ILLIA: Enabling k-Anonymity-Based Privacy Preserving Against Location Injection Attacks in Continuous LBS Queries

Ping Zhao, Jie Li, Fanzi Zeng, Fu Xiao, Chen Wang, *Member, IEEE*, and Hongbo Jiang, *Senior Member, IEEE*

***Abstract*—With the increasing popularity of location-based services (LBSs), it is of paramount importance to preserve one's location privacy. The commonly used location privacy preserving approach, location k-anonymity, strives to aggregate the queries of k nearby users within a so-called cloaked region via a trusted third-party anonymizer. As such, the probability to identify the location of every user involved is no more than $1/k$, thus offering privacy preservation for users. One inherent limitation of k-anonymity, however, is that all users involved are assumed to be trusted and report their real locations. When location injection attacks (LIAs) are conducted, where the untrusted users inject fake locations (along with fake queries) to the anonymizer, the probability of disclosing one's location privacy could be greatly more than $1/k$, yielding a much higher risk of privacy leakage. To tackle this problem, in this paper we present ILLIA, the first work that enables k-anonymity-based privacy preservation against LIA in continuous LBS queries. Central to the ILLIA idea is to explore the pattern of the users' mobility in continuous LBS queries. With a thorough understanding of the users' mobility similarity, a credibility-based k-anonymity scheme is developed, such that ILLIA is able to defense against LIA without requiring in advance knowledge of how fake locations are manipulated while still maintaining high quality of services. Both the effectiveness and the efficiency of ILLIA are validated by extensive simulations on real world dataset loc-Gowalla.**

***Index Terms*—Continuous location-based service (LBS) query, location injection attack (LIA), location k-anonymity, location privacy, mobility similarity.**

Manuscript received August 25, 2017; revised November 20, 2017 and December 16, 2017; accepted January 25, 2018. Date of publication January 30, 2018; date of current version April 10, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61502192, Grant 61572219, Grant 61671216, Grant 61471408, and Grant 41701479, in part by the China Post-Doctoral Science Foundation under Grant 2017T100556, and in part by the Fundamental Research Funds for the Central Universities under Grant 2016YXMS297 and Grant 2016JCTD118. *(Corresponding author: Chen Wang.)*

P. Zhao and C. Wang are with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: pingzhao2014ph@gmail.com; cwangwhu@gmail.com).

J. Li and F. Zeng are with the College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China (e-mail: jieli.csee@gmail.com; zengfanzi@hnu.edu.cn).

F. Xiao is with the College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210013, China, and also with the Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, Nanjing 210042, China (e-mail: xiaof@njupt.edu.cn).

H. Jiang is with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China, and also with the College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China (e-mail: hongbojiang2004@gmail.com).

Digital Object Identifier 10.1109/JIOT.2018.2799545

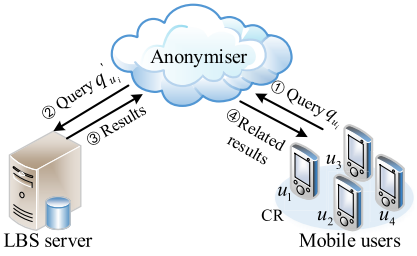


Fig. 1. k-anonymity-based privacy preserving system.

I. Introduction

RECENT years have witnessed an ever-growing number of mobile users with hand-held devices requesting location-based services (LBSs) [1]–[3] (e.g., finding nearby restaurants, or monitoring real-time traffic, etc.) by sending queries to the LBS server. Unfortunately, it is found that, the LBS server may either deliberately or inadvertently disclose the location information involved in queries [4]–[6], which can be used to infer such sensitive personal information as religious activities and health/living habit. According to [7], there are 15 of the 30 investigated Apps, e.g., MySpace and Trapster, disclosing users' locations to advertisement or analytic servers. More severely, the black hat "Peace" disclosed over 167 million users' information (including location information) from LinkedIn in 2013 [8], and about 360 million users' information in MySpace is disclosed in 2016 [9]. In such grim situations, how to preserve users' location privacy has been a hot research topic for years [2], [10]–[12].

Location k-anonymity first introduced by Gruteser and Grunwald [13] is a commonly used approach to protect users' location privacy in LBS, which employs a trusted third-party (dubbed as *anonymizer* [14]–[16]). In a location k-anonymity mechanism, as shown in Fig. 1, a specific user (say u_1) first sends LBS query q_{u_1} to the anonymizer. Then, the anonymizer strives to aggregate the queries of $(k-1)$ users around u_1 (say u_2, u_3, and u_4) and u_1's query q_{u_1} within a so-called cloaked region (CR), and sends the cloaked query q'_{u_i} $(i \in (1, 2, 3))$ to the LBS server. As a result, the user's location cannot be distinguished from $(k-1)$ locations of other users by the LBS server [13], [15], [17].

Fig. 2(a) shows an example of the k-anonymity-based privacy preserving system used in continuous LBS queries, where

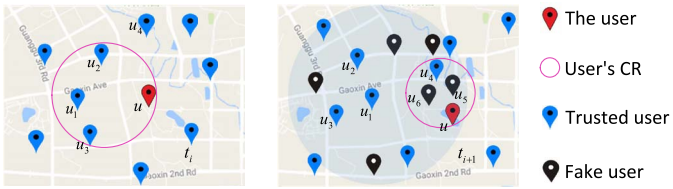


Fig. 2. 4-Anonymity (a) without LIA and (b) with LIA in continuous LBS queries.

all users send requests to the anonymizer. The anonymizer purposely cloaks the locations of users u_1, u_2, u_3 and the user of interest u (a possible victim) in the CR R_{u,t_i} at time t_i (see the purple circle), and the cloaked queries are sent to the LBS server afterward. As a result, the LBS server cannot distinguish those four locations, and the location privacy preservation of u can be achieved.

A fundamental limitation of all existing techniques complying with location k-anonymity [14]–[19], however, is that all users [e.g., u_1, u_2, and u_3 in Fig. 2(a)] are assumed to be inherently trusted and report their real locations. If this assumption no longer holds, the location privacy may be faced with some new threatens as discussed below.

A. LIA in Continuous LBS Queries

In continuous LBS queries, suppose an attacker wants to disclose the user of interest u's location at time t_{i+1}. The attacker first obtains u's CR R_{u,t_i} at t_i [see Fig. 2(a)] from the untrusted LBS server. Then the attacker infers u's district at time t_{i+1} by analyzing R_{u,t_i} and the corresponding maximum moving speed obtained from the speed limit of the road [20] or statistical data [21]. That is, by doing so, the attacker gets to know that the user u should be at some place within the district [the blue transparent area shown in Fig. 2(b)] at time t_{i+1}. Accordingly, at time t_{i+1}, the attacker deliberately creates (e.g., utilizing smartphone applications, such as FakeGPSTracker [22]) many fake users[1] [say u_5 and u_6 in Fig. 2(b)] in u's district, and reports/injects these fake locations (along with fake LBS queries) to the anonymizer. Then the CR $R_{u,t_{i+1}}$ of user u at time t_{i+1} is generated by the anonymizer, as shown in Fig. 2(b) (the purple circle).

As a result, the user of interest u that relies on k-anonymity mechanism for privacy preservation could become a victim. To be more concrete, when there is no fake location, the probability to infer the location of u in the 4-anonymity [see Fig. 2(a)] is $1/4$ at time t_i. In contrast, the fake locations of u_5 and u_6 are injected in the CR $R_{u,t_{i+1}}$ at time t_{i+1} [see Fig. 2(b)], and thus the probability to distinguish the location of u increases from $1/4$ to $1/2$ (since the attacker has the knowledge that u_5 and u_6 are fake users), indicating a higher risk of the location disclosure of u in continuous LBS queries. This violation of user's location privacy in terms of in-distinguishability among a set of users is the so-called location injection attacks (LIAs) [23].

[1]Hereafter, fake users refer to both the users generated by the attacker and the untrusted users.

B. Existing Work Related to LIA

So far, little progress has been made along the line of privacy preservation against LIA—most existing work are vulnerable to LIA. Jin *et al.* [23] presented the first work that characterizes LIA and demonstrates LIA' effectiveness (on disclosing users' location privacy). Unfortunately, they do not provide a corresponding solution to protect users' location privacy against LIA.

