# Activity Tracking: A New Attack on Location Privacy

Xihui Chen*, Andrzej Mizera†, Jun Pang*†

*Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg
†Faculty of Science, Technology and Communication, University of Luxembourg

***Abstract*—The exposure of location information in location-based services (LBS) raises users' privacy concerns. Recent research reveals that in LBSs users concern more about *the activities that they have performed* than *the places that they have visited*. In this paper, we propose a new attack with which the adversary can accurately infer users' activities. Compared to existing attacks, our attack provides the adversary not only with the places where users perform activities but also with the information when they stay at each of these places. To achieve this objective, we propose a new model to capture users' mobility and their LBS requests in continuous time, which naturally expresses users' behaviour in LBSs. We then formally implement our attack by extending an existing framework for quantifying location privacy. Through experiments on a real-life dataset, we show the effectiveness of our new tracking attack.**

## I. Introduction

Nowadays, most people are equipped with mobile devices which can obtain their locations in real time. Location-based services (LBS) benefit from this technical progress and bring great convenience to people's daily life by offering responses customised to their whereabouts. However, the frequent exposure of whereabouts in LBS requests may leak users' personal information and thus breach their privacy [1], [2]. As a consequence, many *location privacy preserving methods* (LPPM) [3], [4] have been proposed. In general, the purpose of an LPPM is to break the link between users and their locations. Their main idea is to *anonymise* user identities and *obfuscate* location coordinates by replacing them with pseudonyms and regions, respectively. Meanwhile, many attacks on location privacy have been proposed and successfully demonstrated that attackers can still manage to associate users to their locations in the presence of LPPMs [5], especially when users' background information (e.g., velocity, occupation) is explored. For instance, Shokri et al. [6] implement a tracking attack using user mobility profiles through which the adversary can learn the locations of a user when he requested LBSs. Existing attacks in the literature mostly target at deriving 'where users actually visited'. However, recent research requires us to revisit this objective from the viewpoint of attackers in practice. Namely, what the adversary is really curious about, with respect to location privacy, is what users did during their movement, i.e., their *activities* [3]. Users' location privacy is thus exposed to a new threat which has not been studied in the literature. In this paper, we address how to effectively formalise and implement this threat.

A number of methods have been proposed to fill the gap between where users visited and what activities they performed. The concept of *points of interest* (PoI) is introduced to represent places where a user may stay and perform activities. PoIs subsequently leads to the exploitation of *location semantics* (e.g., hospital, school) to infer the possible activities that a user can perform [7], [8], [9], [10]. Since PoIs, such as a commercial centre, are usually annotated with multiple semantics, e.g., cinema and supermarket, probability distributions over location semantics are extracted to decrease the adversary's uncertainty about the real location semantics [11]. Although PoIs and their semantics are widely accepted in the literature, they are not sufficient to accurately infer users' activities. The adversary still has a high probability to be incorrect as the distributions only capture users' choice of activities statistically and do not consider the purposes of a single visit for individual users.

**Motivations.** Services offered by a PoI normally possess certain patterns with respect to their busy time and the amount of delivery time. With this observation, we can further reduce the adversary's uncertainty by making use of such temporal patterns. Therefore, to accurately infer users' activities, the adversary should take into account when they *enter* each PoI and the amount of time for which they *stay* in it. We cannot simply explore existing attacks, e.g., the ones described by Shokri et al. [6] to achieve this goal: these attacks target at users' locations at given time points and thus users' profiles are constructed to capture their behaviour (e.g., mobility) in discrete time space. Furthermore, the computational overhead grows in the numbers of targeted time points and locations [6]. To achieve a similar goal, the attacks should be constructed with (i) sufficiently fine-grained locations to correctly identify PoIs; (ii) a fine-grained time space, e.g., in seconds, to capture the exact entering and exiting time of a PoI. These two requirements make the attacks [6] infeasible in practice, especially with long periods of movements in a large area.

**Contributions.** In this paper, we propose a new *activity tracking* attack. It enables the adversary to *directly* learn a new form of trajectories: *activity trajectories* which contain not only the sequence of PoIs where a user performed activities but also his entering and exiting time at each of the PoIs. With activity trajectories, the adversary can deduce users' activities with a high correctness rate. To perform this attack, we propose a new model for *user profiles* capturing users' *mobility patterns* and *temporal patterns* when issuing LBS requests. Compared to existing models, our new model has two main differences: (i) users' behaviour is modelled with continuous time; (ii) users' movements are modelled with transitions between PoIs instead of locations. We make use of the former to describe users' behaviour in a more flexible and natural manner while the latter makes our model more expressive as it allows us to explicitly specify users' transitions between activities and their temporal patterns in movements, i.e., the stay time in PoIs and transition time between two PoIs. We formally define our new tracking attack by extending a well-recognised formal framework in the literature [6]. For validation, we implement and apply our activity tracking attack on a real-life trajectory dataset. The experimental results demonstrate that our model and tracking attack are rather effective.

## II. Related Work

In this section, we briefly discuss the state-of-the-art from the following two perspectives.

**Protecting location privacy.** In general, location privacy preserving mechanisms (LPPM) can be divided into two classes. The first class exploits cryptographic methods to encrypt LBS requests [12] and to prevent locations from being exposed or eavesdropped. However, this class of methods incur additional computational overhead due to the cryptographic primitives used. The second class of LPPMs adopt a different approach which modifies LBS requests and hides the link between users and their accurate locations. We can further categorise the methods of this class into two types: *anonymisation* and *obfuscation*. Anonymising LPPMs replace users' identities with pseudonyms while obfuscating LPPMs modify the geographic information in LBS requests. Cloaking [3], [4] and perturbation [13] are two of the most used obfuscating methods. The former reduces the precision of locations while the latter adds other locations as noise. Hiding request and adding dummy requests [13] are another two obfuscating methods. In practice, such methods are implemented in many different ways, which lead to various LPPMs. For instance, a user can choose to use different pseudonyms according to the types of LBSs or just use one common for all requests, while the precision of locations can be tailored to kilometres or metres. The implementations usually determine the amount of privacy ensured by the LPPMs.

**Attacking location privacy.** Despite of the development of LPPMs, location privacy can still be violated, especially when background knowledge is exploited by the adversary. For example, it has been shown that anonymous trajectories can be associated with their originators' home and work addresses [14]. Mulder et al. [15] show that it is possible to de-anonymise trajectories once users' previous movements are made available. In this attack, *location profiles* of Markovian models are extracted and used.

