

Information Technology and Quantitative Management , ITQM 2013

## A Framework of Evaluation Location Privacy in Mobile Network

Chao Lee<sup>a,b,\*</sup>, Yunchuan Guo<sup>a</sup>, Lihua Yin<sup>a</sup>

<sup>a</sup>*Institute of Information Engineering Chinese Academy of Sciences, Beijing, 100185, China*

<sup>b</sup>*Graduate University of Chinese Academy of Sciences, Beijing, 100049, China*

---

### Abstract

In modern mobile networks, people increasingly tend to use location-based services, which would share their location information with third-party servers. Such communications leaking users' location information may be observed by adversaries, and then user's individual privacy may be threatened. To solve the problem, researchers have proposed various location privacy preserving methods. The goal of this paper is to build a common model to describe the location privacy protecting mechanism and show how to evaluate the location privacy in our framework. In this paper, we present a framework to model location privacy preserving mechanism, and abstract the properties characterizing the effectiveness of these approaches. We give an adversary model to reconstruct the actual trace through observation and initial knowledge. We give a method to quantify location privacy by comparing the difference between initial uncertainty and remaining uncertainty of suspected trace before and after adversary's attack.

© 2013 The Authors. Published by Elsevier B.V. Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Selection and peer-review under responsibility of the organizers of the 2013 International Conference on Information Technology and Quantitative Management

*Keywords:* location privacy; location privacy metric; obfuscation; anonymity; hidden markov model; mutual information

---

### 1. Introduction

Today, the increasing numbers of objects with computational power and digital communications are surrounding human beings. These devices pervade people everyday life and could be used to improve the quality of their lives. When utilizing these ubiquitous devices and related applications, location is an important characteristic, since location information is often considered as a contextual parameter that the applications based on. For example, if a person wants to find a gas station nearest her, she needs to send the query containing the location information, which is obtained from GPS modular embedded in her mobile phone or vehicle, to a location-based service (LBS) server through the 2G/3G network. So, user's activities are closely related to the location information. If location information is leaked, individual's traveling route, buying habit, state of health, political views and other sensitive information could be inferred, and persons would face several undesirable effects, such as location-based spam, personal safety and intrusive inference. Therefore, privacy of person's location information is becoming an important new issue in privacy protection. Beresford and Stajano [1] define location privacy as the ability to prevent other parties from learning one's current or past location. To solve this problem, many researchers have focused on designing the location privacy preservation methods that would protect user's location privacy [1, 2, 3, 4, 5].

---

\*Corresponding author.

E-mail address: [lichao@nelmail.iie.ac.cn](mailto:lichao@nelmail.iie.ac.cn).

When designing these methods, metrics must be proposed to evaluate the effectiveness of these approaches. The location privacy preserving mechanism is designed based on the metrics, therefore the progress in location privacy research depends on the ability of metrics to quantify location privacy, and the method of how to evaluate location privacy protecting approaches appropriately. In the literature, there are various metrics have been proposed to measure location privacy. However, there is not yet a standard for this, and it is rare for even two different research projects to use the same method of quantification [6].

In this paper, we proposed a quantification framework for modeling and evaluating location privacy. We make the following contributions. We propose a three-dimension model to describe location privacy. The model abstracts mobile users as events, formalizes location privacy preserving mechanisms as noisy channels that modifies the information communicated from the users to the observer, generalizes the concrete method protecting location privacy as two functions of anonymity and obfuscation, and give two properties of location privacy. We build a framework to evaluate location privacy protecting approaches taken into consideration of adversary’s reconstruction ability. We present the approach to quantify the ability of location privacy protecting mechanism.

The structure of the paper is as follows. In Section 2, we described the proposed model of location privacy, Section 3 presents how to reconstruct user’s suspected location via inference attack based on the Hidden Markov Model, and utilize mutual information between the initial uncertainty and remaining uncertainty after inference attack, to describe effectiveness of location privacy protecting mechanism. Finally, we make a discussion and summarize our work in Section 4.

## 2. System model

In this paper, we focus on location based service (LBS) systems that provide location privacy through *location privacy preserving mechanism*. The system model we consider, thus consists of the following three dimensions of entities: *user*, *location privacy mechanism* and *property metric*, which is shown in Figure 1. Simply speaking, the model shows that users’ location privacy is protected by location privacy preserving mechanism, which utilizes *anonymity* and *obfuscation* techniques, and the effectiveness of location privacy protection is evaluated by two property metrics composed of *uncertainty*, *inaccuracy*.