A straightforward method against LIA could be enlarging the privacy parameter k of a specific user when k-anonymity is performed. Intuitively, a larger k can lead to a stronger defense, but it is more likely to result in a larger CR, which will give rise to higher overhead and more inaccurate querying results. More seriously, enlarging k will bring about a significantly lower success cloaking rate, and thus cannot fully protect users' location privacy against LIA. As will be shown in Section IV, it has been verified via simulations that when k is enlarged from 6 to 12, the success cloaking rate can be reduced by roughly 39.4%, while almost 63% users are still suffered from LIA.

C. Our Approach

In this paper, we propose ILLIA, the first work that enables location k-anonymity-based privacy preservation against LIA in continuous LBS queries. The intuition of ILLIA stems from the observation that the attackers tend to attack some specific users they are interested in (we refer to those users as *high-risk users*). Despite the way of manipulating the fake locations that the attacker follows, the attacker can disclose the location privacy of the high-risk users only when some injected fake locations are cloaked in the same CR with the locations of high-risk users.

Therefore, our main idea is to explore users' mobility patterns, and look deeper into the mobility similarity between high-risk users and fake users, so that the suspected locations (i.e., fake locations) can be identified. Though the basic idea is simple, the particular problem we are facing, however, is quite challenging.

1) Traditional mobility models [24] based on the exact locations are not practical in this paper, as the attacker only has the knowledge of high-risk users' CRs and conducts LIA based upon CRs rather than the exact locations of high-risk users. To address this problem, we propose a *CR-based mobility model* that models users' mobility trajectories in terms of CRs instead of exact locations. This new model can thus provide the opportunity of analyzing all the users' (including high-risk users and fake users) mobility patterns.
2) It is nontrivial to identify real locations and fake locations, as the spatial distribution of the fake locations is often of indeterminacy [23], [25]. To tackle this challenge, we deduce from the CR-based mobility model a novel metric, the *mobility similarity*, which characterizes the transition probability and the distribution of users' locations, to distinguish real locations and fake locations.

3) It is difficult to balance the quality of service (QoS) and the location privacy preservation [15], [17]. Enhancing the location privacy preservation will lead to less users being cloaked, larger CRs, as well as higher computation cost, and thus decrease the QoS. To that end, we develop a *credibility-based k-cloaking algorithm*. By giving priority to cloaking users with higher credibility, we can achieve a better tradeoff between the QoS and the privacy preservation for all users.

To the best of our knowledge, ILLIA conducts the first work toward location privacy preservation against LIA in continuous LBS queries. It also offers several salient features. First, it is a general solution against LIA, without requiring in advance knowledge of how fake locations are manipulated. Second, it is scalable, since it is robust to the number of users and lightweight in terms of the processing time. What is more, ILLIA is able to maintain a better tradeoff between QoS and location privacy preservation. Extensive simulations on real world dataset loc-Gowalla have validated both the effectiveness and the efficiency of ILLIA.

The remainder of this paper is organized as follows. Section II describes the system model. Section III describes ILLIA in detail. Section IV presents the evaluation results, and finally Section V concludes this paper.

II. Preliminary

A. System Model

Fig. 3 shows our system model, which is composed of three entities: 1) the users; 2) the anonymizer; and 3) the LBS server.

1) Users: A user sends to the anonymizer the LBS query $q_1 = \{id_1, (x_1, y_1), (k_1, A_{\min,u_1}, \xi_{u_1}), C_1\}$, where id_1, (x_1, y_1) and C_1 are user's identity, location, and query content, respectively. $(k_1, A_{\min,u_1}, \xi_{u_1})$ are privacy parameters, where k_1 is the privacy requirement that u_1's query should be cloaked with no less than $(k_1 - 1)$ other users' queries; A_{u_1} is the CR's minimum area; and ξ_{u_1} is the threshold of the mobility similarity. Here, the communication between the user and the anonymizer is assumed to be secure [14]–[16].

2) Anonymizer: The anonymizer is assumed trusted [14]–[16]. When receiving the LBS query q_1 from the user, the anonymizer first searches for at least $(k_1 - 1)$ users (e.g., $u_2, u_3, \ldots, u_k$), cloaks them in a CR (e.g., CR) by executing the ILLIA algorithm, and then sends the cloaked LBS queries $q' = \{[id'_1, \ldots, id'_k], [(x_1, y_1), C_1], \ldots, [(x_k, y_k), C_k], \text{CR}\}$ to the LBS server. Here, id'_j is the dummy ID of u_j assigned by anonymizer ($j \in (1, \ldots, k)$). Note that the anonymizer keeps records of η historical queries and CRs for each user to model the user's mobility.

3) LBS Server: The LBS server is assumed untrusted [14]–[16]. Upon receiving the cloaked query q' from the anonymizer, the LBS server seeks for query results (e.g., restaurants, shopping malls, etc.), and sends them to the anonymizer. Afterwards, the anonymizer sends the related results back to the corresponding users (e.g., C_1 to u_1). Note that the LBS server plays almost the same role as in conventional *k*-anonymity techniques, and thus is less involved in this paper.

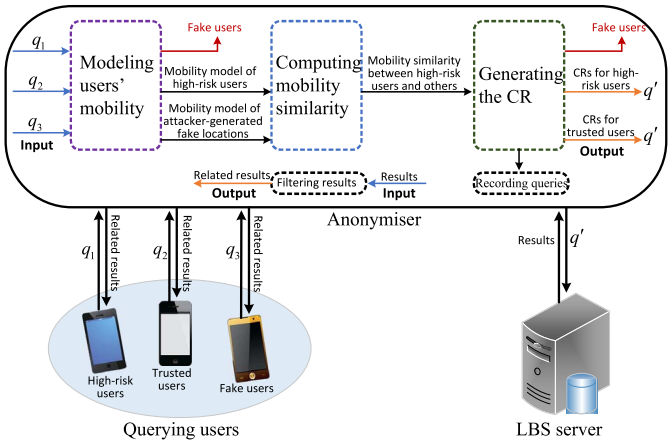


Fig. 3. System architecture of ILLIA.

B. LIA Model

The attacker first gets access to the historical CRs of high-risk users, as the untrusted LBS server may deliberately disclose the CRs to the attacker, or be hacked by the attacker [4], [5]. Then the attacker intelligently manipulates a number of fake locations and launches LIA by sending the fake locations along with LBS queries to the anonymizer, aiming to identify the exact locations of the high-risk users.

Despite the way of manipulating the fake locations that the attacker follows, the trajectories of fake users consist of some plausible trajectory fragments that exhibit similar mobility pattern as a user [e.g., meeting the constraints of the maximum moving speed, bypassing the traffic jams, and seeking for the shortest path (hereafter physical factors), and exhibiting consistent lifestyle], and some obviously fake trajectory fragments.

Note that the anonymizer may assign different pseudonym IDs for different queries of a specific user to prevent the attacker from correlating his consecutive CRs. What needs to be stressed here is that, even unaware of the IDs, the attacker can still identify the user's CRs by, e.g., employing multitarget tracking [26] or clustering techniques on the basis of one's living habits or spatial distributions of locations [27].

III. ILLIA System

The intuition of ILLIA is to identify fake users (also the corresponding fake locations) on the basis of some specific users they are interested in (i.e., the aforementioned high-risk users, e.g., users always requesting Western-style food from the LBS server are more likely to be breached, receiving precision marketing about European cuisine). However, seeking the high-risk users is out of the scope of this paper. We refer interested readers to the existing techniques [28], [29] that detect users suffering from attacks and can be used in conjunction with our proposed approach. Overall, ILLIA mainly involves the following three steps to preserve location privacy against LIA, which is shown in Fig. 3.

