Vol.2010-DPS-143 No.18
Vol.2010-MBL-54 No.18
2010/5/20

Protecting Privacy of GPS Traces against Inference Attacks

Kazuhiro Minami †1 and Nikita Borisov†2

Recently, many location tracking services for GPS-enabled mobile phones are available. However, when users share their GPS traces with other users, there is significant concern on user privacy since the locations of users imply their private activities, such as having a secret meeting. In this paper, we argue that traditional access-control schemes that only protect exact private locations of users are not satisfactory since an unauthorized user can predict users' future movements from their previous ones. Our preliminary results with a GPS location predictor based on the Markov model show that such an unauthorized user can infer the target user's visiting a private place with high accuracy.

1. Introduction

Nowadays, most mobile phones are equipped with a GPS capability, and many location-based services (LBSs) has become available on various platforms for mobile phones, such as Symbian, iPhone, and Android. Initially, LBSs start with applications that provide a user with a turn-by-turn navigation on a map or support location-based queries to find the nearest business or service, such as an ATM or restaurant. In those early systems, location information of users is mostly maintained in their mobile phones and is never shared with other users of the systems. However, more recent LBSs[4),5),9),11)] for locating people or objects (e.g., Google latitude[5)]) allow a user to track other users on a map. Therefore, the location of the user could be released to other users continously. The same situation holds with increasingly popular location-based social networking services, such as BrightKite[3)] and Twitter, which announced its support for location sharing[15)] recently; Those services associate each message of the user with his GPS location when it is posted.

†1 National Institute of Informatics
†2 University of Illinois at Urbana-Champaign

LBSs that supports location sharing raise significant concern on location privacy[1)] since users' locations often imply their private activities. For example, visitng a hospital indicates that the user has some medical problem. Or, a co-location of multiple users from different organizations implies a secret business meeting. Therefore, many LBSs for location sharing allow users to define simple access control policies to limit access to their location information. For example, Google Latitude allows a user to hide his location from another user requesting for his location information. Google Latitude also allows a user to manually specify his location, which is different from the actual location. Glympse[4)] allows a user to specify a time period during which the system can release her location to other users. Such simple release policies supported by the current LBSs are too coarse to protect users' location privacy. For example, the current LBSs cannot handle the following situation adequately. Suppose that Alice is willing to disclose her location to Bob, her colleague, when she is in her office, but she is reluctant to do so when she is in a hospital. In this case, a LBS should be able to allow Alice to hide her location in selected private places including the hospital.

This issue has been actually studied the field of pervasive computing, and several researchers[7),8),10),12),13)] proposed more fine-grained access control schemes to allow users to define privacy policies considering their situations such as their current location, the time of day, and so on. In those schemes, users define access-control policies as a set of logical rules where their contextual information is encoded as logical facts, and a system evaluating those policies makes access-control decisions about which locations of the user can be released to another user. We could define access-control policies saying that *Bob is granted to access to Alice's location only if she is in her office.* Such a policy satisfies Alice's privacy requirements in the above example.

However, the previous approaches to location privacy still fail to protect users' privacy if an unauthorized user has external knowledge about the mobility patterns of a target user. The unauthorized user can obtain such knowledge easily by physically observing the movements of the target user or by inferring from the common behaviors of people in a group. For example, every group member of the same project gets together every week for a group meeting. Such an adversary (i.e., an unauthorized user) can predict that a target user is going to visit a pri-

Vol.2010-DPS-143 No.18
Vol.2010-MBL-54 No.18
2010/5/20

vate location if he learns that user's previous movements before visiting the exact private location. Therefore, we need extra mechanisms for preventing inference attacks using previous mobility patters of a user; we should protect not only the exact set of the user's private locations, but also some prior locations on the user's paths directing towards those private locations. Figure 1 shows the actual path of a user visiting a hospital, which the user considers as *private*. If there is an alternate possible path from the intersection in the middle to the library, which is considered as *public*, we can safely disclose the user's movements until the point at the intersection. An unauthorized user cannot determine which of the two locations the user will visit. However, if we disclose any further point on the actual path, an adversary (i.e., an unauthorized user) can figure out that the user is surely visiting the hospital.