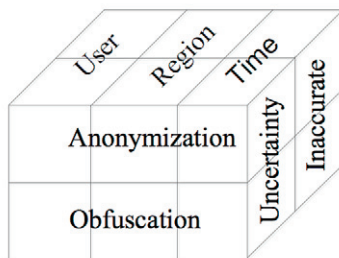


Fig. 1. The System Model of location privacy

### 2.1. Mobile User Model

These are mobile users who send (or have the ability to send) messages to LBS through the wireless infrastructure. These messages can be the interesting request(i.e, where is the restaurant nearest me) and application data, containing their location. For each request, users may identify themselves to the LBS using proper credentials. LBSs provide users with services using the location information from requests.

We denote  $\mathcal{U}$  as the set of users who are members of the mobile network. Users are considered to be able to communicate by wireless network. Further, a user has a set of properties, such as name, gender, occupational, etc. We call the properties as *attributes* of the user. These attributes describe a user’s profile.

In the mobile network, users report their locations periodically, or sporadically when they expect to get some information from LBSs. We denote  $\mathcal{L}$  the set of locations of users. Location  $l \in \mathcal{L} = \{l_1, l_2, \dots, l_n\}$  can be a two-dimensional point which represents a GPS coordinates, or a region that containing the point. This is dependent on

the granularity of application that LBSs divide the area (i.e., each region is divided from an area by each  $100m^2$ ). We denote  $\mathcal{T} = \{t_1, t_2, \dots, t_n\}$  the set of discrete times that when users report their location.

In order to represent the case that a user  $u_i \in \mathcal{U}$  reports her/his location  $l_i \in \mathcal{L}$  to the LBSs at the time  $t_i \in \mathcal{T}$ , we introduce the notion of *event*.

**Definition 1.** An *event* is a triple  $e = \langle u, l, t \rangle$ , where  $u \in \mathcal{U}$  is a user,  $l \in \mathcal{L}$  is a location, and  $t \in \mathcal{T}$  is a discrete time.

In case user  $u$  reports her/his location  $l_i$  at time  $t_j$ , we write  $e_u(t_j) = \langle u, l_i, t_j \rangle$ . The set of events belonged to user  $u$  denoted by  $tr_u$ , is called the *event trace* of  $u$  or *trace* of  $u$  for short. We denote  $tr_u = (e_u(1), e_u(2), \dots, e_u(n))$  to record the events of  $u$  occurred at time  $1, 2, \dots, n$ .

### 2.2. Location privacy preserving mechanism model

When mobile users send their request to LBSs, the location information is contained in the request. Mobile users do not want to expose their location to eavesdropping attackers. However, if the identity or location is not distorted, adversary can know the user did something at a specific place. Therefore, location privacy mechanism is needed. To protect a user’s location privacy, it alters or distorts the information observable by the adversary.

The aim of location privacy protecting mechanism is to prevent an adversary against directly observing person’s actual location, which can be used to infer person’s other attributes. The essential idea of location privacy protecting mechanism is that for a given input event, it outputs one or more events from which a person can get the acceptable quality service from LBSs while the adversary cannot deduce the person’s actual location and identity, as shown in Figure 2.

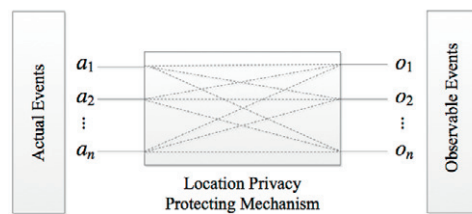


Fig. 2. Location privacy protecting Mechanism channel

We model the location privacy protecting mechanism as an information-theoretic noisy channel, in which the actual events denoted by  $\mathcal{A} = \{a_1, a_2, \dots, a_n\}$  are the input of the channel, and the observed events denoted by  $\mathcal{O} = \{o_1, o_2, \dots, o_m\}$  are the output of the channel. Location privacy protecting method can be seen as the noise of the channel. As the noise in the channel, the input and the output must be have some differences. The difference is the goal of location privacy protecting mechanism making effort to achieve. The more different the location privacy protecting mechanism provided, the more difficult for the adversary to get person’s actual location is.

There are several papers in the field of location privacy protecting. According to Krumm’s work [6], we abstract these approaches into two classifications: *anonymity* and *obfuscation*.