Due to the diversity of user profiles and attacks, a unified framework is needed to implement attacks in a way which allows quantitative evaluation of LPPMs. Shokri et al. in [6] made the first successful and impressive attempt. Next, their framework was extended and explored in a number of ways [6], [16], [17]. In [16] the original framework was augmented to allow the modelling of sporadic requests to LBSs: the actual issuing of an LBS request at each of the considered time points was governed by a Bernoulli distribution. The original framework was also adopted in [18] to calculate the optimal parameters for LPPMs where strategic adversaries were assumed to be aware of the implementation details of the deployed LPPMs and to know user mobility profiles. This work was further refined by Herrmann et al. in [17] by considering the bandwidth constraints in cases where dummy requests are issued to perturb the real requests.

## III. System Model

In this section, we describe our extension of the formal framework in [6] to model the components required to define our new tracking attack. The framework can be denoted as a quadruple $\langle \mathcal{U}, LPPM, ADV, \mathcal{M} \rangle$ where $\mathcal{U}$ is a set of users, $LPPM$ represents the set of deployed LPPMs, $ADV$ is the adversary and $\mathcal{M}$ denotes privacy metrics.

**Users.** We consider a set of users $\mathcal{U} = \{u_1, \ldots, u_n\}$ who subscribe certain LBSs and move within a common area, e.g., a city. We use $\mathcal{L}$ to denote all possible geographical locations in the area which users can visit. Their formats and accuracy are determined by positioning devices. Since users request LBSs and expose their locations whenever needed, we should not exclude any time instant from being a possible location exposure time. Thus, we model the issuing time of an LBS request as a random variable whose value is chosen from a subset of non-negative real line. We denote this subset by $\mathcal{T}$ and we use $[t, t']$ ($t, t' \in \mathcal{T}$ and $t \leq t'$) to represent a time interval from time $t$ to $t'$ ($t$ and $t'$ included).

The *trajectory* of user $u$ is the path that the user follows through the considered area in time. We model it as a function mapping a time point in $\mathcal{T}$ to the location where user $u$ was located at that time instance, i.e., $\mathit{traj}_u : \mathcal{T} \to \mathcal{L}$.

A user moves from one place to another for some purposes, e.g., shopping or working. This indicates that the user has to perform different activities to achieve these purposes. Thus, users' trajectories actually record users' geographical traces when they move around to perform such activities. To capture users' activities, in addition to trajectories, we propose a new concept to record users' activities performed in a given time interval. An activity of a user includes at least three types of information: where it is performed, when it starts and ends. We introduce the concept of *points of interest* (PoI) from the literature to represent the places where users can perform an activity. Specifically, a PoI stands for a region, which is actually a set of adjacent locations in $\mathcal{L}$. We use $\mathcal{P} \subset 2^{\mathcal{L}}$ to denote the set of all PoIs of all users.[1] With respect to the time when users start and finish the activities in a PoI, we use the entering time and exiting time of the PoI to approximate

[1] $2^{\mathcal{X}}$ denotes the power set of the set $\mathcal{X}$.

them. With all these three types of information, we denote an activity as a quadruple $\langle u, p, [t_b, t_e]\rangle$, where $u \in \mathcal{U}, p \in \mathcal{P}$ and $t_b, t_e \in \mathcal{T}$. It represents the fact that user $u$ was performing an activity in PoI $p$ in the time interval from $t_b$ to $t_e$. We represent all activities of user $u$ that he performed within a time interval $[t, t']$ with the following sequence:

$$a_u^{t,t'} = (\langle u, p_1, [t_b^1, t_e^1]\rangle, \ldots, \langle u, p_n, [t_b^n, t_e^n]\rangle),$$

where $t_b^i < t_e^i$ and $t_e^i < t_b^{i+1}$ for any $1 \leq i < n$. Note that users may have different states at time $t$ which are entering PoI $p_1$, staying in $p_1$ or on the way to $p_1$. Formally,

$$(t_b^1 \leq t < t_e^1) \vee (t_b^1 \geq t \wedge \nexists \langle u, p, [t_b, t_e]\rangle, t < t_e < t_b^1).$$

At $t'$, we have three similar states: just exiting $p_n$, staying in $p_n$ or on the way to the next PoI after $p_n$ which the user reaches after time $t'$. Formally,

$$(t_b^n < t' \leq t_e^n) \vee (t_e^n \leq t' \wedge \nexists \langle u, p, [t_b, t_e]\rangle, t_e^n < t_b < t').$$

Therefore, $a_u^{t,t'}$ can be considered as the *shortest* sequence of activities which cover the time in $[t, t']$ which are not spent on transitions between activities. We call this sequence the *activity trajectory* of user $u$ in $[t, t']$.

In the setting of LBSs, a user exposes his locations to request LBSs. We call such an action an *exposure event* and denote it by a triple $\langle u, t, \mathit{traj}_u(t)\rangle$ if user $u$ requested LBSs at time $t$ while being at $\mathit{traj}_u(t)$. We call the time ordered sequence of exposure events of user $u$ his *exposed trajectory*. Given a time interval $[t, t']$, we can denote the exposed trajectory of user $u$ as follows:

$$e_u^{t,t'} = (\langle u, t_1, \mathit{traj}_u(t_1)\rangle, \ldots, \langle u, t_k, \mathit{traj}_u(t_k)\rangle)$$

where $t \leq t_i < t_{i+1} \leq t'$ $(1 \leq i < k)$.

We observe that in popular LBSs such as Foursquare and Facebook, a user's exposure events are intensively issued in some small regions. For instance, users make check-ins in places they visit for certain purpose, e.g., restaurants and bars. In order to capture this new characteristic in currently popular LBSs, we make an assumption in this paper that users request LBSs in PoIs and thus all locations in exposure events belong to certain PoIs. Formally, for any $1 \leq i \leq k$, there always exists a PoI $p \in \mathcal{P}$ such that $\mathit{traj}_u(t_i) \in p$.

**LPPMs.** LPPMs are designed to protect users' privacy by processing their LBS requests before sending them to LBS providers. In other words, an LPPM takes users' exposure events as input and modifies or distorts certain information involved. In practice, LPPMs can be implemented locally on user devices or remotely on other trusted agents. In addition, according to the time when LBS requests are processed, LPPMs can be divided into two classes: *on-line* and *off-line* LPPMs. On-line LPPMs provide real-time services while off-line services process all requests at a time. Since in practice, people require real-time services, we only consider on-line LPPMs in this paper. For the same reason, we follow the assumption of Shokri et al. [6] that LPPMs do not modify temporal information.

Focusing on popular LBSs, we consider three types of LPPMs: *anonymisation*, *cloaking*, and *perturbation*. For obfuscating mechanisms, we do not take into account the two frequently studied LPPMs: hiding and adding dummies, as they either result in frequent loss of access to LBSs or cause extra communication overhead which compromises user experience and wastes data allowance.