1) *Modeling Users' Mobility:* To credibly imitate the mobility of users in LIA, we propose a CR-based mobility model, by modeling users' mobility trajectories based on CRs instead of exact locations.
2) *Computing Mobility Similarity:* Given users' mobility model, we propose the metric—mobility similarity, to grasp the correlation between the high-risk users' mobility and other users's mobility.
3) *Generating the CR:* Based on the mobility similarity, we assign each user a credibility value. Then we propose a credibility-based k-cloaking algorithm, which cloaks the high-risk user with the credible users, and cloaks other users scoring approximation credibility in the same CR so that all users' location privacy can be protected.

A. Mobility Model

We model users' mobility as a time-dependent first-order Markov chain on the set of locations. In his sense, the mobility profile $\langle P_u, L_u\rangle$ for a given high-risk user u (or $\langle P_a, L_a\rangle$ for a fake user) is a transition probability matrix of the Markov chain that characterizes the user's visiting probability distribution over his locations. Note that all users except for high-risk users, i.e., trusted users and fake users, are regarded as fake users at first. Then we distinguish trusted users and fake users with the help of the mobility similarity defined based upon our proposed mobility models.

1) Mobility Model of High-Risk Users: Intuitively, we should model the mobility of high-risk users in the perspective of the attacker, in order to accurately grasp the relationship between locations of high-risk users and fake users. Since the attacker only has prior knowledge of the previous CRs and can infer the maximum movement boundaries (MMBs)[2] of high-risk users, we propose the following CR-based mobility model for high-risk users.

Denote L_u locations of a high-risk user u, and T_u the time when u issues LBS queries. The possibility that the location of u is $l_{u,i}$ ($i \in (1, 2, \dots)$ at time t_i is denoted by $P_{l_{u,i},t_i}$. The possibility that u locates at $l_{u,i}$ at time t_i is $P^{l_{u,i},t_i}_{l_{u,i-1},t_{i-1}}$, given the location $l_{u,i-1}$ at time t_{i-1}. Then we can get

$$P_{l_{u,i},t_i} = P\{L_u = l_{u,i}, T_u = t_i\} \tag{1}$$

$$P^{l_{u,i},t_i}_{l_{u,i-1},t_{i-1}} = P\{L_u = l_{u,i}, T_u = t_i / L_u = l_{u,i-1}, T_u = t_{i-1}\}. \tag{2}$$

Similarly, given $L_u = l_{u,i}$ at time t_i, the possibility that the location $L_u = l_{u,i-1}$ at time t_{i-1} is

$$P^{l_{u,i-1},t_{i-1}}_{l_{u,i},t_i} = \frac{P\{L_u = l_{u,i}, T_u = t_i; L_u = l_{u,i-1}, T_u = t_{i-1}\}}{P_{l_{u,i},t_i}}. \tag{3}$$

Next, we focus on computing the transition probability $P^{l_{u,i},t_i}_{l_{u,i-1},t_{i-1}}$, $P^{l_{u,i-1},t_{i-1}}_{l_{u,i},t_i}$, and $P_{l_{u,i},t_i}$. We first present the following results.

Lemma 1: Suppose the last location is $l_{u,i-1}$ at time t_{i-1}, the possibility that the current location is $l_{u,i}$ at time t_i is $(1//S_{\mathrm{MMB}_{i-1,i}})$, where $S_{\mathrm{MMB}_{i-1,i}}$ is the area of the MMB $\mathrm{MMB}_{i-1,i}$. Conversely, given current location, the possibility that the last location is $l_{u,i-1}$ at time t_{i-1} is given by $(1/S_{R_{u,t_{i-1}}})$, where $S_{R_{u,t_{i-1}}}$ is the area of the CR $R_{u,t_{i-1}}$ at time t_{i-1}.

Proof: When the high-risk user u issues LBS query at time t_i, u must be in the MMB $\mathrm{MMB}_{i-1,i}$, as shown in Fig. 4(a). That is to say, u locates at any point in $\mathrm{MMB}_{i-1,i}$ with the equal probability in the view of attackers. Thus,

$$P^{l_{u,i},t_i}_{l_{u,i-1},t_{i-1}} = \frac{1}{S_{\mathrm{MMB}_{i-1,i}}}. \tag{4}$$

Given the location of u at time t_i, u must locate in the maximum arrival boundary (MAB)[3] $\mathrm{MAB}_{i-1,i}$ at time t_{i-1}, as shown in Fig. 4(b). Furthermore, u must locate in the CR $R_{u,t_{i-1}}$ at time t_{i-1}, since attackers have a knowledge of the CR $R_{u,t_{i-1}}$ at time t_{i-1}. So,

$$P^{l_{u,i-1},t_{i-1}}_{l_{u,i},t_i} = \frac{1}{S_{R_{u,t_{i-1}}}}. \tag{5}$$

To sum up, Lemma 1 holds. ■

According to Lemma 1, we get the following results.

Theorem 1: The possibility that the high-risk user locates at $l_{u,i}$ at time t_i is given by $(S_{R_{u,t_1}}/S_{\mathrm{MMB}_{1,2}}) \cdots (S_{R_{u,t_{i-1}}}/S_{\mathrm{MMB}_{i-1,i}})(1/S_{R_{u,t_1}})$.

Proof: Combining with (2) and (3), we can get

$$P^{l_{u,i},t_i}_{l_{u,i-1},t_{i-1}} P_{l_{u,i-1},t_{i-1}} = P^{l_{u,i-1},t_{i-1}}_{l_{u,i},t_i} P_{l_{u,i},t_i}. \tag{6}$$

Combining Lemma 1 and (6), we get

$$\frac{1}{S_{\mathrm{MMB}_{i-1,i}}} P_{l_{u,i-1},t_{i-1}} = \frac{1}{S_{R_{u,t_{i-1}}}} P_{l_{u,i},t_i}. \tag{7}$$

Then we can deduce

$$\begin{cases} \frac{1}{S_{\mathrm{MMB}_{1,2}}} P_{l_{u,1},t_1} = & \frac{1}{S_{R_{u,t_1}}} P_{l_{u,2},t_2} \\ \frac{1}{S_{\mathrm{MMB}_{2,3}}} P_{l_{u,2},t_2} = & \frac{1}{S_{R_{u,t_2}}} P_{l_{u,3},t_3} \\ \qquad\qquad\vdots & \\ \frac{1}{S_{\mathrm{MMB}_{i-1,i}}} P_{l_{u,i-1},t_{i-1}} = & \frac{1}{S_{R_{u,t_{i-1}}}} P_{l_{u,i},t_i}. \end{cases} \tag{8}$$

Therefore, the possibility that u locates at $l_{u,i}$ at time t_i is

$$P_{l_{u,i},t_i} = \frac{S_{R_{u,t_1}}}{S_{\mathrm{MMB}_{1,2}}} \frac{S_{R_{u,t_2}}}{S_{\mathrm{MMB}_{2,3}}} \cdots \frac{S_{R_{u,t_{i-1}}}}{S_{\mathrm{MMB}_{i-1,i}}} P_{l_{u,1},t_1}. \tag{9}$$

Since the CR R_{u,t_1} is known to attackers, the possibility $P_{l_{u,1},t_1}$ that $L_u = l_{u,1}$ when u issues LBS query at first time t_1 is $P_{l_{u,1},t_1} = \frac{1}{S_{R_{u,t_1}}}$. Therefore, we get

$$P_{l_{u,i},t_i} = \frac{S_{R_{u,t_1}}}{S_{\mathrm{MMB}_{1,2}}} \frac{S_{R_{u,t_2}}}{S_{\mathrm{MMB}_{2,3}}} \cdots \frac{S_{R_{u,t_{i-1}}}}{S_{\mathrm{MMB}_{i-1,i}}} \frac{1}{S_{R_{u,t_1}}}. \tag{10}$$

In summary, Theorem 1 holds. ■

It can be observed from (10) that a larger i results in less possibility that u locates at $l_{u,i}$ at time t_i, since $S_{R_{u,t_{i-1}}}$ is obviously less than $S_{\mathrm{MMB}_{i-1,i}}$. This is in accordance with the fact that the attacker is not aware of the exact location of u whenever u issues LBS query, as he only has the knowledge of previous CRs of u.

