYN-BR), sequential best response dynamic (SEQ-BR) and Pareto improvement path (PI) respectively.

5.1 Calculate Best Response

Before we describe the algorithms, we first describe the solution for calculating node i's best response when given all his neighbors' strategies in Algorithm 1. Node i's neighbor set is denoted as $\mathcal{N}(i)$, and $|\mathcal{N}(i)|$ denotes the cardinality of set $\mathcal{N}(i)$. $BR(a_i, a_{-i})$ denotes node i's best response when given the other nodes' strategies in vector a_{-i}.

We consider the following three cases to calculate the best response for node i:

- When $\frac{K}{|\mathcal{N}(i)|} \leq min_{j\in\mathcal{N}(i)} a_j$ (i.e., node i's neighbors have relative high accuracy than expected), setting $a_i = \frac{K}{|\mathcal{N}(i)|}$ will maximize the utility function.

- When $\sum_{j\in\mathcal{N}(i)} a_j \leq K$ (i.e., the summation of node i's neighbors' granularity cannot reach K), the best response of node i is to match the maximum of the accuracy of its neighbors.

- In other cases rather than the two cases discussed above, node i's best response is a value between two of his neighbors' accuracy value. If we sort all the node i's neighbors' accuracy value, node i's best response is between two consecutive accuracy values a_{j_k} and $a_{j_{k+1}}$ and $\sum_{j\in\mathcal{N}(i)} min(a_i, a_j) = K$. To calculate node i's best response in this case, we use the following fact: when $c \leq 1$ (as defined in previous section) and $\min_{j\in\mathcal{N}(i)} a_j \leq a_i \leq \max_{j\in\mathcal{N}(i)} a_i$: if $\sum_{j\in\mathcal{N}(i)} \min(a_j, a_i) \leq K$, the utility function is non-decreasing with a_i; on the other hand, if $\sum_{j\in\mathcal{N}(i)} \min(a_j, a_i) \geq K$, the utility function is decreasing with a_i.

[1]When $c \geq 1$, there is not enough incentive for the users with degree less than c to participate in the game. We leave the discussion of the case where $c \geq 1$ to future work.

The algorithm to calculate the best response for a node is presented in Algorithm 1. The best responses can be calculated in a distributed manner. The complexity of a node i to compute its best response is $O(n \log n)$ (where $n = |\mathcal{N}(i)|$ is the number of neighbors of node i) when choosing proper sorting algorithm.

Algorithm 1 Calculate Best Response for Player i : $BR(a_i, a_{-i})$

```
if K/|N(i)| ≤ min_{j∈N(i)} a_j then
  return BR(a_i, a_{-i}) = K/|N(i)|;
else
  if Σ_{j∈N(i)} a_j ≤ K then
    return BR(a_i, a_{-i}) = max_{j∈N(i)} a_j;
  else
    sort a_j (∀j ∈ N(i)) in ascending order, denote the
    order as a_{j_1}, a_{j_2}, ..., a_{j_{|N(i)|}};
    find BR(a_i, a_{-i}) such that a_{j_k} ≤ BR(a_i, a_{-i}) ≤
    a_{j_{k+1}} and Σ_{q=1}^{k} a_{j_q} + (|N(i)| − q)BR(a_i, a_{-i}) = K;
    return BR(a_i, a_{-i});
  end if
end if
```

5.2 Synchronized Best Response Dynamic

SYN-BR is the easiest learning dynamic in game theory. This algorithm assumes that all players take action simultaneously and periodically. In each iteration, all players give their best responses to the other players' actions in last iteration. Algorithm 2 illustrates the steps to do SYN-BR. Note that SYN-BR does not guarantee convergence. To avoid infinite loops, we set a large number $maxIter$ as an upper bound for the iterations. However, if the algorithm converges, it will converge to one arbitrary Nash equilibrium. We also notice that $a_i = 0$ $\forall i$ is a trivial Nash equilibrium. To avoid this trivial Nash equilibrium, we set the initial state as all nodes at value K.[2] Let a_i^t denote player i's strategy at iteration t.

Algorithm 2 Synchronized Best Response Dynamic

```
Initialization: a_i^0 = K; t = 1;
while Not-Converged AND (t < maxIter) do
  for Every Node i do
    a_{-i}^t = a_{-i}^{t-1};
    Calculate BR(a_i^t, a_{-i}^t);
  end for
  t++;
end while
```

Here we give an example to show that SYN-BR algorithm does not converge in some cases. In Figure 1, nodes C,D and E converge after the second iteration and keep stable from then on, while nodes A and B will never converge. The values of nodes A and B keep oscillating forever.

[2]The initial state for player i in SYN-BR can be randomized in the range. Different initial states might lead to different Nash equilibrium at the end. Different initial states might also affect the convergence time.

