Location Privacy and Game-Theoretic Analysis

Tingshan Huang

***Abstract*—In the recent mobile wireless network, location privacy is serious situation. Users' identities can be easily exposed and their positions can be easily tracked by eavesdropping or other illegal usage of location-aware applications. In this paper, we first discuss the potential privacy violation in using location-aware applications. Then we present and compare three *pseudonym*-based techniques for location privacy protection, and two different ways for anonymity measurements. Next, a user-centric model that are widely used is discussed. Consider that most techniques require users to cooperate, we present a non-cooperative location privacy protocol and analyze the achievable gain in anonymity using game theory.**

I. Introduction

Recently, technologies for locating and tracking individuals are emerging at speeding pace. These location-aware applications provide useful service by tracking people's movement and providing appropriate information for each movement. For example, students in the campus may want to track the location of his/her friends so they can meet sooner, or they may want to track the location of the campus shuttle bus using cell phones in their dorms so that they won't miss the bus when it is 12F outside.

On the other side, we can see these location-aware applications have the potential to track every movement of some device or some one, and they can be more pervasive in the future. Also, we don't want these applications to communication with each other and get our position revealed to all applications. For example, we may use some location-aware application to make our favorite coffee shops to know our visits during working hours, but we may not want the location-aware applications in our company to know we go out for a coffee all the time. Since we don't trust these applications and we have no control over them, we should always assume that these applications would collude against us.

Considering this tension, we need to be concerned about our location privacy before these applications become too pervasive. Here the notion of location privacy is defined as the ability to prevent other parties from learning one's current or past location. Of course, we don't want to stop all the location-aware services, since they could use our position information to provide useful services. What we need to do is to control access to our position information for privacy protection and at the same time make use of these location-aware applications.

One solution to protect location privacy is to build or modify location-aware applications to use *pseudonym*s to hide true user identities. Pseudonym is simply a bit string that serves as public key for identification and end-point authentication. For location sighting, location-aware applications communicate users by tracing their pseudonyms, so users become anonymous in user sightings for applications. It is also encouraged that users should adopt different pseudonyms for different applications to prevent applications from sharing information with each other. However, simple application of anonymity cannot solve security issues since using a long-term pseudonym for each user does not provide much privacy. This is because a malicious attacker can violate anonymity by intercepting messages, stealing data from service providers and *et al.* and expose the real identity of a user.

Many strategies have been developed to protect personal location information by using anonymity. One kind of strategy is to control access to personal location information. One such strategy is proposed by the Geographic Location/Privacy (Geopriv) Working Group [4], which uses pseudonym for identity and end-point authentication. In this design the linking between the pseudonym and its holder is initially known only to the holder himself and a trusted server of the user, after that users are allowed to deliver pseudonym encrypted location information using rule-based policies. Another strategy is proposed by Hengartner and Steenkiste [5], which add digital certificates to improve strategy by Geopriv.

Another kind of strategy is to degrade location information in a controlled way before releasing it. Gruteser and Grunwald developed *Adaptive-Interval Cloaking Algorithm* in [6] which works to increase location privacy by reducing the resolution of location presented to location-aware applications. This algorithm subdivides the area around a subject's position until the number of subjects in the area falls below certain constraint k so that the location information of this subject is indistinguishable from the location information of at least k other subjects. Chaum proposed the technique of *multiple pseudonym* and constructed *mix network* in [3], where each user changes its pseudonyms alternatively to hide its identity. Beresford *et al.* developed *mix zone* in [7] and [8].

The structure of this paper is as follows. In Section II, we present three techniques for location privacy protection. In Section III, we discuss two metrics and compare them base on their effectiveness in measuring anonymity. In Section IV, we presente a cooperative model where users can cooperate with each to protect their location privacy. In Section V, we describe some basic concepts in game theory and use these concepts to analyze a non-cooperative model in Section VI. We conclude the paper in Section VII.