[2]MMB is a circle that extends the CR by a radius of $v_{\max}\Delta t$, where $v_{\max}$ is the maximum moving speed. MMB $\mathrm{MMB}_{i-1,i}$ is a circle that extends the CR at time t_{i-1} by a radius of $v_{\max}(t_i - t_{i-1})$.

[3]MAB is a circle that extends the CR by a radius of $v_{\max}\Delta t$. MAB $\mathrm{MAB}_{i-1,i}$ is a circle that extends the CR at t_i by a radius of $(t_i - t_{i-1})$.

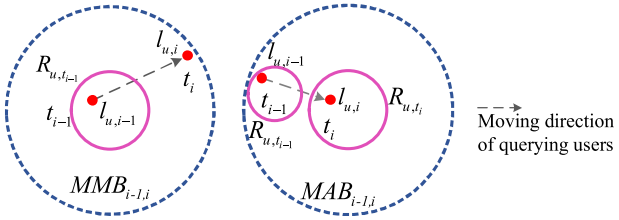


Fig. 4. Illustration of computing the transition possibility for high-risk users. (a) The probability that u locates in $\text{MMB}_{i-1,i}$, given the CR at t_{i-1}. (b) The probability that u locates in $\text{MAB}_{i-1,i}$, given the CR at t_i. (c) Legend.

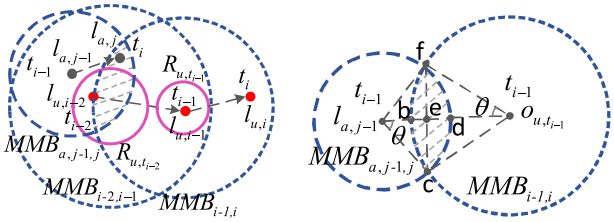


Fig. 5. Illustration of computing the transition possibility for attackers. (a) The effect of both $\text{MMB}_{i-1,i}$ and $\text{MMB}_{a,j-1,j}$ on the mobility of attackers. (b) The geometrically formalization of the effect of $\text{MMB}_{i-1,i}$ and $\text{MMB}_{a,j-1,j}$.

2) Mobility Model of Attacker-Generated Fake Locations: Despite the way of manipulating the fake locations that the attacker follows, the trajectories of fake users consist of some plausible trajectory fragments that deliberately imitate a user's mobility pattern (e.g., meeting the constraints of physical factors and exhibiting consistent life style), and some obviously fake trajectory fragments. Fortunately, it has been proved in the existing work [14], [24], [30] that the fake trajectory fragments are vulnerable to the location inference attacks that can easily filter out fake trajectories. In addition, as for the plausible trajectories, to breach the location privacy of high-risk users, the fake locations in plausible trajectories should be cloaked in the same CR with the locations of high-risk users. As a result, previous CRs of high-risk users have influence on the plausible trajectories of fake users. In summary, we only need to model the plausible trajectories, considering the previous CRs of high-risk users and physical factors, which is introduced in detail as follows.

Denote $P_{l_{a,j},t_i}$ the possibility that $L_a = l_{a,j}$ ($j \in (1, 2, \ldots)$) at time t_i. The possibility that $L_a = l_{a,j}$ at time t_i is $P^{l_{a,j},t_i}_{l_{a,j-1},l_{u,i-1},t_{i-1}}$, given $L_a = l_{a,j-1}$ and $L_u = l_{u,i-1}$ at time t_{i-1}. Then we can get

$$P_{l_{a,j},t_i} = P\{L_a = l_{a,j}, T_a = t_i\} \tag{11}$$

$$P^{l_{a,j},t_i}_{l_{a,j-1},l_{u,i-1},t_{i-1}} = P\{L_a = l_{a,j}, T_a = t_i / L_a = l_{a,j-1}, T_a = t_{i-1} / L_u = l_{u,i-1}, T_u = t_{i-1}\}. \tag{12}$$

Thereafter, we concentrate on computing the transition probability $P^{l_{a,j},t_i}_{l_{a,j-1},l_{u,i-1},t_{i-1}}$.

Lemma 2: The attacker can launch LIA to a high-risk user, iff he is in the overlapped region of his MMB[4] and the MMB of the high-risk user.

Proof: On one hand, the attacker must be in his MMB to satisfy the maximum moving speed constraint. On the other hand, the attacker must be in the MMB of the high-risk user so that he can be cloaked with the high-risk user. In summary, Lemma 2 holds. ■

[4]The attacker's MMB $\text{MMB}_{a,j-1,j}$ is a $v_{\max}(t_i - t_{i-1})$-radius circle that centers at his fake location $l_{a,j-1}$.

Theorem 2: The possibility for an attacker to launch LIA is shown in (13) at the bottom of this page, where $\gamma_1 = r^2_{u,t_{i-1}} - r^2_{a,t_{i-1}}$, $\gamma_2 = |l_{a,j-1}o_{u,t_{i-1}}|^2$, $o_{u,t_{i-1}}$ is the center of u' CR $R_{u,t_{i-1}}$ at time t_{i-1}, $|.|$ represents the Euclidean distance of two points, and $r_{u,t_{i-1}}$ and $r_{a,t_{i-1}}$ are the radius of $\text{MMB}_{i-1,i}$ and $\text{MMB}_{a,j-1,j}$, respectively.

Proof: According to Lemma 2, to conduct LIA, the fake location should locates in the overlapped region of $\text{MMB}_{a,j-1,j}$ and $\text{MMB}_{i-1,i}$, namely the blurred region, as shown in Fig. 5(a). Thus,

$$P^{l_{a,j},t_i}_{l_{a,j-1},l_{u,i-1},t_{i-1}} = \frac{1}{S_{\text{MMB}_{j-1,j,i-1,i}}} \tag{14}$$

where $S_{\text{MMB}_{j-1,j,i-1,i}}$ is the area of the blurred region.

According to Fig. 5(b), the area of sector $fbco_{u,t_{i-1}}$ and triangle $\triangle o_{u,t_{i-1}}fc$ is

$$S_{fbco_{u,t_{i-1}}} = 2\arccos\frac{\gamma_1+\gamma_2}{r_{u,t_{i-1}}}r_{u,t_{i-1}} \tag{15}$$

$$S_{o_{u,t_{i-1}}fc} = \frac{\gamma_1+\gamma_2}{2\sqrt{\gamma_2}}\sqrt{r^2_{u,t_{i-1}} - \frac{(\gamma_1+\gamma_2)^2}{4\gamma_2}}. \tag{16}$$

Similarity, the area of sector $fdcl_{a,i-1}$ and triangle $\triangle fcl_{a,i-1}$ is

$$S_{fdcl_{a,i-1}} = 2\arccos\frac{\gamma_2-\gamma_1}{r_{u,t_{i-1}}}r_{a,t_{i-1}} \tag{17}$$

$$S_{fcl_{a,i-1}} = \frac{\gamma_2-\gamma_1}{2\sqrt{\gamma_2}}\sqrt{r^2_{a,t_{i-1}} - \frac{(\gamma_2-\gamma_1)^2}{4\gamma_2}}. \tag{18}$$

So we can get the area of the shaded region

$$S_{\text{MMB}_{j-1,j,i-1,i}} = S_{fbco_{u,t_{i-1}}} - S_{o_{u,t_{i-1}}fc} + S_{fdcl_{a,i-1}} - S_{fcl_{a,i-1}}. \tag{19}$$

To sum up, Theorem 2 holds. ■

Based upon Theorems 1 and 2, different spatiotemporal behaviors of high-risk users and attackers (i.e., the possibility of visiting a specific location deduced from Theorems 1 and 2, respectively) can be identified when LIA are launched. Accordingly, we can take advantage of such spatiotemporal

$$P^{l_{a,j},t_i}_{l_{a,j-1},l_{u,i-1},t_{i-1}} = \frac{1}{2\arccos\frac{\gamma_1+\gamma_2}{r_{u,t_{i-1}}}r_{u,t_{i-1}} - \frac{\gamma_1+\gamma_2}{2\sqrt{\gamma_2}}\sqrt{r^2_{u,t_{i-1}} - \frac{(\gamma_1+\gamma_2)^2}{4\gamma_2}} + 2\arccos\frac{\gamma_2-\gamma_1}{r_{u,t_{i-1}}}r_{a,t_{i-1}} - \frac{\gamma_2-\gamma_1}{2\sqrt{\gamma_2}}\sqrt{r^2_{a,t_{i-1}} - \frac{(\gamma_2-\gamma_1)^2}{4\gamma_2}}} \tag{13}$$

behaviors to model different characteristics of real locations and fake locations, as will be discussed below.