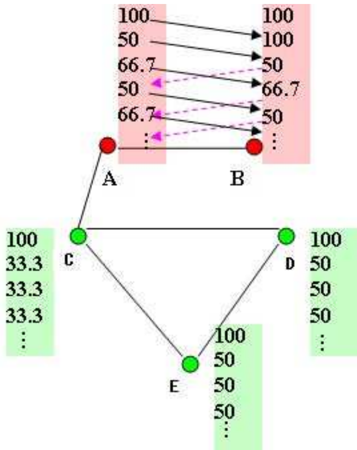

Figure 1: An example to show SYN-BR algorithm does not always converge

5.3 Sequential Best Response Dynamic

The example in Figure 1 illustrates that the SYN-BR algorithm cannot guarantee convergence. To eliminate the oscillation of the strategies among players, we propose SEQ-BR. In SEQ-BR, players update their strategies sequentially according to some pre-agreed order. The order for updating is called the sequential index. When a player i calculates his best response, he considers two sets of his opponents' strategies. For those players who have lower sequential index (i.e., the players who have already updated their strategies before player i), player i takes their strategies in "current" iteration into consideration. For the remaining nodes, player i uses the information from the last iteration. A formal description of this algorithm is in Algorithm 3.

Algorithm 3 Sequential Best Response Dynamic

Initialization: $a_i^0 = K$; $t = 1$;
while Not-Converged AND $(t < maxIter)$ **do**
  for i from 1 to N **do**
    for $j \in \mathcal{N}(i)$ **do**
      if j < i **then**
        $a_j^t = a_j^t$;
      else
        $a_j^t = a_j^{t-1}$;
      end if
    end for
    Calculate $BR(a_i^t, a_{-i}^t)$;
  end for
  $t++$;
end while

We have empirically observed that the sequentially best response dynamic converges in all simulations. However, the Nash equilibrium it converges to is an arbitrary Nash equilibrium based on the initial state. In the following section, we propose a Pareto Improvement algorithm which guarantees convergence and results in a Pareto optimal solution.

5.4 Move the Nash Equilibrium

Before we propose our algorithm, we first introduce the basic concepts of Pareto improvement and Pareto optimality.

Given a set of alternative allocations, a movement from one allocation to another that can make at least one individual better off without making any other individual worse off is called a Pareto improvement. Specifically, in this game, a Pareto improvement strategy vector $(\tilde{a}_1, \tilde{a}_2, ..., \tilde{a}_{|\mathcal{N}|})$ for strategy vector $(a_1, a_2, ..., a_{|\mathcal{N}|})$ satisfies the following two conditions:

- $\exists i \in \mathcal{N}$ such that $U(\tilde{a}_i, \tilde{a}_{-i}) > U(a_i, a_{-i})$
- $\forall i \in \mathcal{N}$, $U(\tilde{a}_i, \tilde{a}_{-i}) \geq U(a_i, a_{-i})$

When no further Pareto improvement can be made for a joint strategy vector, the strategy vector is called Pareto optimal or Pareto efficient. For our privacy game, we already pointed out that there might exist multiple Nash equilibria. In this case, finding a Pareto efficient Nash equilibrium becomes an interesting problem. We propose a polynomial-time algorithm to find a Pareto optimal Nash equilibrium starting from the all-zero trivial Nash Equilibrium in the following.

We need the following lemmas to hold for the correctness of the Pareto improvement algorithm described in Algorithm 4.

Lemma 1: Given a Nash equilibrium $(a_1, a_2, ..., a_{|\mathcal{N}|})$, no Pareto Improvement can improve node i's utility in this Nash equilibrium if $\sum_{j \in \mathcal{N}(i)} min(a_j, a_i) = K$. Such a node i is called a **saturated** node. Let $\mathcal{S}$ denote the set containing all saturated nodes.[3]

Proof (by contradiction): Suppose that exists a Pareto Improvement PI such that after the PI process, $U(a_i^{PI}, a_{-i}^{PI}) > U(a_i, a_{-i})$ and $\forall j \in \mathcal{N}(i)$, $U(a_j^{PI}, a_{-j}^{PI}) \geq U(a_j, a_{-j})$. Since $U(a_i, a_{-i}) = K - ca_i$ and K is the maximal possible value for the positive part, the only way to increase the node i's utility is to decrease a_i, i.e.,$a_i^{PI} < a_i$. We consider two cases here.

Case 1: $a_i \leq min_{j \in \mathcal{N}(i)} a_j$. Since (a_i, a_{-i}) is a Nash equilibrium, we have $\sum_{j \in \mathcal{N}(i)} min(a_j, a_i) = |\mathcal{N}(i)| a_i = K$. $a_i^{PI} < a_i$ infers that $\sum_{j \in \mathcal{N}(i)} min(a_j^{PI}, a_i^{PI}) \leq |\mathcal{N}(i)| a_i^{PI}$. Hence

$$\begin{aligned} U(a_i^{PI}, a_{-i}^{PI}) &\leq |\mathcal{N}(i)| a_i^{PI} - ca_i^{PI} \\ &= (|\mathcal{N}(i)| - c) a_i^{PI} \\ &< (|\mathcal{N}(i)| - c) a_i \\ &= U(a_i, a_{-i}) \end{aligned}$$

, which is contradictory to the assumption.