II. Techniques for Location Privacy Protection

In this section, we introduce three techniques for location privacy protection, *multiple pseudonym*, *mix network* and *mix zone*. By the end we compare these three techniques based on their effectiveness in privacy protection.

A. Mix Network

In [3], Chaum constructed *mix network*. A mix network consists of normal message-routing nodes and mix nodes. A

mix node works by mixing the packets it has collected and forward them to others. The way for mixing could be reorder the packets randomly or lexicographically. In this way, the source and destination of a message cannot be linked. In such a store-and-forward manner, the mix network can make communications within the network anonymous. Therefore, malicious attacker $\mathcal{A}$ can not trace the route of a message even though $\mathcal{A}$ knows the topology of the mix network.

However, it is still possible that the mix nodes may collude together with $\mathcal{A}$. If so, the communication is completely exposed and thus is no longer anonymous.

B. Multiple Pseudonym

In [3], Chaum proposed the *multiple pseudonym* approach in the context of Internet communications, and aimed to the traffic analysis problem: how to keep confidential the information of who converses with whom, and when they converse. This technique is based on public key encryption, and uses a public key named digital *pseudonym* for authentication of anonymous applications. It was later applied for location privacy in mobile ad hoc network (MANET) as follows. First an authority creates a list of pseudonyms and assigns a set of pseudonyms to each node. Later on, pseudonyms are used by nodes alternatively for communications with each other or with the authority. In this way, the authority is unable to trace the pseudonym for each user.

Assume a pervasive communication network, where a mobile user periodically broadcast to the outside its information such as the user's authentic information, position and speed. Therefore, users in communication range always know the existence of others. Also assume this network uses multiple pseudonyms and there are N nodes in the network. Before node i enter the network it is preloaded with M pairs of asymmetric keys, $\{K_{i,m}, K_{i,m}^{-1}, m = 1, \cdots, M\}$. Here, $K_{i,m}, m = 1, \cdots, M$ are pseudonyms that serve as public key for communication within the network. From time to time, each user will change its public key to another pseudonym from the set of pseudonyms it has been preloaded with, so that its positions over time is less likely to be traced.

While the approach multiple pseudonym can increase anonymity to some extent, this application requires coordination for pseudonym changes: when should the users change their pseudonyms. One solution to change pseudonyms periodically [9]. When at least two users are within the communication range, they can change their pseudonyms at the time, so that they become indistinguishable to malicious attackers. However, this solution requires synchronization among users, for otherwise there will be only one change at a time and attackers can still track the position of users by observing and recording the change of this lonely user. Also it requires there to be at least one user nearby when it is time for one user to change it pseudonym, for otherwise it will lead to confusion on the attackers' side. In reality, synchronization among users and high population at the same time is rarely met.

C. Mix Zone

In [7] and [8], a mix zone model is introduced where users become anonymous by changing pseudonyms in the mix zones. This technique make users frequently change their pseudonyms so that users are protected form being identified by the locations they visit. In our threat model, the individual does not trust the location-aware application but does trust the raw location system (the sensing infrastructure that can position locatables).

In this model, third-party applications are all assumed to be untrusted, and a trustful event-driven middleware system is assumed to help users hide their identity by positioning between the underlying location systems and untrusted third-part applications. Users don't communicate with applications directly for otherwise their identity will get exposed. Instead, they communication with the middleware system which will then talk with applications. In this way, one application will receive messages from an anonymous user and reply to the same anonymous user.

Before any communications, first users register their interest in applications in a geographic space with the middleware. Then the middleware system divides the whole geographic space into two kind of zones at application level based on the register information: application zones and mix zones. An application zone is an area where at least one user has registered a callback. If a user registers interest in a set of location-aware applications, then the middleware system will assign the corresponding application zones to be the application zone of this user. A mix zone is an area where none of the users have registered a callback. At the same time, middleware tracks the user's location periodically with a period called *update periods*, so that it can detect any possible events of the users. Then the middle waresystem send event callbacks such as enter/exit the application zone to applications.