We, therefore, propose to develop a new access-control scheme that prevents such inference attacks. Our basic approach is to model an adversary as a location predictor that predicts future movements of a target user from his previous movements with certain probabilities. Intuitively, our access control scheme discloses a user's location information only if an unauthorized user cannot predict that the user moves to some private location with a sufficiently high probability. Our approach is the most conservative in the sense that we assume that an adversary knows all the previous movements of the target user. In this paper, we focus on studying an adversary model based on the Markov model, which is the most successful model for predicting people's location movements[14]. The Morkov model captures the probability distributions of location movements from current locations to next ones in a state transition matrix. Our preliminary results with actual GPS traces of a single user show that we can predict the user's next movement with the accuracy of 60% using a first-order Markov model, and we can improve the accuracy by 10% by considering multiple previous movements with a higher-order Markov model.

The rest of the paper is organized as follows. Section 2 introduces our system model of a LBS considered in this paper, and Section 3 describes a location predictor based on the Markov model. We present our preliminary results of experiments in Section 4. We cover related work in Section 5 and finally states our concluding remarks and future plans in Section 6.

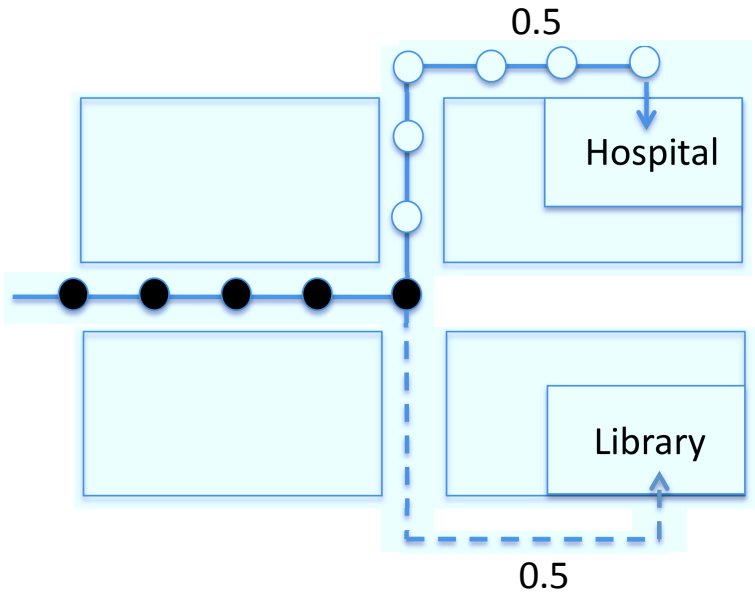

Fig. 1 Example safe disclosure of location information. The solid line represents an actual path of a user visiting a hospital. We assume that the hospital is a private place and the library is a public place. A safe LBS would disclose location points denoted by black nodes. We assume that the user has 50% chance of visiting of the library when he is at the intersection in the middle.

Vol.2010-DPS-143 No.18
Vol.2010-MBL-54 No.18
2010/5/20

2. System model

Figure 2 shows our system model for LBSs. We assume that user p_j is interested in receiving a target user p_i's location movements. User p_i carrying a GPS-enabled mobile device periodically sends LBS a series of location-timestamp pairs (loc_k, t_k) for $k \in \mathcal{N}$; LBS receives a set of all pairs

$$L = \{(loc_k, t_k) \mid k \in \mathcal{N}\}.$$

User p_i also defines its access-control policies in LBS so that LBS can protect p_i's location movements properly. We represent p_i's access-control policies with the function

$$acl : \mathcal{P} \times \mathcal{W} \rightarrow 2^{\mathcal{P}}$$

where $\mathcal{P}$ is a set of all users and $\mathcal{W}$ is a finite set of all locations. The function acl takes a user identity p_i and a location name l_k as inputs and outputs a set of users who are authorized to learn that "p_i is at location l_k." In other words, LBS releases p_i's location movement (l_k, t_k) to principal p_j only if p_j belongs to set $acl(p_i, l_k)$, and thus user p_j receives a subset of events $L' \subseteq L$

$$L' = \{(loc_k, t_k) \mid p_j \in acl(p_i, l_k)\}.$$

Notice that we only consider the case that p_i's access-control policies depend on p_i's location l_k to simplify our discussion in this paper, but we can easily support the general case where access-control policies also considers a timestamp t_k.