Case 2: $a_i > min_{j \in \mathcal{N}(i)} a_j$. Since (a_i, a_{-i}) is a Nash equilibrium, $a_i \leq max_{j \in \mathcal{N}(i)} a_j$.sort a_j $(\forall j \in \mathcal{N}(i))$ in ascending order, denote the order as $a_{j_1}, a_{j_2}, ..., a_{j_{|\mathcal{N}(i)|}}$. $\exists k < |\mathcal{N}(i)|$ such that $a_{j_k} < a_i \leq a_{j_{k+1}}$. If none of a_{j_1} through a_{j_k}

[3]Notice that in this paper, we assume $a_i \in [a_{min}, a_{max}]$ is in the range of $[0, K]$, which is equal to stating that $a_{max} \geq K$. However, all the results in this paper can be easily adapted to the case where $a_{max} < K$. To handle the case $a_{max} < K$, the corresponding change on the definition of saturated nodes should be that the nodes either satisfy $\sum_{j \in \mathcal{N}(i)} min(a_j, a_i) = K$ or the nodes' $a_i = a_{max}$.

increases its accuracy value, since $a_i^{PI} < a_i$,

$$\begin{aligned} U(a_i^{PI}, a_{-i}^{PI}) &\le \sum_{q=1}^{k} a_{j_q} + (|\mathcal{N}(i)| - k)a_i^{PI} - ca_i^{PI} \\ &< \sum_{q=1}^{k} a_{j_q} + (|\mathcal{N}(i)| - k - c)a_i \\ &= U(a_i, a_{-i}) \end{aligned}$$

This contradicts the assumption. Hence, we claim that there exists a q $(1 \le q \le k)$ such that $a_{j_q} < a_{j_q}^{PI}$. Now we investigate the utility for node j_q. According to the definition of Pareto improvement, we have $U(a_{j_q}, a_{-j_q}) \le U(a_{j_q}^{PI}, a_{-j_q}^{PI})$. Notice that $a_{j_q} < a_{j_q}^{PI}$, in the previous Nash equilibrium, node a_{j_q} must NOT be a saturated node (otherwise, increase a_{j_q} can only decrease the node's utility). Therefore, $a_{j_q} = max_{p \in \mathcal{N}(j_q)} a_p \ge a_i$. However, this equation contradicts the previous claim that $a_{j_q} \le a_{j_k} < a_i$. □

Lemma 2: Given a Nash equilibrium $(a_1, a_2, ..., a_{|\mathcal{N}|})$, no Pareto Improvement can be made for node i in this Nash equilibrium if $\forall j \in \mathcal{N}(i)$, $j \in \mathcal{S}$. Such node i is called a **Constrained** node. Let $\mathcal{C}$ denote the set containing all constrained nodes.

Proof:(by contradiction) Assume that there exists a node i that can change its strategy to a_i^{PI} such that $U(a_i, a_{-i}) < U(a_i^{PI}, a_{-i}^{PI})$. According to the definition of Nash equilibrium, no node is willing to change its strategy unilaterally. Hence, in this problem, Pareto Improvement needs to involve at least two neighboring nodes to change strategies simultaneously. Without loss of generality, suppose that node i's neighbor k changes strategy to a_k^{PI} with node i while node i's all other neighbors keep the same strategy. Notice that node k is a saturated node in the given Nash equilibrium. According to the proof of Lemma 1, the only way to keep or increase k's utility is to decrease a_k. That is $a_k > a_k^{PI}$. In the given Nash equilibrium, node i is not saturated. Therefore, $a_i = max_{j \in \mathcal{N}(i)} a_j \ge a_k$. Since in all the neighbors of node i, node k decreases its accuracy value and all other nodes keep same, in order to improve its utility, node i has to increase its accuracy value. That is, $a_i^{PI} > a_i$. A contradiction follows, since $a_i^{PI} = max_{j \in \mathcal{N}(i)} a_j^{PI} \le max_{j \in \mathcal{N}(i)} a_j^{PI} = a_i$. □

Lemma 3: Given a Nash equilibrium $(a_1, a_2, ..., a_{|\mathcal{N}|})$, for node $i \in \mathcal{N} - (\mathcal{S} \bigcup \mathcal{C})$, we can infer that $\sum_{j \in \mathcal{N}(i)} min(a_i, a_j) < K$. Furthermore, $\exists \tilde{j} \in \mathcal{N}(i) \bigcap (\mathcal{N} - (\mathcal{S} \bigcup \mathcal{C}))$ such that $a_i = a_{\tilde{j}} = \max_{j \in \mathcal{N}(i)} a_j$.