In this model, users change their pseudonyms upon the boundary between one application zone and one mix zone. When more than two users change their pseudonyms when in a mix zone, identity of one user is mixed with all other users in the same mix zone, and no application is able to trace their locations once these users exit this mix zone. Attackers are unable to link the entrance and exit of multiple users. Therefore, users in the mix zone model is protected from location-tracking.

However, this method requires all users to know the geometry of the mix zones. Also, factors such as geometry of the mix zone and population within the mix zone will influence the achievable privacy. The impact of population in the mix zone is clear since more users in the mix zone will lead to higher amount of anonymity and thus more confusion to the attacker. The effect of mix zone's geometry can be explained by a simple example.

In the figure, A, B and C are three application zones, and the empty space in the middle is a mix zone. Suppose at some point, attacker $\mathcal{A}$ observes that there is a user n_A in application zone A, and a user n_C in application zone C. After one update period, these two users both disappear into the mix zone, and after another update period, one user n_X is observed to be in application zone B. Since the geometry of the mix zone is large enough, $\mathcal{A}$ will know that $n_X = n_A$ since user n_C cannot make it to be in application zone B in one or two update periods. For this reason, users in proximity have to coordinate both in time and space for pseudonym changes.

III. Metric for Anonymity

In this section, we present two metrics for anonymity, one based on *anonymity sets* and the other based on *entropy*. We also compare these two metrics by the end of this section.

A. The Anonymity Set

Observing that the larger number of users in the same area, the more anonymous are the users in this area, Chaum proposed that we could use the size of *anonymity set* to measure anonymity [14]. Here the anonymity set in some area is the set of users that cooperate to change their pseudonyms, and a larger anonymity set provides more anonymity.

However, the size of anonymity set alone can not measure anonymity exactly if we consider the movement of users and there distributions within an area. Assume in a certain area there are n users, one user locates at the extreme of the area, and all other $n-1$ users locate at the other extreme of the area. Also assume that these users change their pseudonyms from time to time. If we consider using the size of anonymity set as the metric, then this lonely user has large anonymity of size n since it is $n-1$ other users in the same area. However, an attacker could easily find out the identity of the lonely user since the attacker find out lonely pseudonym change all the time.

B. Location Privacy

The metric called *location privacy* is proposed in [8]. Suppose that at time t, there are $n(t)$ users in proximity whose pseudonyms has just changed from the set of pseudonyms $B = \{b_1, b_2, \cdots, b_{n(t)}\}$ to $D = \{d_1, d_2, \cdots, d_{n(t)}\}$. Also suppose there is an attacker who $\mathcal{A}$ who has been tracking the location of user i who used pseudonym $b_i \in B$ before its pseudonym change. After the pseudonym change, $\mathcal{A}$ observes the set of pseudonym D and will need to figure out the new pseudonym for user i. The uncertainty of $\mathcal{A}$ would be the uncertainty of user i's new pseudonym given user i's old pseudonym b_i. Assume the transmission probability of d_j given b_i is $p(d_j|b_i)$, then the uncertainty of $\mathcal{A}$ for node i at time t is:

$$\text{Privacy}_i(t) = -\sum_{j=1} n(t)p(d_j|b_i)\log_2(p(d_j|b_i)), \qquad (1)$$

which is also defined as the *location privacy* of node i after its pseudonym change at time t. When user i is the only one that has changed its pseudonym and suppose its new pseudonym is d_k, then $p(d_j|b_i) = 0$ for all j except for $j = k$. In this case, user i's location privacy $\text{privacy}_i(t)$ is zero, since $\mathcal{A}$ can track its pseudonym change. For this reason, we define the situation that more than two users change their pseudonyms in the same period of time. When every user in proximity coordinates to change their pseudonym so that then $p(d_j|b_i) = 1/n(t)$, then $\text{privacy}_i(t)$ has the maximum value of $\log_2 n(t)$. In this case, user i has maximum location privacy since $\mathcal{A}$ cannot figure out which pseudonym user i is using at t and unable to track user i.