B. Mobility Similarity Metric

In this part, to distinguish real locations and fake locations, we define the mobility similarity metric between locations of high-risk users and locations of other users.

At first, we normalize the possibilities $P^{l_{u,i},t_i}_{l_{u,i-1},t_{i-1}}$ and $P^{l_{a,j},t_i}_{l_{a,j-1},l_{u,i-1},t_{i-1}}$, and denote the normalized possibilities as $q_{u,i}$ and $q_{a,j}$, i.e.,

$$q_{u,i} = \frac{P^{l_{u,i},t_i}_{l_{u,i-1},t_{i-1}}}{\sum_{x=1}^{m} P^{l_{u,x},t_x}_{l_{u,x-1},t_{x-1}}} \tag{20}$$

$$q_{a,j} = \frac{P^{l_{a,j},t_{j+1}}_{l_{a,j-1},l_{u,j},t_j}}{\sum_{x=1}^{n} P^{l_{a,x},t_{x+1}}_{l_{a,x-1},l_{u,j},t_x}}. \tag{21}$$

Then let the probability distributions of W_u and W_a, two discrete random variables, be q_u, q_a, and $P\{W_u = l_{u,i}, T_a = t_i\} = q_{u,i}$, $P\{W_a = l_{a,j}, T_a = t_i\} = q_{a,j}$, $(i \in (1, 2, \ldots), j \in (1, 2, \ldots))$, respectively.

Then we define the mobility similarity as follows.

Definition 1: The *mobility similarity* between a high-risk user u and a specific user $u_{i'}$

$$\mathrm{sim}_{(u,u_{i'})} = 1 - \frac{1}{g}\sum_{i=1}^{\eta} \min_{F_{ij}} \left\{ \sum_{i=1}^{\eta}\sum_{j=1}^{\eta-1} F_{ij} |l_{u,i} l_{u_{i'},j}| \right\} \times \frac{P_{l_{u,i-1},t_{i-1}}}{\sum_{x=1}^{\eta} P_{l_{u,x-1},t_{x-1}}} \tag{22}$$

where F_{ij} is the joint distribution function of $(W_u, W_{u_{i'}})$, which meets $\sum_{i=1}^{\eta} F_{ij} = q_{u_{i'},j}$, $\sum_{j=1}^{\eta-1} F_{ij} = q_{u,i}$, $\sum_{i=1}^{\eta}\sum_{j=1}^{\eta-1} F_{ij} = 1$, and $g = \max_{z}\{\min_{F_{ij}}\{\sum_{i=1}^{z}\sum_{j=1}^{z-1} F_{ij}|l_{u,i}l_{u_{i'},j}|\}\}$ $(z \in (1, 2, \ldots))$. The threshold of mobility similarity is specified by the high-risk user u, denoted as $\mathrm{sim}_{o(u)} = \xi_u$.

We emphasize that, except for the probability distributions, we also include the Euclidean distance $|l_{u,i}l_{u_{i'},j}|$ in mobility similarity, as the probability distributions only indicate the similarity in possibilities of visiting locations at different time, while the Euclidean distance $|l_{u,i}l_{u_{i'},j}|$ does reveal attackers' motivation that they try to enable the fake locations cloaked with high-risk users. Intuitively, a user always locating in the MMBs of a high-risk user with little probabilities is quite likely to be an attacker.

C. Generating the CR

Before digging into generating CRs for users, we first give the definitions as follows.

Definition 2: For any user u_i, we define his *credibility* Cre_{u_i} as follows:

$$\mathrm{Cre}_{u_i} = \begin{cases} 0, & \text{if } v > v_{\max} \\ \mathrm{Cre}_0, & \text{if } v \le v_{\max} \text{ and } \mathrm{num}_{q_i} = 1 \\ \mathrm{Cre}_{u_i} - \varepsilon, & \text{if } v \le v_{\max},\ 1 < \mathrm{num}_{q_i} \\ & \text{and } \mathrm{sim}_{o(u)} < \mathrm{sim}_{(u,u_i)} \\ \mathrm{Cre}_{u_i} + \varepsilon, & \text{if } v \le v_{\max},\ 1 < \mathrm{num}_{q_i} \\ & \text{and } \mathrm{sim}_{(u,u_i)} \le \mathrm{sim}_{o(u)} \\ 1, & \text{if } u_i \text{ is a high-risk user} \end{cases}$$

Algorithm 1 Generating the CR

Input A new query q from u, queries from users $u_1, u_2, \ldots, u_{l_2}$ waiting to be cloaked.

Output CR

1: $Cre_{u_i} \leftarrow$ Definition. 2, $count = 0$, $i = 0$, $\max(k) = k$, $\max(A_{min}) = A_{min,u}$, a set $\Omega = \{u\}$.
2: **if** u is a high-risk user **then**
3: **if** $count \le \max(k)$ **then**
4: **if** $num_{q_i} > 1, 0 \ne Cre_{u_i} \approx Cre_u$ **then** computing sim_{u,u_i}
5: **if** $sim_{(u,u_i)} < sim_{o(u)}$ **then** $count$++, $\max(k) = \max(\max(k), k_i)$, $\max(A_{min}) = \max(\max(A_{min}), A_{min,u_i})$, $\Omega \leftarrow \{u_i\}$.
6: **else**
7: **if** $Cre_{u_i} > 0$ **then** $Cre_{u_i} = Cre_{u_i} - \varepsilon$
 i++
8: **else**
9: **if** count$<\max(k)$ **then**
10: **if** $Cre_{u_i} \approx Cre_u$ **then** count++, i++, $\max(k) = \max(\max(k), k_i)$, $\max(A_{min}) = \max(\max(A_{min}), A_{u_i})$, $\Omega \leftarrow \{u_i\}$.
11: CR = the minimum circle c that contains the locations of users in Ω
12: **if** $s \le \max(A_{min})$ **then** enlarge CR until $\max(A_{min}) \le s$
return CR

where num_{q_i} $(1 \le \mathrm{num}_{q_i})$ is the number of u_i's LBS queries, v is the moving speed of u_i, ε and Cre_0 are a system parameter, and u is a high-risk user.

Definition 3: For any users u_i and u_j, when credibility Cre_{u_i} and Cre_{u_j} meet: $|\mathrm{Cre}_{u_i} - \mathrm{Cre}_{u_j}| \le \min\{\xi_{u_i}, \xi_{u_j}\}$, $\mathrm{Cre}_{u_i} \ne 0$, and $\mathrm{Cre}_{u_j} \ne 0$, u_i and u_j have the *approximation credibility*, denoted by $\mathrm{Cre}_{u_i} \approx \mathrm{Cre}_{u_j}$. τ is a system parameter.

Based upon the definitions of credibility and approximation credibility, we propose the following credibility-based k-cloaking algorithm.

Upon receiving the LBS query of the high-risk user u, we first filter out the obviously fake trajectories around u incorporating the location inference attack [31], and these users filtered out are treated as fake users and scored credibility 0. Then we pick out the users [denoted by $u_1, u_2, \ldots, u_l$ $(l > k)$] form the remaining users that meet the constraints: for $\forall$ u_i $(i \in (1, \ldots, l))$, $\mathrm{Cre}_{u_i} \approx \mathrm{Cre}_u$. Thereafter, we model the mobility for the high-risk user u and $u_1, u_2, \ldots, u_l$, and compute the mobility similarity between u and $u_1, u_2, \ldots, u_l$. Denote these users by $u_1, u_2, \ldots, u_{l'}$ who meet $\forall$ u_i $(i \in (1, \ldots, l'))$, $\mathrm{sim}_{u,u_i} \le \mathrm{sim}_{o(u)}$. Lastly, when $\max\{k, k_1, k_2, \ldots, k_{l'}\} \le l'+1$, we generate the CR for u and these users. Note that if more than two high-risk users are cloaked in one CR (e.g., u and u'), any a user u_i in $(u_1, u_2, \ldots, u_{l'})$ should also meet $\mathrm{sim}_{u',u_i} \le \min\{\mathrm{sim}_{o(u')}\}$.