Anonymisation replaces user identities in exposure events with pseudonyms. Although a different pseudonym can be assigned to each exposure event, as our goal is to address location privacy issues in real-life applications, e.g., Facebook or Twitter, we follow the assumption in [6] that each user is assigned a unique pseudonym. Let $\mathcal{U}' = \{u'_1, \ldots, u'_n\}$ be the set of pseudonyms which users in $\mathcal{U}$ select. Then, an anonymising LPPM can be modelled as a bijective function mapping a user identity in $\mathcal{U}$ to a pseudonym in $\mathcal{U}'$, i.e., $\sigma : \mathcal{U} \to \mathcal{U}'$. In this paper, we follow the minimal information principle. Although users may have their own preference on selecting their own pseudonyms, we assume that such preference is out of our analysis. Thus, the anonymising function is selected uniformly from all candidate functions. Thus, the probability of $\sigma$ (denoted by *anony*$(\sigma)$) is $\frac{1}{|\mathcal{U}|!}$.

Obfuscation mechanisms substitute the locations in exposure events with other locations (perturbation) or set of locations, i.e., regions (cloaking). Such substitutions can be computed in many different ways, which can be divided into two classes based on whether previous exposure events are referred to or not. A straightforward implementation is to treat each exposure event independently. We focus on this implementation in this paper due to its popularity and simplicity. The obfuscating mechanism replaces the location $\ell \in \mathcal{L}$ in an exposure event with a *location pseudonym* $r \in \mathcal{R}$ according to the probability $\mathit{obf}(r \mid \ell)$ where $\mathcal{R} \subseteq 2^{\mathcal{L}}$ is the set of location pseudonyms.

The modified exposure events of a user can be observed by outside observers, e.g., attackers. Therefore, we refer to them as *observed events*. The *observed trajectory* of user $u$ in time interval $[t, t']$ records the list of all observed events exposed within the time period $[t, t']$ ordered in the increasing order of exposure time. We represent this trajectory as $o_{u'}^{t,t'} = (\langle u', t_1, r_1\rangle, \ldots, \langle u', t_k, r_k\rangle)$, where $u' = \sigma(u)$ and $\forall 1 \leq i \leq k, r_i \in 2^{\mathcal{L}}$. Further, we use $o^{t,t'}$ to denote the set of all observed trajectories in the time interval $[t, t']$ of all users.

**The adversary.** The adversary should be modelled to capture attackers in practice in terms of their *objectives*, *knowledge*, and *attacks*. Intuitively, an objective of the adversary is the information she aims to obtain. Knowledge is some *a priori* information being in possession of the adversary. An attack consists of a number of steps in accordance with the adversary's knowledge and executing these steps allows the adversary to achieve her objectives. In this paper, we discuss the objectives related to users' activities, which complement the existing research only targeting at users' whereabouts.

We consider a strong adversary, i.e., we make the following three assumptions with respect to the adversary's knowledge.

- (i) The adversary is aware of the anonymising and obfuscating mechanisms exploited, i.e., *anony* and *obf*, with all relevant details concerning their implementation.
- (ii) The adversary is in possession of users' observed trajectories in the time intervals of interest.
- (iii) For each user, his trajectory has been recorded for a sufficiently long period with a relatively high recording frequency and good accuracy. This information can be obtained by side channel attacks, e.g., breaking into the servers which users trust and expose his detailed movements to.

Based on the above information, the adversary can deduce additional information that enhances her knowledge. In particular, from users' travel history, the adversary can infer characteristic patterns in their movements. We refer to them as users' *mobility profiles*. The mobility profiles are constructed to foster the performing of attacks and thus their specific formats vary between attacks. We propose a new model for mobility profiles in Section IV which facilitates the attacks on users' activities. In the sequel, we use $\mathcal{K}$ to denote the adversary's knowledge about all users in $\mathcal{U}$ and $\mathcal{K}_u$ to represent the knowledge about user $u \in \mathcal{U}'$.

**The metric.** The adversary performs an attack to learn certain information, which constitutes the adversary's objectives. With the purpose of generality of presentation, we model an objective as some generic random variable $Y$ that takes values in the set $\mathcal{Y}$. We assume that given a set of observed trajectories $o$ and the adversary's knowledge $\mathcal{K}$, the actual but unknown information value is distributed in accordance with some posterior probability distribution denoted $\Pr(Y = y \mid o, \mathcal{K})$. This probability distribution is used as the basis for measuring the users' privacy guaranteed by LPPMs against attacks. In particular, we adopt the expected estimation error proposed by Shokri et al. [6] to measure users' privacy. In their work, this metric has been proved to outperform others in terms of accuracy in evaluating users' privacy. If $y' \in \mathcal{Y}$ is the real value of the objective, the estimation error of the adversary is calculated as follows:

$$privacy(o, \mathcal{K}) = \mathbb{E}_{\Pr(Y=y|o,\mathcal{K})}(\parallel y, y' \parallel), \qquad (1)$$

where the expected value is taken with respect to $\Pr(Y = y \mid o, \mathcal{K})$ and $\parallel y, y' \parallel$ denotes the distance between $y$ and $y'$. The exact definition of the distance depends on the type of information the adversary is interested in. For example, if the objective is to discover the real identity of the user $u'$, then $\parallel u, u' \parallel$ is 0 when $u = u'$ and 1, otherwise.

## IV. Profiling Users

In this section, we propose a new model for user mobility profiles and LBS request issuing patterns.

**Mobility profiles.** A user mobility profile captures his movement patterns. It is necessary for location privacy attacks because it facilitates the prediction of future moves performed by the user. Additional information, e.g., observed trajectories, can be further used by the adversary to reduce the uncertainty of the prediction. User mobility profiles should be designed in a way which assures that the movement predictions obtained from them can be efficiently exploited for further inference of user's private information.

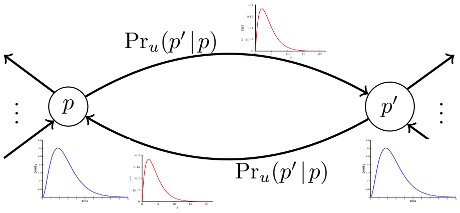


Fig. 1. An example of user $u$'s mobility profile.