#### 2.2.1. Anonymity

The basic idea of anonymity approaches utilizes a pseudonym and creates ambiguity by grouping with other people. These methods can be classified by:

- *pseudonym*
- *k-anonymity*

For *pseudonym*, the identity of an event is altered in order to break the link between a user and his/her events. Using renaming function, the real identity of a user on each event is replaced by another pseudonym, which makes the attacker cannot observe the event of the actual user directly [7]. For *k-anonymity*, the person reports the region containing  $k - 1$  other users. Person’s identity is indistinguishable from the location information of at least  $k - 1$  other subjects [2].

2.2.2. Obfuscation

The basic idea of obfuscation approaches is used to protect user’s location privacy through *degrading* the quality of information of user’s location data, and at the same time, the degrading location data must be satisfied to service quality which user accepted. This is done by adding noise to the location and/or time-stamp of the events or by coarse graining them. Obfuscation is achieved mostly through *perturbation* or *generalization* algorithms. They can be divided by:

- *hide location*
- *adding noise*
- *reducing precision*

**Definition 2** (Location privacy protecting mechanism). We define a privacy protecting mechanism as a triple

$$\langle \mathcal{A}, \mathcal{O}, \phi \rangle$$

where  $\mathcal{A}$  is a set of actual events, and  $\mathcal{O}$  is a set of observations.  $\phi$  is a probability distribution function.

For each input, there might be different outputs after LPPM processing. To introduce probabilities we associate the random variables  $A$  for  $\mathcal{A}$  and  $O$  for  $\mathcal{O}$  respectively.  $A$  is a random variable representing the input event with  $n$  values  $\{a_1, a_2, \dots, a_n\}$ .  $O$  is a random variable representing the observable outputs events  $m$  values  $\{o_1, o_2, \dots, o_m\}$  after the processing of location privacy protecting mechanism.  $\phi$  denotes the conditional probability between the two random variables of actual events and observed events.

Therefore, the location privacy protecting algorithms for each user  $u$  can be abstracted as

$$f_{a_u}(o_u) = P\{O_u = o_u | A_u = a_u\} \tag{1}$$

The event of user  $u$  is denoted by  $a_u$ , which is the input of the location privacy protecting algorithm. After processing of the algorithm, the degraded events  $o_u$  of user  $u$  are output, which can be observed by adversary. It is worth to mention that, for each  $a_u$ , there can be multiple  $o_u$ .

The probability values  $p(o|a)$  for every output/input pair constitutes the matrix. Typically, the input events are arranged by columns and the output events are arranged by rows. Therefore, the above triple can be represented by the LPPM matrix shown in Table 1. The element of the matrix  $p(o_j|a_i)$  is the condition probability that is the chance of observing  $o_j$  (after LPPM processing, the observable output event that the adversary is able to see) given  $a_i$  as input (the actual location send to the LPPM to process).

Table 1. LPPM matrix

	$o_1$	$o_2$	$o_3$	...	$o_m$
$a_1$	$p(o_1 a_1)$	$p(o_2 a_1)$	$p(o_3 a_1)$	...	$p(o_m a_1)$
$a_2$	$p(o_1 a_2)$	$p(o_2 a_2)$	$p(o_3 a_2)$	...	$p(o_m a_2)$
$a_3$	$p(o_1 a_3)$	$p(o_2 a_3)$	$p(o_3 a_3)$	...	$p(o_m a_3)$
⋮	...	...	...	...	...
$a_n$	$p(o_1 a_n)$	$p(o_2 a_n)$	$p(o_3 a_n)$	...	$p(o_m a_n)$

2.3. Measurement

In literatures, metrics for location privacy are usually mechanism-oriented. For instance, Gruteser and Grunwald [2] introduce *k-anonymity* for location privacy. The metric for evaluating location privacy is the  $k$ . They use  $k$  to measure the level of privacy. In Duckham’s work [8], he uses the number of different coordinates sent by the user in a single query as the metric to evaluate the level of privacy.

If the metric is mechanism-oriented, it is difficult to compare location privacy protecting methods which belongs to different types. In this section, we provide location privacy metric from adversary’s perspective, which is not mechanism-oriented.

**Definition 3** (Suspected location). Adversary utilizes the locations he observed to inference person’s actual location. The location obtained by inference attack is called suspected location, which is denoted by  $a^s$ .  $a^s$  may contains the actual location  $a$ , or may have some distances to  $a$ .