It is important to note that the fake users that each of them issues one LBS request are assigned low credibility Cre_0, and thus these fake users are not cloaked with the high-risk user u.

The solution above focuses on selecting more credible users to be cloaked with the high-risk users to protect high-risk users' location privacy. As regard to the location privacy of other trusted users that are not cloaked with high-risk users, we

propose to cloak users owning the approximation credibility in the same CR.

In summary, Algorithm 1 shows the pseudocode of the credibility-based *k*-cloaking algorithm above.

D. Discussion

ILLIA can protect the location privacy against LIA, and guarantee the QoS in terms of the computation cost.

Theorem 3: ILLIA can protect users' location privacy against LIA, regardless of the way of manipulating fake locations that the attacker follows.

Proof: Regardless of the way of manipulating fake locations, the obviously fake trajectories are susceptible to location inference attacks, and thus are first filtered out via incorporating the state-of-the art location inference attack [31]. In addition, the plausible trajectories that deliberately imitate a user's mobility pattern are distinguished based on the credibility that deduced from the mobility similarity. In summary, ILLIA can identify fake users without requiring in advance knowledge of how fake locations are manipulated. ■

Theorem 4: The computation cost in ILLIA is at most $O(N)$, where N is the total number of users.

Proof: The first step is to model users' mobility. It incurs $O(N(u))$ computation cost to model the mobility of high-risk users, where $N(u)$ is the number of high-risk users. Then we model the mobility of other users, which incurs $O(N(a)+N(t))$ computation cost ($N(a)$ and $N(t)$ are the number of fake users and trusted users).

The second step is to compute the mobility similarity to search for qualified users that can be cloaked with high-risk users. Thus the time complexity is $O(\sum_{i=1}^{N(u)} k_i)$.

The last step is to generate CR for users, which is shown in Algorithm 1. The most expensive operation is to search the qualified users for high-risk users (lines 2–7) and other trusted users (lines 8–10), which take $O(\sum_{i=1}^{N(u)} k_i)$ and $O(\sum_{i=1}^{N(a)+N(t)} k_i)$ computation cost at most, respectively.

In summary, the time complexity is $O(N)$ at most, since $N(u) < N$, $N(a) < N$, and $N(t) < N$.

Overall, Theorem 4 holds. ■

IV. Performance Evaluation

A. Evaluation Dataset

We employ the location-based social network dataset, loc-Gowalla [32] to validate ILLIA. loc-Gowalla collects 6.4 million check-ins (i.e., locations) from Feb. 2009 to Oct. 2010, consisting of 196 591 nodes (i.e., users) and 950 327 edges (i.e., friendships).

The statistics of high-risk users, trusted users, and fake users are shown in Table I. First, we randomly select n_i users as the high-risk users, and the remaining users are regarded as the trusted users. Denote $m_{i,t}$ the number of friendship between high-risk users and trusted users. Then n_a locations are randomly chosen as n_a fake users' initial locations. Thereafter, during each updating of fake users' locations, we randomly select n_a locations as the current locations of fake users. Denote l_i and l_a the number of high-risk users' locations and fake locations. Each fake user can randomly attack a high-risk user whenever updating his locations.

TABLE I
Mobility Similarity in Various Data

n_i	l_i	$m_{i,t}$	n_a	l_a	$sim_{(u,a)}$	$sim_{(u,t)}$
65530	2577156	95885	131061	3865734	0.103	0.015

Other default settings are as follows: k is set randomly in the range of $[2, 12]$, $A_{\min}$ is set $[0.005, 0.01]$ percent of the space, $\varepsilon = 0.05$, $\text{Cre}_0 = 0.1$, and ξ is set 0.05.

B. Evaluation Metrics

To demonstrate the effectiveness and efficiency of ILLIA, we compare ILLIA with OPT and EPK. OPT cloaks users without considering LIA, while EPK straightforwardly enlarges privacy parameter k to $2k$. In addition, we focus on four metrics for performance evaluation: 1) the average attack success rate δ_a; 2) the average cloaking success rate δ_c; 3) the average processing time ω; and 4) the cumulative distribution function (CDF) of CRs. δ_a and δ_c are defined as follows.

Definition 4: A high-risk user or trusted user u suffers from the *successful LIA* when $N - N(a) < k_u$, where N is the total number of users in the CR of u, and $N(a)$ is the number of fake locations in the CR.

Definition 5: Assume $N_1(u)$ high-risk users and $N_2(u)$ trusted users are cloaked, and $n_1(u)$ high-risk users and $n_2(u)$ trusted users suffer from successful LIA, then the *attack success rate* is $\delta_a = (n_1(u) + n_2(u))/(N_1(u) + N_2(u))$.

Definition 6: Assume $N_1(u)$ high-risk users, $N_2(u)$ trusted users and $N_3(a)$ attackers are cloaked, and the total number of users is N, then the *cloaking success rate* is $\delta_c = (N_1(u) + N_2(u) + N_3(a))/N$.

C. Mobility Similarity Analysis

Before digging into the performance of ILLIA, we focus on characterizing the mobility similarity in default settings, which is shown in Table I. The mobility of high-risk users and trusted users are not strongly similar to each other, with the mobility similarity $\text{sim}_{(u,t)}$ less than 0.009. In contrast, high-risk users and attackers share a more similar mobility $\text{sim}_{(u,a)}$.

D. Performance Varies With k and n_a

In this part, we investigate the impact of privacy parameter k and the number of attackers n_a. We vary each user's k from 2 to 12. We set the number of attackers $n_a = 131\,061$ (the number of fake locations $l_a = 3\,865\,734$) and $n_a = 98\,295$ ($l_a = 2\,577\,156$), respectively.

1) δ_a Varies With k and n_a: Fig. 6(a) shows that average attack success rate δ_a varies with k and n_a. It can be observed that ILLIA yields lower δ_a than the other two schemes, and the superiority of ILLIA is even more prominent as k and n_a increase. Specifically, δ_a in ILLIA is 2.83 times lower than that in OPT and 1.04 times lower than that in EPK at most. The

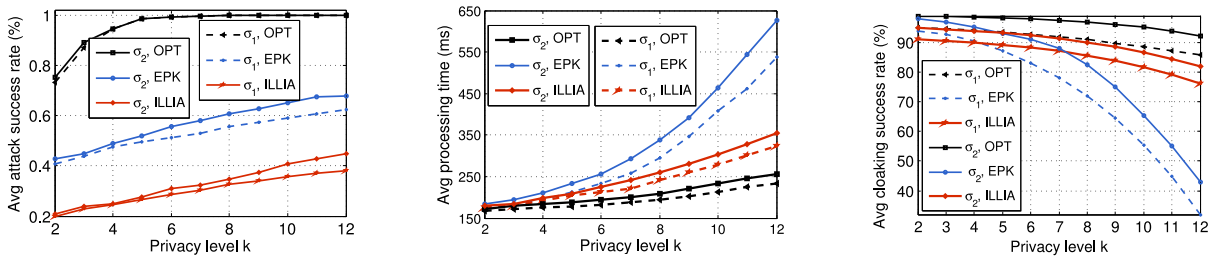


Fig. 6. δ_a, ω, and δ_c vary with k and n_a; $\sigma_1 = 98\,295$ and $\sigma_2 = 131\,061$ are the numbers of attackers (i.e., n_a). (a) δ_a varies with k and n_a. (b) ω varies with k and n_a. (c) δ_c varies with k and n_a.