From the definition of location privacy, we can see the location privacy achieved by a pseudonym change is upper bounded by $\log_2 n(t)$. We can also see the influence of user popularity at the point of pseudonym change, reflected by $n(t)$, and cooperation of other users as well as the knowledge of attacker $\mathcal{A}$, reflected by $p(d_j|b_i)$. Therefore, when we design a scheme for pseudonym change we need to increase uncertainty to increase user's location privacy.

IV. User-Centric Model

Assume a network system uses multiple pseudonym and silent mix zone. In this setting users change their pseudonyms from time to time to avoid long-term tracking. Also, before a user choose to change its pseudonym, it turns off its transmitter and stops sending messages for certain period of time. By using Swing protocol, when one of the users, say user i, finds its own location privacy level is too low and wants to change its pseudonym, this user sends to its surrounding users a message indicating its pseudonym change and a request for pseudonym change from other user. Upon receiving this message, all other users turn off their transmitters for some period T, during which they will decide whether or not they would also like to change their pseudonyms. Once a user makes the decision, it will change its pseudonym during the silent period. Also, none of the users know the decision of others. In this way, the pseudonym changes of all users seem simultaneously, and the identity of one user who has just changed its pseudonym are mixed with other cooperated users since malicious attacker $\mathcal{A}$ cannot receive any message during the silent period T.

In a distributed setting, each user locally monitor its location privacy level [10] [11] [12]. Based on this information, the user decides to change its pseudonym and start the Swing protocol by sending out request for pseudonym change to other users in proximity, and decides whether or not it should change pseudonym upon a request for pseudonym change. Since the decision is made with local information only, this system is also user-centric.

In a user-centric system, user-centric location privacy level at time t of user i, $\text{privacy}_i(t)$, is influenced by the local privacy of node i after its last pseudonym change at time T_i^p and location privacy loss since last pseudonym change,

$$\text{Privacy}_i(t) = \text{Privacy}_i(T_i^p) - \text{loss}_i(t, T_i^p). \qquad (2)$$

where $\beta_i(t, T_i^{\text{p}})$ represent the location privacy loss function of user i since last pseudonym change at time T_i^{p}. Upon a successful pseudonym change, the loss function is reset to

be zero. After that, the loss function increases based on the user's estimation for the tracking power of malicious attacker $\mathcal{A}$. One simplification for loss function of user i between one successful pseudonym change at time T_i^{pre} and its succeeding successful pseudonym change at time T_i^{s} could be a linear function in time t:

$$\textbf{loss}_i(t, T_i^{pre}) = \begin{cases} \lambda(t - T_i^{\text{p}}), & \text{for } 0 \le t - T_i^{\text{p}} < t_0 \\ \text{privacy}_i(T_i^{\text{p}}), & \text{for } t_0 \le t \le T_i^{\text{s}} \end{cases}$$

where λ is user i's estimation for $\mathcal{A}$'s tracking power, and $t_0 = \text{Privacy}_i(T_i^{\text{p}})/\lambda$ is the time loss function reaches its maximum value $\text{Privacy}_i(T_i^{\text{p}})$.

As we can see from Eq. 2, user's location privacy level decreases in time if it does not change its pseudonym. If the user does not change its pseudonym before t_0, its location privacy level becomes zero after t_0. In other words, the identity of this user will get exposed if one user uses the pseudonym for too long. Therefore, users in the user-centric model will consider pseudonym change before its location privacy level is too low.

Also, upon a successful pseudonym change, location privacy level increases sharply to a new value. This is consistent with the fact that a user can become anonymous by mixing its identity with cooperated users after a successful pseudonym change. Also, if the number of cooperated users are large enough, this user can obtain higher location privacy level and does not need to change its pseudonym too frequently since it can use the same pseudonym for a longer period. On the other side, users will need to change pseudonyms more frequently when the number of cooperated neighbors is small, which leads to low location privacy level achieved by a successful pseudonym change.