The idea of our new model is inspired by an intuitive observation. Namely, a user always moves with certain purposes which actually determine the places the user will visit. Furthermore, to accomplish a purpose, a user usually stays in a PoI. Therefore, a user's trajectory can be divided into two types of segments: *stay at a PoI* and *transition between PoIs*. This division subsequently leads to our new model for mobility profiles:

- $\mathcal{P}_u$: the set of PoIs of users in $\mathcal{U}$.
- $\Pr_u(p' \mid p)$: the probability that a user $u$ will move to PoI $p'$ after leaving $p$.
- $\Gamma_{stay} : \mathcal{P}_u \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$: for any $p \in \mathcal{P}_u$, $\Gamma_{stay}(p, \delta)$ is the probability density of user $u$ staying at PoI $p$ for the amount of time $\delta$ when he visits the PoI. It holds that $\int_{\delta \geq 0} \Gamma_{stay}(p, \delta) \mathrm{d}\delta \ = 1$.
- $\Gamma_{tran} : \mathcal{P}_u \times \mathcal{P}_u \times \mathbb{R}_{>0} \rightarrow \mathbb{R}_{\geq 0}$: for any $p, p' \in \mathcal{P}_u$, $\Gamma_{tran}(p, p', \delta)$ is the probability density of a user spending the amount of time $\delta$ transiting from PoI $p$ to $p'$. In consequence, we have $\int_{\delta > 0} \Gamma_{tran}(p, p', \delta) \mathrm{d}\delta \ = 1$.
- $\delta_{min} : \mathcal{P}_u \times \mathcal{P}_u \rightarrow \mathbb{R}_{>0}$: for any $p, p' \in \mathcal{P}_u$, $\delta_{min}(p, p')$ is the minimum amount of time required to transit from $p$ to $p'$. For any $p \in \mathcal{P}_u$, $\delta_{min}(p, p) > 0$. Having $\delta_{min}(p, p) = 0$ implies that a user does not actually leave $p$.

Figure 1 depicts part of user $u$'s mobility profile in terms of two PoIs. Intuitively, after user $u$ enters a PoI $p$, he stays at $p$ for a certain amount of time $\delta_s$ according to $\Gamma_{stay}(p, \delta_s)$. Then, user $u$ selects his next PoI $p'$ with probability $\Pr_u(p' \mid p)$ and the transition time $\delta_t$ from $p$ to $p'$ follows the density function $\Gamma_{tran}(p, p', \delta_t)$.

In this model, we have made the following assumptions. First, a user determines his next destination based on his past location. In other words, the sequence of visited PoIs follows a first-order Markov chain. Second, the time spent in a PoI is only related to the PoI itself. This is reasonable because the amount of stay time in a PoI is determined by the activities performed. Third, the transition time between two PoIs depends on the source and destination. This is reasonable because transition time is mainly determined by the distance between two PoIs and factors affecting movement, e.g., traffic and weather. Last, users require a minimum time to move between two PoIs which is restricted by the distance and available means of transport. Similar to [6], our mobility profiles do not consider the fact that users may behave distinctively in various time periods. For instance, an accountant works in the shopping mall where he usually does shopping on weekends. Obviously, he will stay much longer on weekdays than on weekends because the purposes of the visits are different. This can be solved by constructing separate mobility profiles for different time periods.

The starting PoI of an activity trajectory is determined by two probability distribution: $\pi_u(t\,|\,p)$ and $\mathrm{Pr}_u(p)$. The former is the probability density of the user entering PoI $p$ at time $t$ while the latter is the probability of the user visiting the PoI.

Given user $u$'s mobility profile, we can calculate the probability density of $a_u^{t,t'} = (\langle u, p_1, [t_b^1, t_e^1]\rangle, \ldots, \langle u, p_k, [t_b^k, t_e^k]\rangle)$ can be expressed as follows:

$$\begin{aligned} f(a_u^{t,t'}\,|\,\mathcal{K}_u) = {} & \pi_u(t\,|\,p_1)\mathrm{Pr}_u(p_1) \\ & \cdot \Big( \prod_{1<i\leq k} (\mathrm{Pr}_u(p_i\,|\,p_{i-1}) \cdot \Gamma_{tran}(p_{i-1}, p_i, t_b^i - t_e^{i-1})) \\ & \cdot \prod_{1\leq i\leq k} \Gamma_{stay}(p_i, t_e^i - t_b^i). \end{aligned}$$

Note that in the rest of this paper, we use $f(\cdot\,|\,\cdot)$ to denote the conditional probability density of a continuous variable and $\mathrm{Pr}(\cdot\,|\,\cdot)$ to denote the conditional probability distribution of a discrete variable for the purpose of being concise.

**Request issuing patterns.** Users have some patterns with respect to when and where they prefer to request LBSs. For instance, the local-search services of nearby restaurants are usually requested during lunch or dinner time from residential areas. Especially, with the recent development of LBS leads to new features, two of which can be described as follows: (i) a user may request LBSs at any time point with certain preference; and (ii) the time interval between two consecutive requests issued in a PoI is not uniformly distributed. This indicates that the issuing time of a request is related to that of the previous request issued in the same PoI. These two observations lead to the following new approach to the modelling of time when users issue LBS requests.

In the extension of the framework in [6] to sporadic LBSs [16], a binary probability distribution is assigned to each time point and it governs the issuing of an LBS request at this time point. This model is reasonable for continuous LBSs where users issue requests independently. In our framework we relax the assumption and propose a more general model. We allow a user to issue a request at any time during his stay in a PoI and we assume that the process of issuing requests is controlled by certain continuous probability density function.

We model the time between two consecutive LBS requests from the same PoI to be exponentially distributed with parameter $\lambda$ with the shift $\Delta$:

$$f_{int}(t\,|\,u) = \begin{cases} \lambda \cdot \mathrm{e}^{-\lambda(t-\Delta)} & t \geq \Delta \\ 0 & t < \Delta \end{cases}$$

Notice that if the available knowledge advocates the use of different distributions, the calculation presented in the remaining of this section can be easily adapted.

Suppose that user $u$ enters a single PoI $p$ at time instance $t_b$ and exits it at time instance $t_e$. Furthermore, we suppose that the user is assigned a pseudonym $u'$ in accordance with some anonymising strategy $\sigma$. With these information, we can calculate the probability density that user $u$, anonymised as $u'$, issues a sequence of observed events $o_{u'}^{t_b,t_e} = (\langle u', t_1, r_1\rangle, \ldots, \langle u', t_k, r_k\rangle)$ within the time interval $[t_b, t_e]$ of the user's presence in PoI $p$:

$$\begin{aligned} & f(o_{u'}^{t_b,t_e}\,|\,p, \sigma(u) = u', \mathcal{K}_u) \\ & = \underbrace{f_{int}(t_1 - t_b\,|\,u) \cdot \prod_{i=2}^{k} f_{int}(t_i - t_{i-1}\,|\,u)}_{\textbf{Part I}} \cdot \underbrace{c(t_e - t_k)}_{\textbf{Part II}} \cdot \\ & \underbrace{\sum_{\ell\in p} obf(r\,|\,\ell) \cdot \prod_{2\leq i\leq k} \sum_{\ell\in p} \mathrm{Pr}(\ell\,|\,p) \cdot obf(r_k\,|\,\ell_k)}_{\textbf{Part III}}, \end{aligned}$$

where

$$c(t) = \begin{cases} \mathrm{e}^{-\lambda(t-\Delta)} & t \geq \Delta \\ 1 & t < \Delta \end{cases}.$$