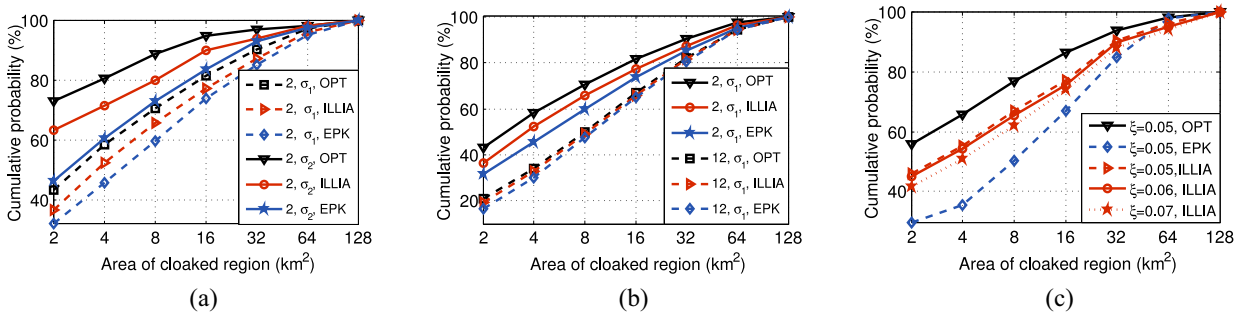


Fig. 7. (a) CDF varies with n_a; 2 and 12 in the legend are the values of k; $\sigma_1 = 98\,295$, $\sigma_2 = 131\,061$. (b) CDF varies with k. (c) CDF varies with ξ. (a) CDF varies with n_a. (b) CDF varies with k. (c) CDF varies with ξ.

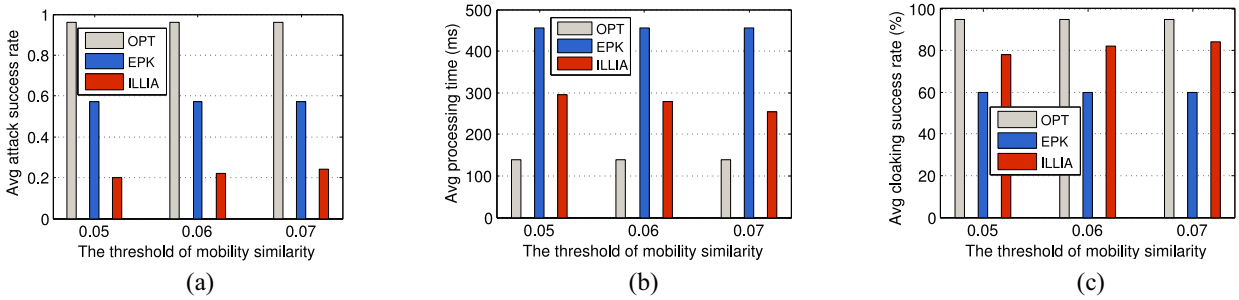


Fig. 8. δ_a, ω, and δ_c vary with ξ.

better performance of ILLIA benefits from the mobility similarity between locations of attackers and high-risk users and the credibility-based k-cloaking algorithm. In contrast, OPT does not consider LIA, and EPK does not concern the credibility of users' locations. Additionally, it can be seen that δ_a in these algorithms increases when increasing k and n_a. This is not much surprising as a larger k or n_a means a higher possibility to cloak a fake location with locations of trusted users or high-risk users, thus incurring much higher δ_a.

2) ω Varies With k and n_a: The processing time ω is shown in Fig. 6(b). ω in ILLIA is less than that in EPK, and more than that in OPT. One reason is that EPK enlarges k and has to cloaks more locations, and OPT does not consider LIA. Also as expected, the overall trends in all cloaking algorithms are similar: ω increases with the increase of k and n_a. That is because increasing k means a more constrained privacy requirement, and a larger n_a means all cloaking algorithms have to cloak more locations of users.

3) δ_c Varies With k and n_a: In Fig. 6(c), it can be observed that δ_c in these algorithms decreases with k and n_a. This is not much surprising as increasing both the privacy level k and the number of fake locations will result in longer ω, and therefore lead to more LBS queries expired and not successfully cloked. Besides, compared with OPT, only a success rate of 10.13% at most in ILLIA is sacrificed for protecting against LIA, while 53.8% success rate in EPK is sacrificed. This is mainly because ω in EPK is much longer than that in OPT and ILLIA, incurring more LBS queries expired.

4) CDF Varies With k and n_a: In Fig. 7(a) and (b), we can see that CRs' area in ILLIA is smaller than that in EPK, but larger than that in OPT. In addition, CRs' area is positively correlated with k [see Fig. 7(b)], as larger k means algorithms have to cloak further users to complete location k-anonymity, and thus results in larger CRs. Moreover, the size of CRs decreases with n_a [see Fig. 7(a)], because more users around

a specific user and algorithms can generate a smaller CR so that it contains no less than $(k-1)$ other users.

E. Performance Varies With ξ

In this section, we study the impact of the threshold of mobility similarity ξ, and set ξ 0.05, 0.06, and 0.07, respectively. Other parameters are as the default settings.

1) δ_a Varies With ξ: We can see δ_a in ILLIA increases when increasing ξ, while δ_a in OPT and EPK nearly keeps steady, which is shown in Fig. 8(a). The reason is that more fake locations can be cloaked with locations of high-risk users and trusted users when we enlarge ξ in ILLIA. In contrast, both OPT and EPK do not consider the credibility of locations, and thereby δ_a in the two algorithms is not affected by ξ. Additionally, ILLIA outperforms EPK, resulting in a 25%–35% improvement. The reasons are as explained in Fig. 6(a).

2) ω Varies With ξ: ω in ILLIA decreases when increasing ξ, as shown in Fig. 8(b). In contrast, ω in OPT and EPK are not affected. In addition, we can clearly see that ω in ILLIA is much lower than that in EPK, but a bit higher than that in OPT, since OPT does not consider LIA, and EPK enlarges k.

3) δ_c Varies With ξ: As we expected, δ_c in OPT and EPK is not affected by $\text{sim}_{o(u,a)}$, because they do not consider the credibility of users' locations, which is shown in Fig. 8(c). While δ_c in ILLIA increases when increasing ξ, it is because a larger ξ results in a less processing time, and thus a higher cloaking success rate.

4) CDF Varies With ξ: As shown in Fig. 7(c), CR's area in ILLIA decreases when increasing ξ, but it is not affected in OPT and EPK. Besides, CR's area in ILLIA is obviously less than that in EPK.

V. Conclusion

In this paper, we have presented ILLIA, the first work that can preserve location privacy against LIA in continuous LBS queries. ILLIA is appealing as it is a scalable and general solution with location privacy-guaranteed and QoS-guaranteed simultaneously, requiring no in advance knowledge of the way of manipulating fake locations used by attackers. Extensive simulations on loc-Gowalla dataset demonstrate the effectiveness and efficiency of ILLIA.

References

[1] H. Lu, J. Li, and M. Guizani, "Secure and efficient data transmission for cluster-based wireless sensor networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 3, pp. 750–761, Mar. 2014.

[2] J. Y. Koh, I. Nevat, D. Leong, and W.-C. Wong, "Geo-spatial location spoofing detection for Internet of Things," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 971–978, Dec. 2016.

[3] C. Wang, H. Lin, and H. Jiang, "CANS: Towards congestion-adaptive and small stretch emergency navigation with wireless sensor networks," *IEEE Trans. Mobile Comput.*, vol. 15, no. 5, pp. 1077–1089, May 2016.

[4] S. Landau, "Control use of data to protect privacy," *Science*, vol. 347, no. 6221, pp. 504–506, 2015.

[5] K. Fawaz, H. Feng, and K. G. Shin, "Anatomization and protection of mobile apps' location privacy threats," in *Proc. USENIX Security Symp.*, Washington, DC, USA, 2015, pp. 753–768.

[6] X. Wu *et al.*, "Social network de-anonymization with overlapping communities: Analysis, algorithm and experiments," in *Proc. IEEE INFOCOM*, May 2018.

[7] W. Enck *et al.*, "TaintDroid: An information-flow tracking system for realtime privacy monitoring on smartphones," *ACM Trans. Comput. Syst.*, vol. 32, no. 2, pp. 1–29, 2014.