Proof: From Lemma 1 and the definition of saturated node, node i does not saturated is the same as the condition $K > \sum_{j \in \mathcal{N}(i)} min(a_i, a_j)$. According to the best response calculation, if node i is not a saturated node, in the Nash equilibrium, $a_i = max_{j \in \mathcal{N}(i)} a_j$. From the fact that node i is not constrained, we can infer that there exists a node within neighborhood of node i that not saturated, denote the node as $\tilde{j}$. We have $a_i \ge a_{\tilde{j}}$. Since the neighborhood range is symmetric, node j and node i are each other's neighbor. Since node i is not saturated, node j is not a constrained node. That is, $\tilde{j} \in \mathcal{N} - (\mathcal{S} \bigcup \mathcal{C})$. Therefore, $a_{\tilde{j}} = max_{j \in \mathcal{N}(\tilde{j})} a_j \ge a_i$. These facts allow us to conclude that $a_i = a_{\tilde{j}} = \max_{j \in \mathcal{N}(i)} a_j$. □

Lemma 3 states that if a node i is neither saturated or constrained, there must exists at least one other node (denote as node $\tilde{j}$) within the neighborhood range of node i that is neither saturated nor constrained. Further, such a node $\tilde{j}$ has the same strategy as node i, which is the largest granularity among all their neighbors.

Let $\mathcal{PN}(i) = \mathcal{N}(i) \bigcap (\mathcal{N} - (\mathcal{S} \bigcup \mathcal{C}))$, lemma 3 suggests an approach to improve an unsaturated node to a saturated node. The step size of increase for node i and its improvable neighbors to make node i saturated is $Inc(i) = \frac{K - \sum_{j \in \mathcal{N}(i)} a_j}{|\mathcal{PN}(i)|}$[4].

Lemma 4: Given a Nash equilibrium $S = (a_1, a_2, ..., a_{|\mathcal{N}|})$, let set $\mathcal{PN}$ contain all the improvable nodes in the network. Consider a strategy profile $\tilde{S}$ where $\forall i \in \mathcal{PN}$, $\tilde{a}_i = a_i + Inc$ and $\forall i \notin \mathcal{PN}$, $\tilde{a}_i = a_i$. $\tilde{S}$ is also a Nash equilibrium if the non-negative number Inc is such that the following condition holds:

$$\forall i \in \mathcal{PN}, \quad \sum_{j \in \mathcal{N}(i)} \tilde{a}_j \le K$$

We omit the proof for this lemma here for brevity. This lemma can be proved using the previous lemmas and considering the users' best responses.

Theorem 1: Starting from a Nash equilibrium, after one iteration of improvement described in Algorithm 4, the resulting strategy vector is still a Nash equilibrium.

This theorem states the correctness of Algorithm 4 to find one Pareto optimal Nash equilibrium. It can be directly derived from the above four lemmas by taking into account that the initial state of the algorithm is a Nash equilibrium.

Corollary 1: Given a Nash equilibrium, after applying Algorithm 4, the resulting strategy vector $(a_1, a_2, ..., a_{|\mathcal{N}|})$ satisfies the condition that $\mathcal{S} \bigcup \mathcal{C} = \mathcal{N}$, this strategy vector is a Pareto optimal Nash equilibrium.

This corollary checks the end state of algorithm 4. When all nodes are either saturated or constrained (or both) in a Nash Equilibrium, no Pareto improvement can be made according to Lemma 1 and Lemma 2. Theorem 1 keeps the result after each iteration as one Nash equilibrium. Therefore, if the algorithm converges, it will converge to a Pareto optimal Nash equilibrium.

We would like to point out that in arbitrary games, there might not exist a strategy profile that is both Nash equilibrium and Pareto optimal. However, we show that a Pareto optimal Nash equilibrium exists in this game by finding the strategy vector. We provide Algorithm 4 to obtain a Parato optimal Nash equilibrium. The algorithm moves a given Nash equilibrium along a Pareto improvement path to achieve both Pareto efficiency and stability (i.e., Nash equilibrium).

Proposition 1: Algorithm 4 guarantees convergence.

Notice that in each PI iteration, we add at least one more node to saturated nodes set $\mathcal{S}$. Since the saturated nodes will keep saturated afterwards, the PI algorithm is guaranteed to converge in $|\mathcal{N}|$ steps in worst case. The running time of the PI algorithm is $O(|\mathcal{N}|^2)$.

We need to point out that there might exist multiple Pareto optimal solutions. Different initial Nash equilibria might result in different Pareto optimal solutions. For example, in the algorithm description, the initial state is the trivial all-zeroes Nash equilibrium. We can also set the initial state as the converged result of SEQ-BR algorithm, which

[4]If $a_{max} < K$, we just need to change the corresponding part of algorithm 4 as $Inc(i) = \min(a_{max} - a_i, \frac{K - \sum_{j \in \mathcal{N}(i)} a_j}{|\mathcal{PN}(i)|})$. All the lemmas and theorems are still hold after the modification.