The above calculation can be divided into three parts which are labelled by numbers. The first two parts are to calculate the probability density that user $u$ exposed *exactly* $k$ locations at $t_1, \ldots, t_k$ during the stay at $p$. Specifically, in Part I, we calculate the probability density of user $u$ issuing $k$ requests within the time interval $[t_b, t_e]$ at time points $t_1, \ldots, t_k$. Part II gives the probability that no requests are issued in the remaining time interval $[t_k, t_e]$. Parts I and II together provide the probability density of user $u$ issuing exactly $k$ requests within the time interval $[t_b, t_e]$ at time points $t_1, \ldots, t_k$. Part III is the joint probability that the location pseudonyms $r_i'$ $(1 \leq i \leq k)$ are output by the deployed obfuscating LPPM. The probability $\mathrm{Pr}(\ell\,|\,p)$ is the likelihood that user $u$ is located at $\ell$ given that he is in PoI $p$. If no further information is available, we assume that $\mathrm{Pr}(\ell\,|\,p) = \frac{1}{|p|}$, i.e., we assume a uniform distribution on all $\ell$ in $p$.

## V. Activity Tracking Attack

In this section, we implement a *tracking attack* in which the adversary tries to infer the *most likely* activity trajectories of users using user profiles in our new model.

In the attack, we assume that the adversary has learnt the observed trajectories of all users in $\mathcal{U}$ in the time period $[t, t']$, i.e., $o^{t,t'}$. Furthermore, we assume that all users are in a PoI at both time $t$ and $t'$. We make this assumption for two reasons: (i) to give a concise presentation; (ii) to simulate the adversary's curiosity of users' daily activities which in most of cases start and end in PoIs, e.g., home. With users' profiles constructed based on users' behaviour in this period, we can approximately interpret that users enter a PoI at $t$ and exits a PoI at $t'$. Our attack presented in this section can also be generalised to more generic situations where users' states at $t$

and $t'$ are not known. However, this will increase calculation complexity because we have to consider three possible user states at the two ends: just entering or exiting a PoI, in a PoI and during transition. Formally, this attack can be formulated as the following optimisation problem:

$$\arg\max_{a,\sigma} f(a,\sigma \,|\, o^{t,t'})$$

where $a$ represents all possible sets of activity trajectories of users in $\mathcal{U}$. As mentioned above, we user $f(x\,|\,y)$ to denote the conditional probability density of $x$ when $y$ is available. Thus $f(a,\sigma\,|\,o^{t,t'})$ is the probability density that users travel activity trajectories $a$ when $o^{t,t'}$ is observed. In this attack, we target at two types of information: users' activity trajectories and the bijective mapping relation between users' activity trajectories and their observed trajectories.

With this understanding, we split this tracking attack into two steps. At the first step, we calculate the most likely anonymising strategy called *de-anonymising* attack. At the second step, we proceed calculating users' most probable activity trajectories given the anonymising strategy obtained in the first step. Note the independence of the two steps, which facilitates the split.

### A. De-anonymisation

The purpose of de-anonymising attack is to find the most likely owner of each observed trajectory in $o^{t,t'}$. In other words, the goal of the adversary is to find the most probable anonymising mapping function $\sigma^*$ . This can be formulated as follows: $\sigma^* = \arg\max_\sigma \ \Pr(\sigma\,|\,o^{t,t'},\mathcal{K})$. We use $f(o^{t,t'}\,|\,\sigma,\mathcal{K})$ to represent the density of $o^{t,t'}$ given the anonymising strategy $\sigma$ and the adversary with the knowledge $\mathcal{K}$. By applying the Bayesian theorem, we have that

$$\Pr(\sigma\,|\,o^{t,t'},\mathcal{K}) = \frac{f(o^{t,t'}\,|\,\sigma,\mathcal{K})\cdot\Pr(\sigma)}{f(o^{t,t'}\,|\,\mathcal{K})}. \tag{2}$$

Note that since the denominator $f(o^{t,t'}\,|\,\mathcal{K})$ can be seen as a proportionality factor, it is independent of $\sigma$ and can thus be considered constant. Due to the assumption that the choice of the anonymising strategy follows a uniform distribution, we have that $\Pr(\sigma)$ is $\frac{1}{|\mathcal{U}|!}$. Thus, the optimisation is reduced to finding $\sigma$ that maximises $f(o^{t,t'}\mid\sigma,\mathcal{K})$. Since users are independent of each other when travelling and their exposure events are anonymised and obfuscated independently, $f(o^{t,t'}\mid\sigma,\mathcal{K})$ can be factorised into a product of probability density functions as follows:

$$f(o^{t,t'}\,|\,\sigma,\mathcal{K}) = \prod_{u'\in\mathcal{U}'} f(o_{u'}^{t,t'}\,|\,\sigma,\mathcal{K}_{\sigma^{-1}(u')}).$$

Notice that $\sigma$ is a bijection between $U$ and $U'$, so $\sigma^{-1}$ is well defined. Thus, the problem is further reduced to calculating the mapping function $\sigma^*$ which maximises the above product.

This problem can be solved as the *minimum weight assignment problem*, which assigns a single task to each agent in a group and guarantees the minimum total cost. This is because the de-anonymising attack can be interpreted as assigning pseudonyms in $\mathcal{U}'$ to users in $\mathcal{U}$. If we take $-\log f(o_{u'}^{t,t'}\,|\,\mathcal{K}_u)$ to be the cost of assigning $u'$ to $u$, then the existing solutions to the minimum weight assignment problem can be used due to the fact that $\arg\max_\sigma f(o^{t,t'}\mid\sigma,\mathcal{K})$ is equivalent to $\arg\min_\sigma -\log f(o^{t,t'}\,|\,\sigma,\mathcal{K})$.

This indicates that for each user $u\in\mathcal{U}$, we have to calculate the density of each observed trajectory $o_{u'}^{t,t'}\in o^{t,t'}$ given user $u$'s profiles, i.e., $f(o_{u'}^{t,t'}\,|\,\mathcal{K}_u)$. In the following, we proceed to present an efficient method to calculate users' patterns with respect to their observed trajectories, i.e., $f(o_{u'}^{t,t'}\,|\,\mathcal{K}_u)$.

Let $\mathcal{A}$ be the set of all possible activity trajectories of user $u$ in the time interval $[t,t']$. Then we have

$$f(o_{u'}^{t,t'}\,|\,\mathcal{K}_u) = \int_{\alpha\in\mathcal{A}} f(\alpha, o_{u'}^{t,t'}\,|\,\mathcal{K}_u)\mathrm{d}\alpha$$

which marginalises out activity trajectories. We proceed to show how to calculate the marginalisation.