We next define which locations are *private* to user p_i formally.

Definition 1 (Private location.) We consider that a user p_i's location l is private with respect to another user p_j if:

$$l \in \{l' \mid p_j \notin acl(p_i, l')\}.$$

We consider that a LBS preserves a user p_i's privacy if p_j cannot infer that p_i was at some private location l from the information p_j receives from LBS. We formalize this concept below.

Definition 2 (Preservation of location privacy.) We say that a LBS preserves a user p_i's location privacy against another user p_j if p_j cannot infer p_i's movement (l, t) where l is p_i's private location from a set of location-timestamp pairs L'.

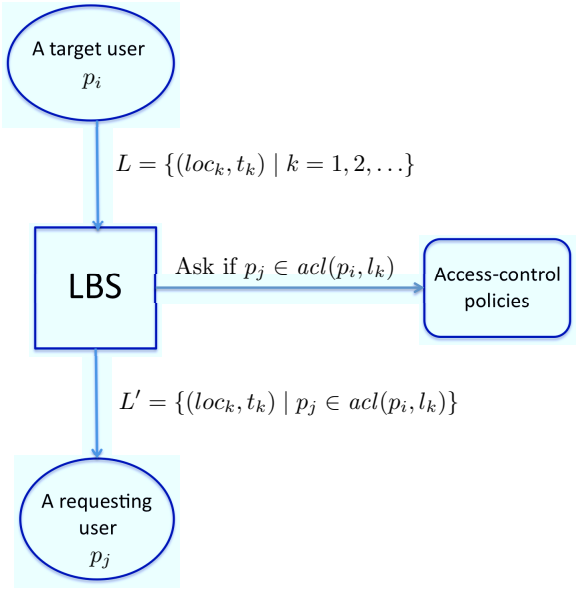

Fig. 2 System model.

In next section, we describe how an unauthorized user p_j performs inference with a location predictor based on the Markov model.

3. Location predictor based on the Markov model

We consider a Markov chain with a sequence of random variables

$$X_1, X_2, X_3, \ldots$$

where each X_i has a value drawn from the finite set of locations $\mathcal{W}$. We here assume that location l_k is published periodically, and we thus omit timestamp t_k in tuple (l_k, t_k). We also assume that the Markov chain is time-homogeneous. So, if we consider a Markov chain of order 1,

$$Pr(X_{n+1} = l_i | X_n = l_j) = Pr(X_n = l_i | X_{n-1} = l_j).$$

We maintain the probability of moving from location l_i to l_j in (i, j)th element

of a state transition matrix $M_{i,j}$ as follows:

$$Pr(X_{n+1} = l_i | X_n = l_j) = M_{i,j}.$$

for every pair of l_i and l_j in set $\mathcal{W}$. The probability of moving from location l_i to l_j in n time steps can be computable by multiplying the transition matrix M n times as follows:

$$Pr(X_n = l_i | X_0 = l_j) = M^{(n)}_{i,j}.$$

Since it is likely that we can improve the accuracy of location predictions by considering multiple previous movements, we also consider a location predictor based on a Markov model of a higher order. If we use a Markov model of 2 order, a state transition matrix M must maintain the probability $Pr(X_{n+1} = l_i | X_n = l_j, X_{n-1} = l_k)$ in $((j,k),i)$th element of M; that is,

$$Pr(X_{n+1} = l_i | X_n = l_j, X_{n-1} = l_k) = M_{(j,k),i}.$$

We make the most conservative assumption that an adversary can observe all the previous movements of a target user and compute a state transition matrix M of an arbitrary order n before predicting the target user's next movement. We now define the preservation of location privacy against an adversary with a state transition matrix M of the 1-order Markov model as follows:

Definition 3 (Preservation of (M,t)-location privacy.) Suppose that a user p_i' current location is l_i and that t is a probability threshold where $0 \leq t \leq 1$. We say that a LBS preserves a user p_i's (M,p)-location privacy against another user p_j if, for every private location $l_k \in \mathcal{W}$ with respect to p_j, the following condition holds

$$M^{(n)}_{i,k} \leq t \text{ for } n = 1, 2, \ldots.$$

Intuitively speaking, the above definition requires that an unauthorized user p_j cannot predict that the target user p_i is at some private location l_k in some future time with probability p, which is greater than the threshold value t. Although the above definition only covers the case with the 1-order Markov model, we can easily generalize the definition to consider a Markov model of order n.

4. Experimental results

We conducted experiments with actual GPS traces to study how accurately we can predict future location movements using location predictors based on the Markov model. One of the authors collected GPS traces by carrying a GPS device for fifty days. We consider GPS data whose coordinates resides within a rectangular region, which covers the campus of University of Illinois and its surrounding off-campus areas. The dimension of the region is 4.8 kilometers times 4.0 kilometers. We divide each coordinate into 40 units and define 1,600 different locations, each of which has about the size of a building in town.

We used half of the data to construct a state transition matrix M and used the other half to compute the accuracy of the predictions with matrix M. Figure 3 shows our experimental results. The X-axis shows how many steps we predict ahead, and the Y-axis shows the accuracy of our predictions. We computed the accuracy of predicting every next location and took its average. When we predict a next location in a single time step with a 1-order Markov model, our predictions are about 60% accurate. However, as we try to predict a location reachable in greater number of steps, the prediction accuracy decreases. We compare the results of Markov models of three different orders. As we can see, when we predict locations reachable in a fewer number of time steps, we can improve the accuracy by 10% by using a higher-order Markov model, which considers multiple previous movements. However, when we predict a location multiple steps ahead, using a higher-order model is not useful. We speculate that the accuracy degrades because earlier movements are not relevant to final destinations. However, we need further investigation regarding this issue.

We believe that the accuracy of predictions could be improved significantly if we predict the eventual final destinations because we can identify a small number of stationary locations in the GPS traces. Since we currently try to predict a location in n steps, the accuracy of the predictions was not as good as we initially expected.

5. Related work

Several researchers[7),8),10),12),13)] propose rule-based access-control schemes for

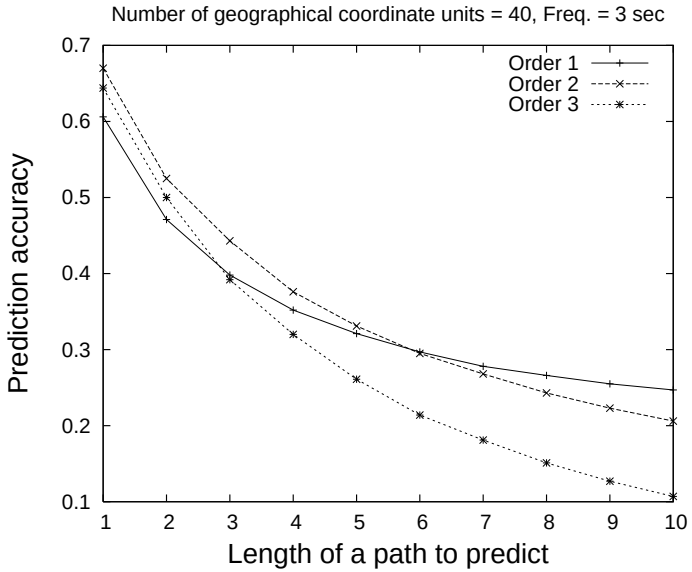

Fig. 3 Accuracy of location predictions.

protecting user location in pervasive environments. Hengartner[7] supports access-control policies considering the granularity of location information and time intervals. Myles[12] provides a XML-based authorization language for defining privacy policies that protect users location information. Users must trust a set of validators that collect context information and make authorization decisions. Those schemes allows a user to define fine-grained access-control policies. Apu[10] provides users with an intuitive way of defining access control policies, which represent physical boundaries surrounding the users. However, no previous scheme considers the issue of inference based on the mobility patterns of users.