Algorithm 4 Pareto Improvement

Initialization: $a_i^0 = 0$; $t = 1$;
while $t \leq |\mathcal{N}|$ **do**
  Check and Flag Saturated Nodes for vector $(a_1^{t-1}, a_2^{t-1}, ..., a_{|\mathcal{N}|}^{t-1})$, Put in Set $\mathcal{S}$;
  Check and Flag Constrained Nodes for vector $(a_1^{t-1}, a_2^{t-1}, ..., a_{|\mathcal{N}|}^{t-1})$, Put in Set $\mathcal{C}$
  if $|\mathcal{S} \bigcup \mathcal{C}| = |\mathcal{N}|$ **then**
    return $(a_1^{t-1}, a_2^{t-1}, ..., a_{|\mathcal{N}|}^{t-1})$ and report convergence;
  end if
  for Each Node $i \in \mathcal{N} - (\mathcal{S} \bigcup \mathcal{C})$ **do**
    Calculate the Improvement $Inc(i) = \frac{K - \sum_{j \in \mathcal{N}(i)} a_j^{t-1}}{|\mathcal{PN}(i)|}$, where $\mathcal{PN}(i) = \mathcal{N}(i) \bigcap (\mathcal{N} - (\mathcal{S} \bigcup \mathcal{C}))$;
  end for
  Pick the minimum value among the increment list $minPI = \min_{i \in \mathcal{N} - (\mathcal{S} \bigcup \mathcal{C})}(Inc(i))$
  for Each Node $i \in \mathcal{N} - (\mathcal{S} \bigcup \mathcal{C})$ **do**
    $a_i^t = a_i^{t-1} + minPI$
  end for
  $t++$;
end while

is guaranteed to be one Nash equilibrium. Both cases will converge to Pareto optimal results after applying the above algorithm. However, the two Pareto optimal results are not necessarily identical.

5.5 Distributed Pareto Improvement

Note that in algorithm 4, users need extra message exchanges to find out the minimum increasing value among all the possible increase.

One centralized way to solve this problem is to involve the base station in selecting the minimum increase value. In each iteration, each node $i \in \mathcal{N} - (\mathcal{S} \bigcup \mathcal{C})$ send a message to the base station, reporting its increasing value. The base picks the minimum value and multi-cast the information to each improvable node.

Another option to get the minimum increase value among all flexible nodes is to use the *FloodMin* algorithm [26]. This algorithm makes the calculation totally distributed at the cost of more message exchange. In this algorithm, nodes send messages to their neighbors reporting the increase numbers. The minimum number will be chosen after the messages flooding throughout the network. The details of the FloodMin algorithm are described in Algorithm 5.

Algorithm 5 FloodMin

Initialization: $min\_inc = inc(i)$; $t = 1$; $tmax$ is the network diameter;
msgs$_i$:
if $t < tmax$ **then**
  send min_inc to all $j \in \mathcal{N}(i)$;
end if
trans$_i$:
$t = t + 1$;
let U be the set of increase values that arrive at node i;
$min\_inc = \min(\{min\_inc\} \bigcup U)$;
if $t == tmax$ **then**
  return $min - inc$;
end if

6. SIMULATION

We conduct the simulations using Matlab. In the simulations, there are 10 sets of different node locations. In each set of node deployment, 20 nodes are randomly located on a 20×20 square. A distance based model is used in the simulation to generate network topology. In this model, each node has same neighborhood range. The benefit upper bound K is set to be 100 in all the simulations and the penalty factor c is set to be 0.1.

Figure 2 and Figure 3 show an example of nodes deploying with nodes' indexes, network topology when setting neighborhood range $R = 5$, converged results for SEQ-BR and PI algorithms respectively. The accuracy value beside each node in Figure 3 is rounded to integers for the sake of clear illustration.

From the algorithm results, we observe a property that the user with more neighbors are likely to have better privacy preservation (i.e., shared accuracy value is low). For example, node 20 has maximal degree of 7 and minimal sharing granularity at level 13. This is a desirable property for the Aegis application. Intuitively, the more neighbors a user has, the more likely that the user is in a safe place. In this case, the user does not need to compromise his privacy to other users to improve his safety.

Another observation is that the two Nash equilibria calculated by SEQ-BR algorithm and PI algorithm are not necessarily the same. Node 1 and node 17 in the triangle at the right part of the graph have improved granularity from level 50 to 67. We can verify that the result of SEQ-BR algorithm is a Nash equilibrium. For node 1 and 17, unilaterally change their strategies will decrease their utility. However, changing their strategy simultaneously can result in utility improvement for both nodes and let the system stay in a new stable state. This improvement is essential for this particular application. It increases both user 1 and user 17's safety level.

Figure 4 illustrates the number of constrained but not saturated nodes (i.e., the cardinality of set $\mathcal{C} - \mathcal{S}$) in the network when node's neighborhood range varies as integers from 5 to 34. All the statistical results presented below are averaged over 10 different node deployment for each neighborhood range value. This result shows that the number of constrained but not saturated nodes decreases with increase in a node's neighborhood range R.