We notice that $f(\alpha, o_{u'}^{t,t'}\,|\,\mathcal{K}_u)$ is zero when the activity trajectory $\alpha$ is not compatible with $o_{u'}$, where the compatibility is understood as follows. We say that an activity trajectory $\alpha$ is *compatible* with $o_{u'}$ if and only if for the time point given by any observed event in $o_{u'}$, user $u$ is at a PoI and not in transition between two PoIs. To make further computation efficient, we introduce a scheme to consider only the activity trajectories that are compatible with $o_{u'}$.

Let $\mathcal{S}$ be the set of all sequences of PoIs that user $u$ could potentially visit in $[t,t']$ which are allowed by the minimum time required to move between successive PoIs. If we use $N(s)$ to denote the length of sequence $s$ and $p_i^s$ be the $i$th PoI in $s$, then $\mathcal{S} = \big\{s \mid \forall_{i=1,\ldots,N(s)}\, p_i^s \in \mathcal{P}, \sum_{i=1}^{N(s)-1} \delta_{min}(p_i^s, p_{i+1}^s) \le t'-t\big\}$. Two remarks are in place. First, for each user $u\in\mathcal{U}$ and $[t,t']$, the set $\mathcal{S}$ is finite since user $u$ has finite PoIs and the minimum transition time between any two PoIs is non-zero. Second, by the definition of activity trajectories, the sequence of PoIs visited by $u$ within any activity trajectory is contained in $\mathcal{S}$.

We proceed to consider how an observed trajectory $o_{u'}^{t,t'}$ could be obtained given that the user visited a sequence of PoIs $s\in\mathcal{S}$. The following restrictions are in place: (i) each observed event in $o_{u'}^{t,t'}$ is issued from a PoI in $s$; (ii) any two consecutive observed events are issued either from the same PoI or the second event is issued from some subsequent PoI in $s$; (iii) $s$ may contain PoIs where no events are issued.

With these restrictions, we can decompose $o_{u'}^{t,t'}$ into $N(s)$ disjoint blocks of contiguous observed events. The $i$th block is the sequence of all observed events issued from PoI $p_i^s$. We use $\Xi_i^{s,o_{u'}^{t,t'}}$ to denote the $i$th block and $\Xi^{s,o_{u'}^{t,t'}} = (\Xi_1^{s,o_{u'}^{t,t'}}, \Xi_2^{s,o_{u'}^{t,t'}}, \ldots, \Xi_{N(s)}^{s,o_{u'}^{t,t'}})$ is a *decomposition* of $o_{u'}^{t,t'}$ with respect to $s$. Note that for the same $s$, $o_{u'}^{t,t'}$ usually has a number of different decompositions with respect to $s$ which is actually exponential in the number of observed events.

We say that an activity trajectory $\alpha$ *complies* with PoI sequence $s$ and $\Xi^{s,o_{u'}^{t,t'}}$ if the $i$th PoI in $\alpha$ is equal to $p_i^s$ and user $u$ enters it before issuing the first request in $\Xi_i^{s,o_{u'}^{t,t'}}$

and exits it after having issued the last request in $\Xi_i^{s,o_{u'}^{t,t'}}$. Let $t_b^i$ and $t_e^i$ are the entering and exiting time points of the $i$th PoI in $\alpha$, then $t_b^i \leq \min\{t'' \mid \exists \langle u', t'', r\rangle \in \Xi_i^{s,o_{u'}^{t,t'}})\}$ and $t_e^i \geq \max\{t'' \mid \exists \langle u', t'', r\rangle \in \Xi_i^{s,o_{u'}^{t,t'}})\}$. These two conditions lead to a small interval for the entering (exiting) time of each PoI in $s$. We use $\mathcal{T}_b^{i,s,o_{u'}^{t,t'}}$ and $\mathcal{T}_e^{i,s,o_{u'}^{t,t'}}$ $(1 \leq i \leq N(s))$ to denote such time intervals for the entering and exiting time of PoI $p_i^s$, respectively. Let $\alpha$ be an activity trajectory that complies with $\Xi^{s,o_{u'}^{t,t'}}$ and $s = (p_1, \ldots, p_k)$. In Equation 3, we calculate the joint probability density of user $u$ travelling $\alpha$ (Part I) and generating the observed trajectory (Part II):

$$f(\alpha, o_{u'}^{t,t'} \,|\, \mathcal{K}_u) = \underbrace{f(a_u \,|\, \mathcal{K}_u)}_{\textbf{Part I}} \cdot \underbrace{\prod_{i=1}^{N(s)} f(\Xi_i^{s,o_{u'}^{t,t'}} \,|\, p_i^s, \mathcal{K}_u)}_{\textbf{Part II}}. \tag{3}$$

With the above density function, in the following, we present a method to marginalise over all activity trajectories and obtain our target $f(o_{u'}^{t,t'} \mid \mathcal{K}_u)$, i.e., the density function of user $u$ issuing an observed trajectory.

We start with constructing the activity trajectories compatible with an observed trajectory $o_{u'}^{t,t'}$. We observe that the following two inference rules hold: (i) if $\alpha$ and $o_{u'}^{t,t'}$ are compatible, there exists $s$ and $\Xi^{s,o_{u'}^{t,t'}}$ that $\alpha$ complies with; (ii) if $\alpha$ *complies* with $s$ and $\Xi^{s,o_{u'}^{t,t'}}$, then $\alpha$ is compatible with $o_{u'}^{t,t'}$. These two rules allow us to construct all activity trajectories that are compatible with the observed trajectory by considering (i) each $s \in \mathcal{S}$; (ii) each possible decomposition of the observed trajectory with respect to $s$; and (iii) every possible combination of the entering and exiting time points within the time intervals determined by $s$ and the decomposition. Therefore, we can write

$$\begin{aligned} & f(o_{u'}^{t,t'} \,|\, \mathcal{K}_u) \\ = & \sum_{s \in \mathcal{S}} \sum_{\Xi^{s,o_{u'}^{t,t'}}} \int_{\mathcal{T}_b^{1,s,o_{u'}^{t,t'}} \times \mathcal{T}_e^{1,s,o_{u'}^{t,t'}} \times \ldots \times \mathcal{T}_b^{N(s),s,o_{u'}^{t,t'}} \times \mathcal{T}_e^{N(s),s,o_{u'}^{t,t'}}} \\ & \quad f(\alpha, o_{u'}^{t,t'} \,|\, \mathcal{K}_u) \mathrm{d}t_e^{N(s)} \; \mathrm{d}t_b^{N(s)} \; \ldots \mathrm{d}t_e^1 \; \mathrm{d}t_b^1 \; . \end{aligned} \tag{4}$$