V. Basic Concepts in Game Theory

In this section, we present some basic concepts in game theory, the *dominant strategy solution*, *pure strategy Nash equilibrium* and *mix strategy Nash equilibrium*.

A. Simultaneous Move Game

A simultaneous move game consists of two parts [13]: the set of *n players* $P = \{1, \cdots, n\}$, n sets of possible *strategies* for each player i $S = \{S_1, \cdots, S_i, \cdots, S_n\}$. To play the game, player i selects a strategy $s_i \in S_i$, which results in the *vector of strategies* chosen by all players $\mathbf{s} = \{s_1, \cdots, s_n\}$. In such a game all users select their strategies in a simultaneous manner, also different players intend to choose different strategies based on their own preference. One way to specify preferences is to assign a value to each strategy using the set of *payoff functions* $U = \{u_1, \cdots, u_i, \cdots, u_n\}$, where $u_i(\mathbf{s})$ shows the payoff of user i when strategy vector $\mathbf{s}$ is selected by these users, and the higher value for $u_i(\mathbf{s})$, the more preferable is the strategy $\mathbf{s}$ to user i. As we can see from the definition, the payoff of each user depends not only on his own strategy, but also on the strategies chosen by all other players.

B. Dominant Strategy Solution

A game is said to have a dominant strategy solution if it has the following property: each player in this game has a unique best strategy, and this outcome is independent of the strategies selected by other players.

More formally, let's denote the strategy selected by player i as s_i and the strategies selected by all other players as s_{-i}, then the payoff of player i can be represented by $u_i(\mathbf{s}) = u_i(s_i, s_{-i})$. One strategy vector $\mathbf{s}^* = \{s_1^*, \cdots, s_n^*\}$ is a formal definition for dominant strategy solution if for any player $i \in P$, any $s_i \in S_i$, we have

$$u_i(s_i^*, s_{-i}) \ge u_i(s_i, s_{-i}) \tag{4}$$

We need to note here that a dominant strategy solution does not necessarily give optimal payoff to any of the players, and there are few games where each player could have a single dominant strategy. A more realistic solution is for users to maximize their own payoff.

C. Pure Strategy Nash Equilibrium

A strategy vector $\mathbf{s}^*$ is a *Nash equilibrium* if for any play $i \in P$, any $s_i \in S_i$, we have

$$u_i(s_i^*, s_{-i}^*) \ge u_i(s_i, s_{-i}^*) \tag{5}$$

In the Nash equilibrium, none of the players can individually improve his welfare by deviating. Nash equilibrium is also not necessarily optimal for the players and may not be unique. This equilibria is also called pure strategy Nash equilibrium, since each player deterministically plays his own selected strategy.

D. Mixed Strategy Nash Equilibria

In mixed strategy Nash equilibria, players are allowed to select strategies at random, and act to maximize the *expected payoff*. Assume that each player selects strategies independently according to a common probability distribution $f_i(\theta)$ based on some assigned type θ. The distribution f_i is called a *mixed strategy*, and a strategy vector $\mathbf{s}^*$ is a mixed strategy Nash equilibria if for any play $i \in P$, we have

$$s_i^*(\theta_i) \in \arg\max_{s_i \in S_i} \sum_{\theta_{-1}} f(\theta_{-i}) u_i(s_i^*, s_{-i}^*(\theta_i)), \text{ for all } \theta_i, \tag{6}$$

The independent random choices of players leads to $f(\mathbf{s})$, the probability distribution of strategy vector $\mathbf{s}$.

VI. Game-Theoretic analysis

In [1], Freudiger and *et al.* analyze the Nash equilibrium in *n*-player complete and incomplete information games and is the first stop towards understanding the effect of non-cooperative behavior in location privacy schemes. In this section, we first introduce the game model used in [1], then game theoretic analysis on the *n*-player complete, and finally game theoretic analysis on incomplete information games.

A. Pseudonym Change Games

Freudiger and *et al.* considers a *pseudonym change game* G where users selfishly change their pseudonyms for identity protection. They use the entropy-like location privacy as the metric for anonymity, and mix zone model is used for location privacy gain.