2.3.1. Uncertainty

Uncertainty is a property used to shows how uniform or concentrated the estimated distribution is. In the context of location privacy, it explains the how difficult for adversaries to pinpoint user’s actual location. The more uncertainty of the suspected location, the more difficult for adversaries to get the point of the actual location. In literatures, uncertainty is usually described by entropy:

$$\hat{H}(a^s) = - \sum_{a^s} \hat{P}(a^s|o) \log \hat{P}(a^s|o) \tag{2}$$

2.3.2. Inaccuracy

Inaccuracy is a property that shows how close between the actual location and the suspected location. Adversary deduce person’s actual location through inference attack. The obtaining suspected location must have some distortions compared to the actual location. We use inaccuracy to describe the distortion by the notion of Euclid distance:

$$\sum_{a^s} \hat{P}(a^s|o) \sqrt{(a_x - a_x^s)^2 - (a_y - a_y^s)^2} \tag{3}$$

The distance between the actual location and the suspected location is further, the adversary’s suspected location is more inaccurate.

3. Proposed measurement model

We propose a framework for quantifying the location privacy, which is shown in Figure 3. In the framework, there are four components: *user actual trace*, *location privacy protecting algorithm*, *attack algorithm*, *evaluation algorithm*. A user actual trace *tr* is a set of events  $\mathcal{A}$ , which has been introduced in section 2.1. Users’ *trs* input to the location privacy protecting algorithm to obfuscate users’ actual events. Location privacy protecting algorithm is a triple  $\langle \mathcal{A}, \mathcal{O}, \phi \rangle$  explained in section 2.2. The condition probabilistic distribution is generated by a given location privacy protecting algorithm. The output of the algorithm is condition probabilistic of observing an output event *o*, given input event *a*. A person who wants to measure her/his location privacy protecting algorithm can embed the algorithm into the location privacy protecting algorithm module, or directly provide the distribution function to the attack module. The attack module consists of observable trace aggregation module and the initial knowledge module. The attacker reconstructs users actual trace through observation and knowledge. After obtaining the reconstructed traces, evaluation module quantify the location privacy by comparing the actual traces *trs* and the reconstructed traces *tr*’s.

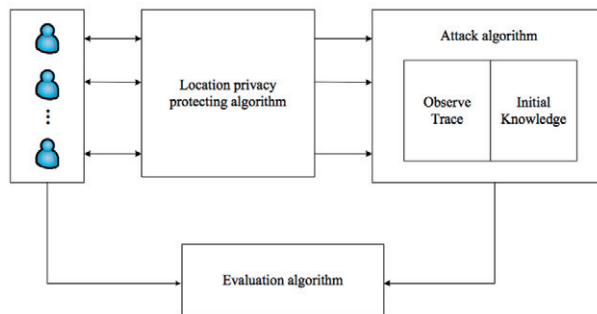


Fig. 3. A framework of evaluation for location privacy

As the modules of user actual traces and location privacy protecting algorithm have been presented in section 2, in this section, we only talk about the attack and evaluation module.

In our framework, the suspected event is the event that reconstructed by an adversary through observations. The goal of the evaluation module is to quantify the difference between the actual event and the suspected event.

Intuitively, if the location privacy protecting algorithm is more power, the adversary is more difficult to reconstruct the accurate event, and the actual event and the suspected event are more different.

Location privacy protecting algorithm can be seen as information-theoretic channel with noise. The noise is a manifestation of the efforts of the location privacy protecting algorithm to hide the link between the inputs and the outputs. The more noise exists in the channel, the more privacy the mechanism is.

The evaluation module is consisted of two part: suspected trace reconstruction module and evaluation algorithm module. The objective of the former one is to reconstruct the user's actual event. And the latter one is to make an evaluation between user's actual event and the reconstructed user's suspected actual event.

### 3.1. Attack module

Adversary can observe the outputs of the information-theoretic channel. We denote the observed event as  $o$ . Adversary can get the suspected event  $a^s$  through inference attack. In this section, we will give the method to reconstruct the suspected trace of users. From the adversary's perspective, he/she has a set of observations, and his/her goal is to deduce user's actual trace. User's actual trace that adversary tries to obtain (suspected trace), is hidden which cannot be observed directly. We can abstract this to the problem of giving an observation sequence  $O = \{o_1, o_2, \dots, o_m\}$ , how to choose a corresponding state sequence (trace)  $\mathcal{A} = \{a_1, a_2, \dots, a_n\}$ , which is optimal to explain the observations. This problem can be modeled by a first order *Hidden Markov Model* (HMM) [9].