### B. De-obfuscation

In this step, the adversary's purpose is to reconstruct a user's most likely activity trajectory when his observed trajectory is learnt. Formally, $\arg\max_{a_u} f(a_u \mid o_{\sigma(u)}^{t,t'}, \mathcal{K}_u)$. With Bayesian theorem, we have

$$f(a_u \,|\, o_{\sigma(u)}^{t,t'}, \mathcal{K}_u) = \frac{f(o_{\sigma(u)}^{t,t'} \,|\, a_u, \mathcal{K}_u) \cdot f(a_u \,|\, \mathcal{K}_u)}{f(o_{\sigma(u)}^{t,t'} \,|\, \mathcal{K}_u)}. \tag{5}$$

Since the denominator is a constant for all activity trajectories, the problem can be reduced to find the activity trajectory that maximises the nominator. This can be seen as global optimisation problem which can be formulated as follows:

$$\mathop{\arg\max}_{s \in \mathcal{S}, \Xi^{s,o_{u'}^{t,t'}}} \quad \mathop{\arg\max}_{(t_b^1, t_e^1, \ldots, t_b^{N(s)}, t_e^{N(s)})} \quad f(\alpha, o_{u'}^{t,t'} \,|\, \mathcal{K}_u)$$

where $\alpha = (\langle u, p_1^s, [t_b^1, t_e^1]\rangle, \ldots, \langle u, p_{N(s)}^s, [t_b^{N(s)}, t_e^{N(s)}]\rangle)$ and $t_b^i \in \mathcal{T}_b^{i,s,o_{u'}^{t,t'}}$ and $t_e^i \in \mathcal{T}_e^{i,s,o_{u'}^{t,t'}}$ for all $1 \leq i \leq N(s)$. To solve this optimisation problem, first, for each $s \in \mathcal{S}$, we consider all of its possible decompositions. Then for each pair $(s, \Xi^{s,o_{u'}^{t,t'}})$, we search for a sequence of entering and exiting time points that maximise the joint probability density function and record the largest probability density. To find the (approximate) optimal sequence, we can refer to a number of algorithms. We use *Simulated Annealing* in our implementation. Last, we choose the pair with the largest probability density and with the corresponding time sequence, in this way the optimal activity trajectory is constructed.

## VI. Validation

In this section, we pursue two goals: (i) constructing user profiles and identify their main features in users' real-life movements; (ii) evaluating the effectiveness of our new activity tracking attack. With respect to the former goal, we focus on users' mobility profiles. We explore a real-life GPS trajectory dataset to justify our model for user mobility profiles which is collected in the *Geolife* project of Microsoft Research Asia [19]. The dataset consists of 17,621 daily trajectories from 182 users in over five years which mainly took place in Beijing, China. The trajectories are collected in a high frequency and over 90% of the locations are recorded in less than every 5 seconds. We select ten representative users based on the number of their collected trajectories in our validation. On average, each user has over 200 daily trajectories.

**Constructing mobility profiles.** We adopt the method and the tool by Chen et al. [8], [9], [20] to dynamically compute users' PoIs. Their method explores the heuristic that a PoI is usually a small region where a user tends to stay for certain amount of time. To extract the transition matrix, we use the maximum likelihood estimation method as described in [21]. In Figure 2 we plot the calculated PoIs of the ten users with circles whose sizes represent their area. On average, the area of these PoIs is 0.317 $km^2$ and each user has 28 PoIs. We can see that the introduction of PoIs leads to fewer states in the Markov chains but meanwhile a relatively high precision.

We explore the Gaussian kernel smoothing method to estimate the probability density functions $\Gamma_{stay}$ and $\Gamma_{tran}$ and the shapes of the estimated functions are similar to Gamma distributions. We then fit Gamma distributions to the data and feed the fitted functions to the Chi-square test to evaluate their goodness of fit. The results are mostly positive meaning that the density functions of stay and transition time can be assumed to be gamma distributions.

The *Geolife* dataset does not contain sufficient amount of data to extract complete mobility profiles for *all the selected users*, especially for the probabilistic density functions related

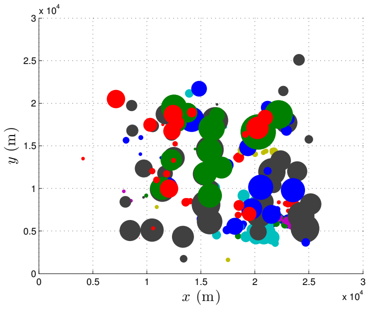


Fig. 2. The distribution of users' PoIs.

to stay time in some PoIs and transition time between some PoI pairs. We thus *partially* simulate users' profiles when the required information is not extractable. This will not impose much impact on our validation since our target focuses on the effectiveness of our tracking attack under the assumption of the availability of user mobility profiles.

**Evaluating the attack.** We implement our activity tracking attack as explained in Section V.

*Experimental setting.* We need to set up our experiments from two perspectives: activity trajectories and observed trajectories. With respect to activity trajectories, for each user we choose 40 of their daily trajectories. The time information of these trajectories are not consistent with user mobility profiles as part of user profiles are simulated rather than extracted directly from the dataset. Thus, we only extract the sequences of PoIs in them and make use of the simulated user profiles to generate the amount of time that users spend in and between the PoIs. The simulated activity trajectories span from 2 to 12 hours depending on the number of PoIs involved. We generate the observed trajectories of the selected users due to the lack of users' exposed trajectories. Given an activity trajectory, we calculate the corresponding exposed trajectory based on the owner's request issuing rate. In order to analyse the influence of the number of issued requests, we generate two exposed trajectories for an activity trajectory with lengths of one and three, respectively. For obfuscating LPPMs, we consider the cloaking mechanism which reduces the precision of the coordinates of the locations in the exposed trajectories due to its popularity. In our experiments, we set two precisions, namely, 0.001 and 0.01, and examine the sensitivity of location privacy to them. These two precisions enlarge a position to a region with area of about 0.02 to 2.25 $km^2$, respectively.

*Location privacy metric.* Recall that in our framework, we make use of the estimation error of the adversary to evaluate the effectiveness of attacks against LPPMs. As we split our attack into de-anonymisation and de-obfuscation attacks, we evaluate their effectiveness separately. The performance of our tracking attack can be guaranteed if they are both effective. For de-anonymisation, given a set of observed trajectories, each of which corresponds to a unique user, we define the adversary's estimation error as the percentage of the user identities that are not correctly assigned to the observed trajectories. For any observed trajectory and its owner, the de-obfuscation attack outputs the most probable activity trajectory. Let $\alpha^*$ be the calculated activity trajectory. According to Equation 1, the location privacy is determined by the distance between $\alpha^*$ and user $u$'s real activity trajectory $\alpha$. Intuitively, if we simply treat the positions during transition as a specific PoI, then two activity trajectories are equivalent if and only if two users are in the same PoIs at any time. This leads us to use the proportion of time when two users are not in the same PoI to assess the de-obfuscation attack.