Figure 5 plots the social welfare for the Nash equilibria obtained by the two algorithms. Social welfare is defined here as the summation of all nodes' utilities throughout the network (20 nodes in our simulations). The summation is a statistic value averaged over 10 different node locations. The initial state of PI algorithm is the Nash equilibrium output by sequential best response dynamic. Two facts are observed from this plot: a) when the graph is relatively sparse, the percentage of improvement on social welfare is about 10%; b) the improvement percentage decreases with the increase of neighborhood range. Overall, PI results in more efficient Nash equilibrium than SEQ-BR algorithm in terms of social welfare.

Figure 6 illustrates the number of instances improved by PI algorithm. In this simulation, we use the output of the SEQ-BR algorithm as the initial state for PI algorithm. The plot shows that when the neighborhood range is small or medium ($R \leq 18$), almost every Nash equilibrium obtained by SEQ-BR algorithm can be improved. SEQ-BR algorithm

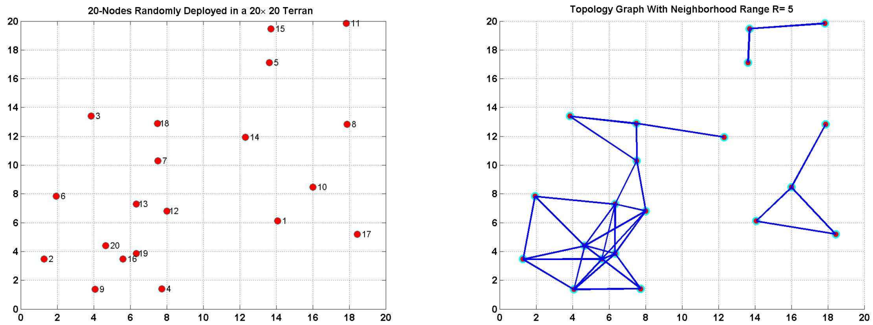

Figure 2: An example: A Random Node Deployment and Corresponding Network Topology when Setting Neighborhood Range $R = 5$

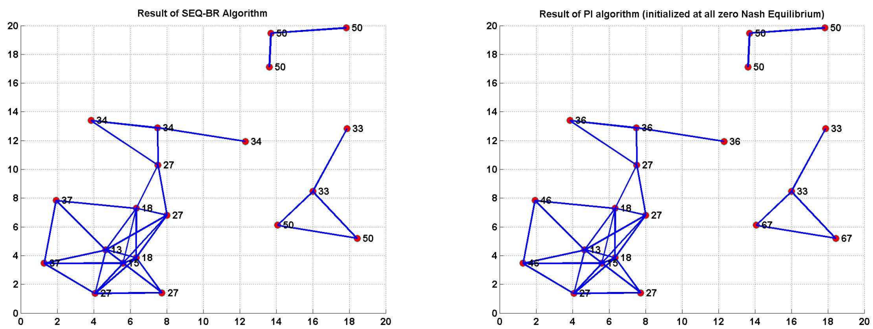

Figure 3: An example: SEQ-BR Algorithm and PI Algorithm Converged Results

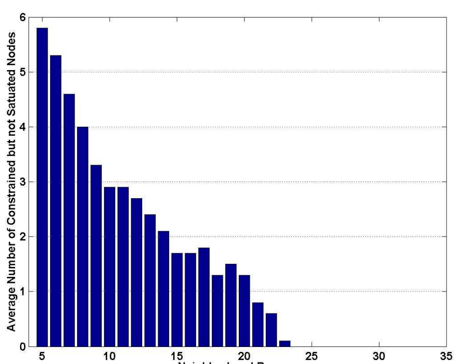

Figure 4: Number of Constrained but not Saturated Nodes VS Node neighborhood range

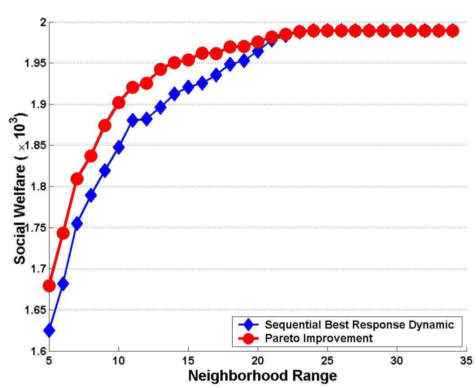

Figure 5: Social Welfare Comparison between SEQ-BR and PI algorithms

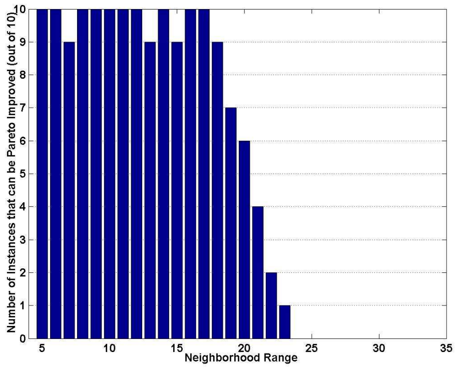

Figure 6: Number of Instances Improved for SEQ-BR Algorithm by PI Algorithm (out of 10)