**Definition 4** (Hidden Markov Model). A first order Hidden Markov Model

$$\theta = (\Pi, A, B) \quad (4)$$

is defined by

- $\Pi = \{\pi_1, \pi_2, \dots, \pi_n\}$ , the initial state distribution.
- $A = \{a_{ij}\}$ , the state transition matrix, where  $a_{ij}$  denotes the probability of a transition from state  $a_i$  to  $a_j$
- $B = \{b_{ij}\}$ , the confusion matrix, where  $b_{ij}$  denotes the probability of observing  $o_j$  in state  $a_i$ .

Given the form of HMM, there are three basic problems of interest that must be solved for the model to be useful in real-world application. These problems are the following:

- Problem 1: Given an observation sequence  $O = \{o_1, o_2, \dots, o_n\}$ , and a model  $\theta = (\Pi, A, B)$  how to compute  $P(O|\theta)$ , the probability that a given HMM  $\theta$  generates an observation  $O$
- Problem 2: Given an observation sequence  $O = (o_1, o_2, \dots, o_T)$ , and a model  $\theta = (\Pi, A, B)$ , how to choose a corresponding state sequence (path)  $Z = (z_1, z_2, \dots, z_T)$ , which is optimal in some meaningful sense (i.e., best "explains" the observations)?
- Problem 3: Given several observation sequences, how to find the HMM parameters that best describe these observations, i.e., how to adjust the model parameters  $\theta = (\Pi, A, B)$  to maximize  $P(O|\theta)$

As our inference attack is based on HMM, we need to first solve the problem 3 of estimating the parameters of the HMM. And then solve the problem 2 by using the estimated HMM and observation event to predict the suspected event.

In Definition 4, a HMM is defined by the triple of  $\langle \Pi, A, B \rangle$ , where  $A$  is the state transition matrix,  $B$  is the confusion matrix, and  $\Pi$  is the vector of the initial state probabilities. In the context of our location privacy protecting model, the user's actual trace can be seen as a sequence of states, in which each event denotes a state. Moving from one region to another can be seen as transmission between two states. Therefore, the user's actual mobility can be easily encoded by  $A$ , in which the element of  $a_{ij}$  states the probabilistic of transmitting from state  $a_i$  to  $a_j$ . In a HMM,  $B$  denotes the confusion matrix (in some literature, it is also named emission matrix). The elements  $b_{ij}$  is the probabilistic of observing  $o_j$  in the state  $a_i$ . In the context we consider, it can be explained by observing location  $o_j$  when a person in location  $a_i$ . This is in accordance with the semantics of location privacy protecting algorithm of  $p(o_j|a_i)$ .

We assume an adversary has some preliminary knowledge, which is the sequence of events about mobility of the users. These events can be recorded by a transmission matrix  $A^I$ , which adversary initially known. The element  $a_{ij}$  of  $A^I$  is the probabilistic of changing event state from  $a_i$  to  $a_j$ . In other words, it can be seen as the probabilistic

that user moves from region  $a_i$  to  $a_j$ . The adversary knows the location privacy protecting mechanism, therefore the distribution of  $P(O|A)$  is known.

According to the assumption, the paths are known for all the observations. Now let  $Z^k = (z_1^k, z_2^k, \dots, z_n^k)$  is the  $k$ th state sequence or the paths of observation  $o_1, o_2, \dots, o_k$  respectively. For this case, we can count the number of times each particular parameter is used in the set of training sequences. We denote these by  $EN(\pi_i), EN(a_{ij}), EN(b_{ij})$

Then the maximum likelihood (ML) estimators for these parameters are given by:

$$\bar{\pi}_i = \frac{EN(\pi_i)}{\sum_{i'=1}^N EN(\pi_{i'})}, i = 1, 2, \dots, n, \tag{5}$$

$$\bar{a}_{ij} = \frac{EN(a_{ij})}{\sum_{j'=1}^N EN(a_{ij'})}, i, j = 1, 2, \dots, n, \tag{6}$$

$$\bar{b}_{ij} = \frac{EN(b_{ij})}{\sum_{j'=1}^M EN(b_{ij'})}, i, j = 1, 2, \dots, m. \tag{7}$$

After obtaining the parameters of HMM  $\theta = (\Pi, A, B)$ , we try to inference the suspected trace or event when getting the observation events. As described above, this can be abstracted to solve the problem 2 of HMM.