[8] *Dark Net Linkedin Sale Looks Like the Real Deal*. Accessed: Dec. 12, 2016. [Online]. Available: http://www.theregister.co.uk/2016/05/18/linkedin/

[9] *Recently Confirmed Myspace Hack Could Be the Largest Yet*. Accessed: Feb. 27, 2017. [Online]. Available: https://techcrunch.com/2016/05/31/recently-confirmed-myspace-hack-could-be-the-largest-yet/

[10] S. Gisdakis, T. Giannetsos, and P. Papadimitratos, "Security, privacy, and incentive provision for mobile crowd sensing systems," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 839–853, Oct. 2016.

[11] B. Han, J. Li, J. Su, M. Guo, and B. Zhao, "Secrecy capacity optimization via cooperative relaying and jamming for WANETs," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 4, pp. 1117–1128, Apr. 2015.

[12] J. Sun, R. Zhang, and Y. Zhang, "Privacy-preserving spatiotemporal matching for secure device-to-device communications," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 1048–1060, Dec. 2016.

[13] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *Proc. ACM MobiSys*, San Francisco, CA, USA, 2003, pp. 31–42.

[14] V. Bindschaedler and R. Shokri, "Synthesizing plausible privacy-preserving location traces," in *Proc. IEEE S&P*, San Jose, CA, USA, 2016, pp. 546–563.

[15] J.-D. Zhang and C.-Y. Chow, "REAL: A reciprocal protocol for location privacy in wireless sensor networks," *IEEE Trans. Depend. Secure Comput.*, vol. 12, no. 4, pp. 458–471, Jul./Aug. 2015.

[16] Y. Zhang, W. Tong, and S. Zhong, "On designing satisfaction-ratio-aware truthful incentive mechanisms for *k*-anonymity location privacy," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 11, pp. 2528–2541, Nov. 2016.

[17] B. Niu, Q. Li, X. Zhu, G. Cao, and H. Li, "Achieving *k*-anonymity in privacy-aware location-based services," in *Proc. IEEE INFOCOM*, Toronto, ON, Canada, 2014, pp. 754–762.

[18] X. Liu, K. Liu, L. Guo, X. Li, and Y. Fang, "A game-theoretic approach for achieving *k*-anonymity in location based services," in *Proc. IEEE INFOCOM*, Turin, Italy, 2013, pp. 2985–2993.

[19] J. Zeng *et al.*, "Mobile r-gather: Distributed and geographic clustering for location anonymity," in *Proc. ACM MobiHoc*, Chennai, India, 2017, Art. no. 7.

[20] G. Ghinita, M. L. Damiani, C. Silvestri, and E. Bertino, "Preventing velocity-based linkage attacks in location-aware applications," in *Proc. ACM SIGSPATIAL GIS*, Seattle, WA, USA, 2009, pp. 246–255.

[21] G. Ghinita, M. L. Damiani, C. Silvestri, and E. Bertino, "Protecting against velocity-based, proximity-based, and external event attacks in location-centric social networks," *ACM Trans. Spatial Algorithms Syst.*, vol. 2, no. 2, pp. 1–36, 2016.

[22] *FakeGPStracker*. Accessed: May 7, 2017. [Online]. Available: http://fakegps.com/

[23] L. Jin, B. Palanisamy, and J. B. D. Joshi, "POSTER: Compromising cloaking-based location privacy preserving mechanisms with location injection attacks," in *Proc. ACM SIGSAC CCS*, Scottsdale, AZ, USA, 2014, pp. 1439–1441.

[24] R. Shokri, G. Theodorakopoulos, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec, "Protecting location privacy: Optimal strategy against localization attacks," in *Proc. ACM CCS*, Raleigh, NC, USA, 2102, pp. 617–627.

[25] G. Wang, T. Wang, W. Jia, M. Guo, and J. Li, "Adaptive location updates for mobile sinks in wireless sensor networks," *J. Supercomput.*, vol. 47, no. 2, pp. 127–145, 2009.

[26] X. Pan, J. Xu, and X. Meng, "Protecting location privacy against location-dependent attacks in mobile services," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 8, pp. 1506–1519, Aug. 2012.

[27] H. Yin *et al.*, "Joint modeling of user check-in behaviors for real-time point-of-interest recommendation," *ACM Trans. Inf. Syst.*, vol. 35, no. 2, pp. 1–44, 2016.

[28] J. Habibi, D. Midi, A. Mudgerikar, and E. Bertino, "Heimdall: Mitigating the Internet of insecure things," *IEEE Internet Things J.*, vol. 4, no. 4, pp. 968–978, Aug. 2017.

[29] Y. Afek, A. Bremler-Barr, and L. Shafir, "Network anti-spoofing with SDN data plane," in *Proc. IEEE INFOCOM*, Atlanta, GA, USA, 2017, pp. 1–9.

[30] L. Wang, J.-H. Cho, I.-R. Chen, and J. Chen, "PDGM: Percolation-based directed graph matching in social networks," in *Proc. IEEE ICC*, Paris, France, 2017, pp. 1–7.

[31] R. Shokri, G. Theodorakopoulos, G. Danezis, J.-P. Hubaux, and J.-Y. Le Boudec, "Quantifying location privacy: The case of sporadic location exposure," in *Proc. Int. Symp. Privacy Enhanc. Technol. Symp.*, Waterloo, ON, Canada, 2011, pp. 57–76.

[32] J. Leskovec and A. Krevl. (Jun. 2014). *SNAP Datasets: Stanford Large Network Dataset Collection*. [Online]. Available: http://snap.stanford.edu/data

Ping Zhao received the B.E. degree from the Tianjin University of Science and Technology, Tianjin, China, in 2013. She is currently pursuing the Ph.D. degree at the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China.

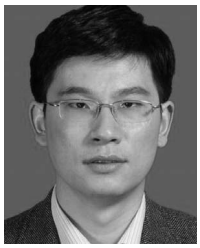
Her current research interests include wireless networking, especially privacy protection in mobile networks.

Jie Li received the B.S. degree from the College of Computer Science and Electronic Engineering, Hunan University, Changsha, China, in 2016, where she is currently pursuing the Ph.D. degree at the College of Computer Science and Electronic Engineering.

Her current research interests include mobile and wireless networks, especially privacy preserving in mobile systems.

Fanzi Zeng received the Ph.D. degree in signal and information processing from Beijing Jiaotong University, Beijing, China, in 2005.

Since 2005, he has been with the School of Information Science and Engineering, Hunan University, Changsha, China, where he is currently a Professor. His current research interests include signal processing for wireless communications, estimation, and detection theory, and cognitive radio technology.



Fu Xiao received the Ph.D. degree in computer science and technology from the Nanjing University of Science and Technology, Nanjing, China.

He is currently a Professor and the Ph.D. Supervisor with the School of Computer, Nanjing University of Posts and Telecommunications, Nanjing. His current research interests include wireless networking and sensor networks.

Chen Wang (S'10–M'13) received the B.S. and Ph.D. degrees from the Department of Automation, Wuhan University, Wuhan, China, in 2008 and 2013, respectively.

From 2013 to 2017, he was a Post-Doctoral Research Fellow with the Networked and Communication Systems Research Laboratory, Huazhong University of Science and Technology, Wuhan. He then joined the faculty of the Huazhong University of Science and Technology, where he is currently an Associate Professor. His current research interests include wireless networking, Internet of Things, and mobile computing, with a recent focus on privacy issues in wireless and mobile systems.

Hongbo Jiang (M'08–SM'14) received the B.S. and M.S. degrees from the Huazhong University of Science and Technology, Wuhan, China, and the Ph.D. degree from Case Western Reserve University, Cleveland, OH, USA, in 2008.

He joined the faculty of the Huazhong University of Science and Technology, where he is currently a Full Professor and the Dean of the Department of Communication Engineering. His current research interests include computer networking, especially algorithms and protocols for wireless and mobile networks.

Dr. Jiang is serving as an Editor for the IEEE/ACM TRANSACTIONS ON NETWORKING, an Associate Editor for the IEEE TRANSACTIONS ON MOBILE COMPUTING, and the Associate Technical Editor for *IEEE Communications Magazine*.