Location privacy has been studied heavily in the context of location data anonymization[2),6)]. The focus of research in this sequence is to ensure that no anonymized data is associated with an individual. For example, Gruteser[6] proposes a scheme that changes the granularity of location information to ensure that each location contains at least k users (i.e., k-anonymity). However, the problem addressed in this paper is different since we consider inference on location data associated with a known individual.

6. Conclusion and future work

In this paper, we address a new issue of inference attacks on GPS traces of mobile users and show the shortcomings of traditional access-control schemes that only protect exact private locations of users. We formally define such an adversary with a location predictor based on the Markov model, and introduce new privacy requirements under the presence of such an adversary. Our preliminary experimental results show that it is possible to predict a mobile user's location with high accuracy. However, we may need to extend our security model to address an adversary who tries to predict eventual user destinations.

We plan to conduct more extensive experiments involving many mobile users. We also consider more general location predictors that consider other parameters, such as the time of day.

References

1) Anthony, D., Henderson, T. and Kotz, D.: Privacy in Location-Aware Computing Environments, *IEEE Pervasive Computing*, Vol.6, No.4, pp.64–72 (2007).
2) Beresford, A.R. and Stajano, F.: Location Privacy in Pervasive Computing, Vol.2, No.1, pp.46–55 (2003).
3) : BrightKite, http://www.brightkite.com.
4) : Glympse, http://www.glympse.com.
5) : Google latitude, http://www.google.com/latitude.
6) Gruteser, M. and Grunwald, D.: Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking, *Proceedings of Mobisys 2003: The First International Conference on Mobile Systems, Applications, and Services*, San Francisco, CA, USENIX Associations (2003).
7) Hengartner, U. and Steenkiste, P.: Access control to people location information, *ACM Transactions on Information and System Security (TISSEC)*, Vol.8, No.4, pp.424–456 (2005).
8) Hong, J.I. and Landay, J.A.: An architecture for privacy-sensitive ubiquitous computing, *Proceedings of the 2nd international conference on Mobile systems, applications, and services (MobiSys)*, New York, NY, USA, ACM, pp.177–189 (2004).
9) : InstaMapper, http://www.instamapper.com.
10) Kapadia, A., Henderson, T., Fielding, J.J. and Kotz, D.: Virtual Walls: Protecting Digital Privacy in Pervasive Environments, *Proceedings of the Fifth International Conference on Pervasive Computing (Pervasive)*, LNCS, Vol.4480, Springer-Verlag,

Vol.2010-DPS-143 No.18
Vol.2010-MBL-54 No.18
2010/5/20

pp.162–179 (2007).
11) : Loopt, http://www.loopt.com.
12) Myles, G., Friday, A. and Davies, N.: Preserving Privacy in Environments with Location-Based Applications, *IEEE Pervasive Computing*, Vol.2, No.1, pp.56–64 (2003).
13) Sacramento, V., Endler, M. and de Souza, C.: A privacy service for location-based collaboration among mobile users, *Journal of the Brazilian Computer Society*, Vol.14, No.4, pp.41–57 (2008).
14) Song, L., Kotz, D., Jain, R. and He, X.: Evaluating next cell predictors with extensive Wi-Fi mobility data, *IEEE Transactions on Mobile Computing*, Vol.5, No.12, pp.1633–1649 (2006).
15) : How to Tweet With Your Location, http://twitter.zendesk.com/entries/122236.