This game is defined as a triple set of players P, strategies S and payoff functions U, $S = (P, S, U)$.

1) $P = \{1, \cdots, n(t)\}$ is the set of n nodes in proximity of each other at time t. $n(t) > 1$ is assumed so that there is no lonely pseudonym change in a mix zone. Also each user knows the existences of all other nodes.
2) $S = \{s_1, \cdots, s_{n(t)}\}$ is the set of strategies for each user. $s_i = C$ (*Cooperate*) if user i cooperates to change its pseudonym, and $s_i = D$ (*Defect*) if user i choose not to change its pseudonym.
3) $U = \{u_1, \cdots, u_{n(t)}\}$ is the set of payoff function of each user. For user i, the payoff function is defined as

$$u_i(t) = \text{Privacy}_i(t) - \text{Cost}_i(t), \tag{7}$$

where $\text{Privacy}_i(t)$ is the location privacy level of user i as defined in Eq. 2, and $\text{Cost}_i(t)$ represent the cost that user i has spent on changing pseudonyms. In [1], the cost for each pseudonym change is a constant γ for all users, and the cost function is assumed to be linear with α_i, the number of unsuccessful pseudonym changes since last successful pseudonym change. Thus, $\text{Cost}_i(t) = \alpha_i * \gamma$. We can also understand α_i as the number of pseudonyms that are wasted by user i. When making a pseudonym change decision, a user considers the cost for changing pseudonyms and potential location privacy gain. Recall from Eq. 2, the privacy level of a user decreases in time if the user does not change the pseudonym, therefore a user intends to cooperate by changing its pseudonym more frequently if it finds the privacy level of itself is too low. On the other side, some users intends to defect if it finds the potential location privacy gain is smaller than the cost of a pseudonym change.

Since the potential location privacy gain is related to the number of cooperated users $n_C(s_{-i})$ as a result of the strategies made by all other users, the payoff function is better represented as $u_i(t, T_i^p, s_i, s_{-i})$, and all the information obtained by user i upon one decision making can be updated as follows:

$$\text{if}(s_i == C \&\& n_C(s_{-i}) > 0) \tag{8}$$
$$T_i^p = t; \tag{9}$$
$$\alpha_i(t, T_i^p) = 0; \tag{10}$$
$$u_i(t, T_i^p, D, s_{-i}) = \max\{\text{Privacy}_i(T_i^p) - \gamma, u_i^p - \gamma\}; \tag{11}$$
$$\text{else if}(s_i == C \&\& n_C(s_{-i}) == 0) \tag{12}$$
$$u_i(t, T_i^p, C, s_{-i}) = \max\{0, u_i^p - \gamma\}; \tag{13}$$
$$\alpha_i(t, T_i^p) + +; \tag{14}$$
$$\text{else} \tag{15}$$
$$u_i(t, T_i^p, D, s_{-i}) = \max\{0, u_i^p\}; \tag{16}$$
$$\tag{17}$$

where $u_i^p = \text{Privacy}_i(t) - \gamma * \alpha_i(t, T_i^p)$ is the value for location privacy before the decision making. If user i decides to change its pseudonym, and there is at least one cooperated user, then this pseudonym change is successful. Therefore we can set last time of a successful pseudonym change T_i^p to be t, reset the number of unsuccessful pseudonym changes since last successful pseudonym change $\alpha_i(t, T_i^p)$ to be zero, and update the value for payoff function to be new location privacy subtracted by the pseudonym change cost $\text{Privacy}_i(T_i^p) - \gamma$. If user i is the only one who changes its pseudonym, then this pseudonym change is unsuccessful. Therefore, we only need to decrease the payoff by pseudonym change cost and increase the unsuccessful pseudonym change by 1. If user i choose not to change its pseudonym, then nothing changes.

4) Mixed strategy is used when each user has incomplete information about the strategies made by all others. This is more realistic, since some users may not want to share their decisions with others. Each user in the incomplete-information game in [1] is assumed to know probability distribution $f_i(s_i)$, the probability for user i to select strategy s_i, for all $i \in P$.