The posterior probability of the state  $a_i$  at time  $t$  for given observation sequences is defined by:

$$\gamma_t(i) = P(z_t = a_i | O, \theta). \tag{8}$$

i.e., the probability of being in the event  $a_i$  at time  $t$ , when observation event sequence is  $O$ .

We can express the equation by the *forward-backward* variables.

$$\begin{aligned} \gamma_t(i) &= P(z_t = a_i | O, \theta) = \frac{P(O, z_t = a_i | \theta)}{P(O | \theta)} \\ &= \frac{P(o_1, o_2, \dots, o_t, o_{t+1}, \dots, o_T, z_t = a_i | \theta)}{P(O | \theta)} \\ &= \frac{P(o_1, o_2, \dots, o_t, z_t = a_i | \theta) P(o_{t+1}, \dots, o_T | o_1, o_2, \dots, o_t, z_t = a_i | \theta)}{P(O | \theta)} \\ &= \frac{P(o_1, o_2, \dots, o_t, z_t = a_i | \theta) P(o_{t+1}, \dots, o_T, z_t = a_i | \theta)}{P(O | \theta)} \\ &= \frac{\alpha_t(i) \beta_t(i)}{P(O | \theta)} = \frac{\alpha_t(i) \beta_t(i)}{\sum_{i'=1}^N [\alpha_t(i') \beta_t(i')]} \end{aligned} \tag{9}$$

Using  $\gamma_t(i)$ , we can solve for the individually most likely state  $\hat{a}_t^s$  at time  $t$ , as

$$a_t^s = \arg \max_{i=1,2,\dots,N} \gamma_t(i), t = 1, 2, \dots, T. \tag{10}$$

### 3.2. Evaluation module

We express the difference as the notion of leakage. The leakage is modeled by the difference between the initial uncertainty and the remaining uncertainty after the observation.

For instance, a user's location  $r_9$  is obfuscated and return a region  $R$ , suppose the region  $R$  is a  $4 \times 4$  grid. For a observer's view, the user may be in any cell of the grid with equal probability. After adversary's inference attack, the suspected locations are  $r_1, r_7, r_{12}, r_{15}$  with different probabilities. As shown in figure 3.2, only 4 regions are the suspected locations for the actual one. The remaining uncertainty is much less than the initial uncertainty. It is easy to see that the less uncertainty is decreased, the more power the algorithm is.

Since the uncertainty is represented by the entropy, the leakage can be represented by the notion of mutual information, which is the difference between the entropy of the input and the conditional entropy of the input given the output.

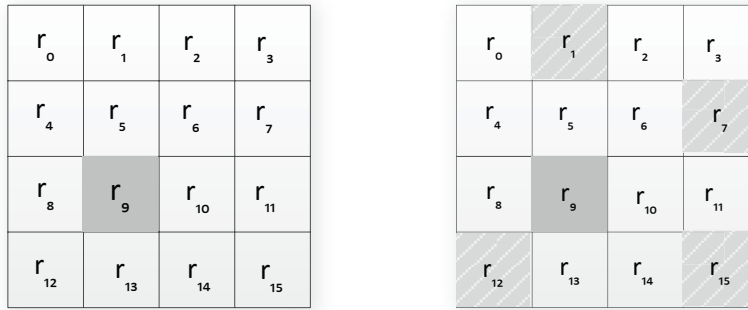


Fig. 4. (a) suspected location before inference attack; (b) suspected locations after inference attack

Let  $A^s$  be the random variable that describes the actual event. Let the attacker's estimate of  $A$  through observation of the system and get the  $A|O$ . The information carried through the observation channel provided by the attack is therefore  $I(A|O)$ . The higher this carried information, the more accurate the attack, and thus the worse the LPPM. The effectiveness of LPPM can be described as follows:

$$I(A^s; O) = H(A^s) - H(A^s|O) \quad (11)$$

The minimum value is obtained when  $A^s$  is completely determined by  $O$ . The maximum value is obtained when  $O$  reveals no information about  $A^s$ , i.e. when  $A^s$  and  $O$  are independent.