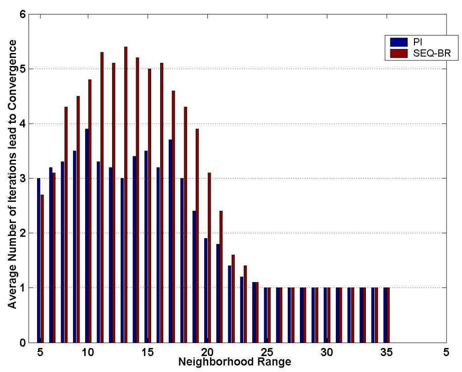

Figure 7: Average Number of Iterations Took to Convergence for SEQ-BR and PI Algorithms

gives Pareto-optimal Nash equilibrium when the topology is presented as a complete graph. We need to point out that the initial state of SEQ-BR (initialized as $a_i = K\ \forall i$) ensures the Pareto-optimal solution for complete graph. Random initialization will not necessarily lead SEQ-BR algorithm to this particular Nash equilibrium.

Figure 7 compares the average iterations to achieve convergence for SEQ-BR and PI algorithms with respect to varied neighborhood range R. On average, the PI algorithm converges faster than SEQ-BR algorithm. For both algorithms, medium density (i.e., $8 \leq R \leq 20$) requires more iterations to converge than in a sparse graph ($R < 8$) or in a dense graph ($R > 20$ where the graph is (or close to) a complete graph).

Together, these simulations lead us to conclude that the PI algorithm is superior. It offers fast, provable convergence in polynomial time, and good quality solution in a distributed manner.

7. CONCLUSION

In this paper, we described a novel community-based mobile application. The application asks the users to share some location information with neighboring friends to enhance security. Considering that the location information is a privacy information which users prefer to preserve, we formulate the application as a game. The utility function gives the users incentives to reveal more accuracy on the location information while there are a few friends around and be conservative on the information accuracy when more friends appear in the neighborhood.

We have illustrated how to calculate the best response for a particular user when fixing all other users' strategies. Furthermore, we investigated several learning dynamics in the system. We point out that the synchronized best response dynamic does not guarantee convergence. To get more control on the resulting equilibrium, we propose an algorithm which can not only guarantee convergence but also is able to move any Nash equilibrium to a Pareto optimal Nash equilibrium. The simulations on different network topologies compare the sequential best response dynamic with the Pareto improvement algorithm. We find that Pareto improvement can give better social welfare in most cases when the network topology is not a complete graph.

In the future, we plan to investigate more general utility functions. We are also interested in solving this problem under dynamic settings when the user configurations change over time.

8. REFERENCES

[1] http://en.wikipedia.org/wiki/Mobile phone

[2] http://news.bbc.co.uk/2/hi/technology/6245798.stm

[3] http://www.nbc11.com/news/15255056/detail.html

[4] A. Adams. Multimedia information changes the whole privacy ballgame. In *Proceedings of the tenth conference on Computers, freedom and privacy*, pages 25–32, 2000.

[5] A. Agah, S.K. Das, and K. Basu. A game theory based approach for security in wireless sensor networks. In *International Conference on Performance, Computing, and Communications*, pages 259-263, 2004.

[6] M. Annavaram, Q. Jacobson and J.P. Shen. HangOut: A Privacy Preserving Social Networking Application Invited paper. To Appear in *the workshop on Mobile Devices and Urban Sensing*, April, 2008.

[7] L. Barkhuus, and A.K. Dey. Location-based services for mobile telephony: a study of users' privacy concerns. In *Proceedings of the 9th International Conference on Human-Computer Interaction*, 2003.

[8] L. Capra, W. Emmerich, and C. Mascolo. A micro-economic approach to conflict resolution in mobile computing. In *Proceedings of the 10th symposium on Foundations of software engineering*, pages 31–40, 2002.

[9] R. Clarke. Internet privacy concerns confirm the case for intervention. *Communications of ACM*, 42(2):60–67, 1999.

[10] 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. In *Proceedings of the conference on Human factors in computing systems*, pages 81–90, 2005.

[11] L.P. Cox, A. Dalton, and V. Marupadi. Smokescreen: flexible privacy controls for presence-sharing. In *Proceedings of the 5th international conference on Mobile systems, applications and services*, pages 233–245, 2007.

[12] S.G. Davies. Re-engineering the right to privacy: how privacy has been transformed from a right to a

commodity. In *Technology and privacy: the new landscape*, MIT Press, 1997.

[13] L. Desmet, W. Joosen, F. Massacci, K. Naliuka, P. Philippaerts, F. Piessens, and D. Vanoverberghel. A flexible security architecture to support third-party applications on mobile devices. In *Proceedings of Workshop on Computer security architecture*, pages 19-28, 2007.

[14] N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Personal Ubiquitous Comput.*, 10(4):255–268, 2006.

[15] D. Fudenberg and J. Tirole. Game Theory. *MIT Press*, 1991.