B. Analysis of Complete Information Game

In the complete information game, each user has the knowledge of payoffs of all other users, and chooses the strategy based on these values. For brevity, here we only list the main result in [1] for complete information game.

1) The 2-player complete information pseudonym change game has two pure-strategy Nash equilibria $\mathbf{s} = (C, C)$ and $\mathbf{s} = (D, D)$, and one mixed-strategy Nash equilibria where player i choose to cooperate with probability $\frac{\gamma}{1-u_{-i}^p}$.
2) The n-player complete information pseudonym change game has at least 1 and at most 2 pure-strategy Nash equilibria. Especillay, the *All Defection* strategy $s_i = D$ for all $i \in P$ is a pure-strategy Nash equilibria. The exits a unique pure-strategy Nash equilibria if thre is a maximal set of cooperation node C^{k^*}, such that $\log_2(|C^{k^*}|) - \gamma > u_i^p \forall i \in C^{k^*}$, and the pure-strategy Nash equilibria is $\mathbf{s}^* = \{s_i^* | s_i^* = C \text{ if } i \in C^{k^*}, s_i^* \text{ if } i \notin C^{k^*}\}$.

The game theoretical analysis on complete information pseudonym change game shows us that each user tries to reduce its consumption of pseudonyms by changing pseudonyms only when their location privacy level is too low and there is at least one user that is willing to cooperate by changing its pseudonym. The n-player game is more asymmetric that the 2-player game for it has more variety over all the users. Also, the All Defection strategy exists for each game since one player cannot gain location privacy by cooperating alone. Furthermore, the Nash equilibria with cooperation is Pareto-optimal, if it exists.

C. Analysis of Incomplete Information Game

For the incomplete information game, players cannot know the payoff of all the other players, and better models the

situations in reality. When making decision, each user choose whether to change its pseudonym or not based on its own location privacy level and its estimation of other users' strategies. Threshold equilibrium is established so that each player select its strategy based on its payoff u_i and a threshold $\tilde{u}_i$:

$$s_i(u_i) = \begin{cases} C, & \text{for } 0 \le u_i^p < \tilde{u}_i \\ D, & \text{for } \tilde{u}_i \le u_i \le \log_2(n) - \gamma \end{cases} \quad (18a)$$

Then the probability for user i to cooperate is:

$$C(\tilde{u}_i) = \Pr(u_i \le \tilde{u}_i) = \int_0^{\tilde{u}_i} f(u_i) du_i \quad (19)$$

and the probability of defection is $D(\tilde{u}_i) = 1 - C(\tilde{u}_i)$.

Suppose at time t there are $n(t)$ players in a incomplete pseudonym change game. Also assume that when players in P
$\{i\}$ choose $\tilde{u}_{-i}$ to be their threshold vector, the probability for k of the other users to cooperate is $\Pr(K = k, \tilde{u}_{-i})$ in user i's knowledge. Then player i's estimation for its average payoff would be:

$$E(u_i(C, \tilde{u}_{-i})) = \sum_{k=0}^{n(t)-1} \Pr(K = k, \tilde{u}_{-i}) u_i(C, \tilde{u}_{-i}); E(u_i(D, \tilde{u}_{-i})) = u_i^p \quad (20)$$

If a player chooses strategies by comparing the average achieved location privacy gain by cooperation and that by detection, then a Bayesian Nash equilibrium (BNE) can be obtained as the solution to the following n equations for variables $\tilde{u}_i$:

$$\sum_{k=0}^{n(t)-1} \Pr(K = k, \tilde{u}_{-i}) u_i(C, \tilde{u}_{-i}^*) = u_i^p, \text{ for } i \in P. \quad (21)$$

The welfare achieved by this BNE is compared with that by *random* strategy and *All Cooperation* in [1] and results are listed as follows:

1) The 2-player incomplete information game has *All Cooperation* and *All Defect* pure-strategy BNE, and every threshold equilibrium $\tilde{\mathbf{u}}^* = (\tilde{u}_1^*, \tilde{u}_2^*)$ is symmetric for continuous distributions. If larger threshold has higher probability, players cooperates more to maintain high privacy. Nodes are less selfish when the cost of a pseudonym change is large. The welfare achieved in the BNE is less than *All Cooperation* strategy and similar to *random* strategy. However, less pseudonyms are consumed in the BNE.
2) The n-player incomplete information game has the *All Defect* pure-strategy BNE and a symmetric threshold equilibrium, while *All Cooperation* strategy is no longer an equilibrium. Higher density of players nearby makes a player feel safe and thus selfish nodes will cooperate less. Also large cost for pseudonym change works as incentive and selfish nodes will cooperate more, the same as in the 2-player incomplete information game.

VII. Conclusion

We have discussed the issue of location privacy in using location-aware applications. Three techniques based on pseudonym changes have been introduced: multiple pseudonym, mix network and mix zone. We have shown the schemes of these three techniques for location privacy protection and have also shown that attackers can still track the users if user population is too low or the spatial-temporal resolution is not satisfied. Next, we presented two ways to measure anonymity of a user, one is based on anonymity set and the other is based on entropy. The entropy is a better metric in that it considers the movement of users as well as the size of anonymity set. We have also presented user-centric model and the non-cooperative pseudonym change game. Using the game theoretic analysis, we have shown the derived equilibria to achieve location privacy in the non-cooperative model, in both complete information and incomplete information pseudonym change game. The summarized results show that the non-cooperative model effectively reduce the consumption of pseudonyms, and the larger cost for changing pseudonyms encourages selfish users to cooperate more.

References

[1] J. Freudiger, M. Manshaei, J.-P. Hubaux, and D. Parkes, "On non-cooperative location privacy: A game-theoretic analysis," in *Proceedings of the 16th ACM Conference on Computer and Communications Security*, Chicago, IL, pp. 324-C337, Nov. 2009.

[2] L. Buttyan and J.-P. Hubaux, "Security and Cooperation in Wireless Networks," *Cambridge University Press*, 2008.

[3] D. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," *Communications of the ACM*, vol. 24, no. 2, pp.84-88, Feb. 1981.

[4] J. R. Cuellar, J. B. Morris, D. K. Mulligan, J. Peterson, and J. Polk, "Geopriv reqs. (IETF Internet draft)," 2003.

[5] U. Hengartner and P. Steenkiste, "Protecting access to people location information," in *Security in Pervasive Computing*, March 2003.

[6] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *ACM/USENIX International Conference on Mobile Systems, Applications and Services (MobiSys)*, 2003.

[7] A. R. Beresford and F. Stajano, "Mix zones: User privacy in location-aware services," in *PerSec*, 2004.

[8] A. R. Beresford and F. Stajano. Location privacy in pervasive computing. IEEE Pervasive Computing, vol. 3, no. 1, pp. 46-55, 2003.

[9] L. Buttyan, T. Holczer, and I. Vajda, "On the effectiveness of changing pseudonyms to provide location privacy in VANETs," in *ESAS*, 2007.

[10] M. Li, K. Sampigethaya, L. Huang, and R. Poovendran, "Swing & swap: User centric approaches towards maximizing location privacy," in *WPES*, 2006.

[11] B. Hoh, M. Gruteser, R. Herring, J. Ban, D. Work, J.-C. Herrera, A. M. Bayen, M. Annavaram, and Q. Jacobson, "Virtual trip lines for distributed privacy-preserving tranc monitoring," in *MobiSys*, pp. 15-28, 2008.

[12] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady, "Preserving privacy in GPS traces via path cloaking," in *CCS*, 2007.

[13] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, "Algorithmic Game Theory," *Cambridge University*, 2007.

[14] D. Chaum, "The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability," *J. Cryptology*, vol. 1, no. 1, pp. 66-75, 1988.