It measures the amount of information about  $A^s$  that we gain by observing  $O$  after inference attack. Therefore, the effectiveness of the attack can be described in terms of the mutual information  $I(A^s; O)$ .

#### 4. Discussion and Conclusions

Location privacy has been heavily studied in privacy-preserving algorithm such as anonymity and obfuscation, and privacy-attack algorithms that inference the user actual location data and other information. Krumm [6] has made a comprehensive survey. However, there is lack of a common model to describe the location privacy, therefore, we try to propose a location privacy model to specify location privacy. The model describes location privacy from three dimensions, which are specified the three aspects closely related to location privacy respectively. We model a user as a sequence of event that describes who did something at some place for a given time. And we model location privacy protecting mechanism as an information-theoretic channel that gets the actual event as the input, and output the distorted event. We abstract metric for evaluating location privacy protecting mechanism as uncertainty, inaccurate, which can be used in the evaluation module of our proposed location privacy quantifying framework. The basic idea of the framework is that quantifying the information leakage between the actual events and the suspected events which are reconstructed from adversary's perspective.

In literature, the metrics of location privacy are usually mechanism-oriented, such as Gruteser [2] and Duckham [10]'s works. Gruteser uses  $k$  as the metric of location privacy, and Duckham uses the number of different location coordinates sent by a user with a single location-based query as the metric of location privacy.

The common metric for describing the level of location privacy protecting mechanism include the works of Hoh [4] and Diaz et al.[11]. Hoh quantifies location privacy as the expected error in distance between a person's true location and an attacker's estimates of that location, which captures how accurate an adversary can estimate a user's position. Diaz utilizes entropy to measure privacy in anonymous communication systems. In entropy of the random variable that is associated with the users' probabilities is considered as the anonymity level of the system. Our work evaluates location privacy from the information leakage perspective that uses mutual information to describe the difference between the initial uncertainty and remaining uncertainty.



In our framework, the designer of location privacy protecting algorithm can embed their code in the our framework directly, or just provide the distribution of  $P(O|A)$ . The distribution is the basic knowledge for building the HMM. In the future work, we will implement some classic location protecting algorithm proposed in the literature in our framework, and makes comparisons among them.

## Acknowledgment

This research is supported by the National Natural Science Foundation of China (61070186) and the Strategic Priority Research Program of the Chinese Academy of Sciences(XDA06030200)

## References

- [1] A. R. Beresford, F. Stajano, Location Privacy in Pervasive Computing, *IEEE Pervasive Computing* 2 (1) (2003) 46–55.
- [2] M. Gruteser, 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, MobiSys '03*, ACM, New York, USA, 2003, pp. 31–42.
- [3] C. Bettini, X. Wang, S. Jajodia, Protecting Privacy Against Location-Based Personal Identification, in: W. Jonker, M. Petkovic (Eds.), *Secure Data Management*, Vol. 3674 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2005, pp. 185–199.
- [4] B. Hoh, M. Gruteser, Protecting Location Privacy Through Path Confusion, in: *Security and Privacy for Emerging Areas in Communications Networks, 2005.*, 2005, pp. 194–205.
- [5] M. Duckham, L. Kulik, Location privacy and location-aware computing, *Dynamic & mobile GIS: Investigating Change in Space and Time* (2006) 34–51.
- [6] J. Krumm, A survey of computational location privacy, *Personal and Ubiquitous Computing* 13 (6) (2009) 391–399.
- [7] L. Sweeney, Achieving k-anonymity privacy protection using generalization and suppression, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10 (5) (2002) 571–588.
- [8] M. Duckham, L. Kulik, Simulation of obfuscation and negotiation for location privacy, in: *Proceedings of the 2005 international conference on Spatial Information Theory, COSIT'05*, Springer-Verlag, Berlin, Heidelberg, 2005, pp. 31–48.
- [9] L. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE* 77 (2) (1989) 257–286.
- [10] M. Duckham, L. Kulik, A Formal Model of Obfuscation and Negotiation for Location Privacy, in: H. Gellersen, R. Want, A. Schmidt (Eds.), *Pervasive Computing*, Vol. 3468 of *Lecture Notes in Computer Science*, 2005, pp. 243–251.
- [11] C. Díaz, S. Seys, J. Claessens, B. Preneel, Towards measuring anonymity, in: *Proceedings of the 2nd international conference on Privacy enhancing technologies, PET'02*, Springer-Verlag, Berlin, Heidelberg, 2003, pp. 54–68.