[16] A.K. Ghosh and T.M. Swaminatha. Software security and privacy risks in mobile e-commerce. *Communications of ACM*, 44(2):51–57, 2001.

[17] T. Grandison and M. Sloman. A survey of trust in internet application. *IEEE Communications Surveys & Tutorials*, (Fourth Quarter), 2000.

[18] M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proceedings of the 1st international conference on Mobile systems, applications and services*, pages 31–42, 2003.

[19] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady. Preserving privacy in gps traces via uncertainty-aware path cloaking. In *Proceedings of the 14th ACM conference on Computer and communications security*, pages 161–171, 2007.

[20] B. Hoh, M. Gruteser, M. Annavaram, Q. Jacobson, R. Herring, J. Ban, D. Work, J. Herrera, and A. Bayen. Virtual trip lines for distributed privacy-preserving traffic monitoring. To Appear in *Proceedings of the 6th international conference on Mobile systems, applications and services*, June, 2008.

[21] J.I. Hong, J.D. Ng, S. Lederer, and J.A. Landay. Privacy risk models for designing privacy-sensitive ubiquitous computing systems. In *Proceedings of the 5th conference on Designing interactive systems*, pages 91–100, 2004.

[22] J.I. Hong and J.A. Landay. An architecture for privacy-sensitive ubiquitous computing. In *Proceedings of the 2nd international conference on Mobile systems, applications, and services*, pages 177–189, 2004.

[23] G. Iachello, I. Smith, S. Consolvo, M. Chen, and G.D. Abowd. Developing privacy guidelines for social location disclosure applications and services. In *Proceedings of the 2005 symposium on Usable privacy and security*, pages 65–76, 2005.

[24] S. Lederer, J. Mankoff, and A.K. Dey. Who wants to know what when? privacy preference determinants in ubiquitous computing. In *extended abstracts on Human factors in computing systems*, pages 724–725, 2003.

[25] H. Liu, B. Krishnamachari, and S. Kapadia. Game Theoretic Tools Applied to Wireless Networks. *Encyclopedia of Ad Hoc and Ubiquitous Computing, Editor D. Agrawal*, World Scientific Publishers, 2008.

[26] N.A.Lynch. Distributed Algorithms. *Morgan Kaufmann Publisher,* 1997.

[27] A. Mackenzie and L. DaSilva. Game Theory for Wireless Engineers. *Morgan & Claypool Publishers*, March 2006.

[28] M.F. Mokbel, C. Chow, and W.G. Aref. The new casper: query processing for location services without compromising privacy. In *Proceedings of the 32nd international conference on Very large data bases*, pages 763–774. 2006.

[29] J. Nasar and P. Hecht and R. Wener. Call if You Have Trouble: Mobile Phones and Safety among College Students, In *International Journal of Urban and Regional Research*, 31(4), pages 863-873, December 2007.

[30] L. Palen and P. Dourish. Unpacking "privacy" for a networked world. In *Proceedings of the conference on Human factors in computing systems*, pages 129–136, 2003.

[31] S. Patil and A. Kobsa. Uncovering privacy attitudes and practices in instant messaging. In *Proceedings of the 2005 international conference on Supporting group work*, pages 109–112, 2005.

[32] A. Patwardhan, F. Perich, A. Joshi, T. Finin, and Y. Yesha. Active Collaborations for Trustworthy Data Management in Ad Hoc Networks. In *Proceedings of the 2nd IEEE International Conference on Mobile Ad-Hoc and Sensor Systems*, November 2005.

[33] A. Patwardhan, F. Perich, A. Joshi, T. Finin, and T. Yesha. Querying in Packs: Trustworthy Data Management in Ad Hoc Networks. *International Journal of Wireless Information Networks*, April 2006.

[34] S. Reddy, J. Burke, D. Estrin, M. Hansen, and M. Srivastava. A framework for data quality and feedback in participatory sensing. In *Proceedings of the 5th international conference on Embedded networked sensor systems*, pages 417–418, 2007.

[35] I. Smith, S. Consolvo, J. Hightower, J. Hughes, G. Iachello, A. LaMarca, J. Scott, T. Sohn, and G. Abowd. Social Disclosure of Place: From Location Technology to Communication Practice. In *Proceedings of the International Conference on Pervasive Computing*, May 2005.

[36] J. Tang, V. Terziyan and J. Veijalainen. Distributed PIN verification scheme for improving security of mobile devices. In *Mobile Networks and Applications*, 8(2), pages 159–175, 2003.

[37] K.P. Tang, P. Keyani, J. Fogarty, and J.I. Hong. Putting people in their place: an anonymous and privacy-sensitive approach to collecting sensed data in location-based applications. In *Proceedings of the conference on Human Factors in computing systems*, pages 93–102, 2006.

[38] M. Voorneveld, P. Borm, F. Van Megen, S. Tijs, and G. Facchini. Congestion games and potentials reconsidered. *International Game Theory Review*, pages 283–299, 1999.