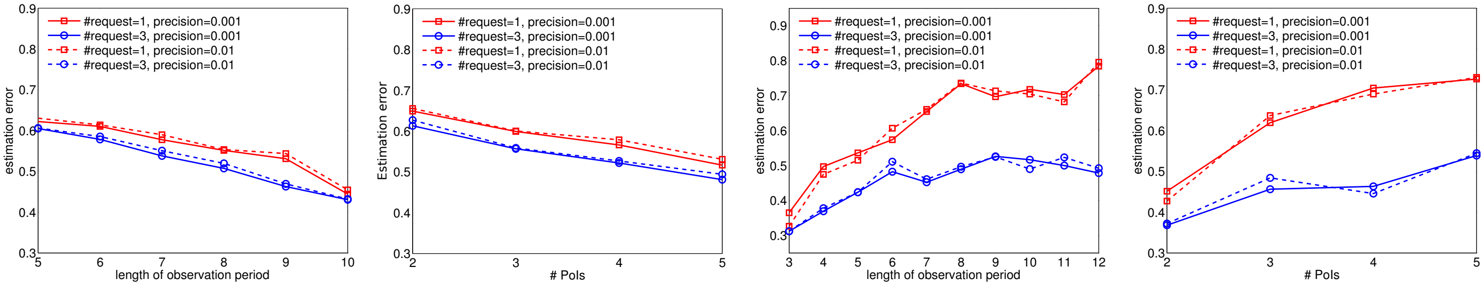
TABLE I
THE DISTRIBUTION OF ESTIMATION ERROR.

| Attacks | $\leq 0.2$ | $0.2-0.4$ | $0.4-0.6$ | $0.6-0.8$ | $\geq 0.8$ |
|---|---|---|---|---|---|
| de-anonym. | 0.01 | 0.16 | 0.46 | 0.32 | 0.04 |
| de-obfuscation | 0.11 | 0.18 | 0.26 | 0.31 | 0.14 |

*Experimental results.* In this section, we present the experimental results of the two attacks from two aspects: their effectiveness and their sensitivity to other factors. In the attacks, we assume that the adversary has access to the time periods when users travelled their activity trajectories, which are called *observation periods* in the following discussion. We construct 5,000 daily observations for the selected users in each of which a user has a unique observed trajectory and feed them to the de-anonymisation attack. The de-obfuscation attack aims to infer information from observed trajectories whose owners have been already learnt.

Table I summarises the percentages of samples whose estimation errors fall into different intervals when the reduced precision is set to 0.001. Generally, most of the estimation errors range from 0.2 to 0.8. We can see that for over 60% of the samples, the adversary has at least a probability of 0.4 to get users' right pseudonyms. Our obfuscating attack can even ensure at least a correctness of 0.6 for 30% users. Therefore, from these statistics, we can see that our attacks are rather effective even if they are combined in our tracking attack.

We study four parameters that may have impact on users' location privacy: the number of issued LBS requests, the length of observation periods, the number of visited PoIs and the reduced precision. Our first observation is that more issued requests lead to more privacy loss. The estimation errors when only one request is issued (annotated by red curve) are always larger than those when three requests are issued. Our second observation is that the length of observation periods has different influence on the effectiveness of our attacks (see Figures 3(a) and 3(c)). We group observed trajectories according to the (mean) length of their observation periods and depict the mean estimation error of the trajectories in each case. In the de-anonymisation attack, the mean estimation error decreases along with with the length of observation periods. This can be explained by the fact that users are more likely to travel distinctive trajectories in a longer time period. Meanwhile, it is opposite in the de-obfuscation attack. The estimation error increases significantly when users travel a longer time. This is because in a longer period, users have more flexibility to arrange their visits to PoIs as well as the



(a) Mean E.E. vs. length of period (DA)

(b) Mean E.E. vs. # PoIs (DA)

(c) Mean E.E. vs. length of period (DF)

(d) Mean E.E. vs. # PoIs (DF)

Fig. 3. Estimation error (E.E.) of the attack (DA for de-anonymisation and DF fro de-obfuscation).

stay time in them. Our third observation (see Figures 3(b) and 3(d)) is that in both attacks the number of visited PoIs has a similar impact to that of the length of observation periods. This is because of the fact that a longer period indicates more PoIs that can be visited. Last, we observe that the increase of reduced precision from 0.001 to 0.01 does not have a visible improvement for users' location privacy. This is because the exploration of PoIs has decreased the adversary's uncertainty eliminated the impact of the reduced precision, especially when the precision is not reduced significantly enough.

## VII. Conclusion

In this paper, we proposed a new model for user profiles and a new tracking attack with the aim to provide the adversary with more information to accurately infer users' activities. Namely, the places where users visits and the entering and stay time of these places can be obtained through our attack in a direct way. Other attacks on location privacy in [6] can be formalised in our framework as well [22]. Compared to existing works on user mobility profiles, our model can describe users' patterns with respect to mobility and requesting LBS in continuous time. By making use of PoIs, our attack has a reasonable efficiency and can be extended to cover more general cases with little increase in computational overhead.

## References

[1] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady, "Enhancing security and privacy in traffic-monitoring systems," *IEEE Pervasive Computing*, vol. 5, no. 4, pp. 38–46, 2006.

[2] M. Decker, "Location privacy - an overview," in *Proc. 7th International Conference on Mobile Business (ICMB)*. IEEE CS, 2008, pp. 221–230.

[3] B. Lee, J. Oh, H. Yu, and J. Kim, "Protecting location privacy using location semantics," in *Proc. 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. ACM, 2011, pp. 1289–1297.

[4] Y. Wang, D. Xu, X. He, C. Zhang, F. Li, and B. Xu, "L2p2: Location-aware location privacy protection for location-based services," in *Proc. 31st Annual IEEE International Conference on Computer Communications (INFOCOM)*. IEEE CS, 2012, pp. 1996–2004.

[5] S. Mascetti, L. Bertolaja, and C. Bettini, "A practical location privacy attack in proximity services," in *Proc. 14th IEEE International Conference on Mobile Data Management (MDM)*. IEEE CS, 2013, pp. 87–96.

[6] R. Shokri, G. Theodorakopoulos, J.-Y. L. Boudec, and J.-P. Hubaux, "Quantifying location privacy," in *Proc. 32nd IEEE Symposium on Security and Privacy (S&P)*. IEEE CS, 2011.

[7] X. Xiao, Y. Zheng, Q. Luo, and X. Xie, "Finding similar users using category-based location history," in *Proc. 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS)*. ACM, 2010, pp. 442–445.

[8] X. Chen, J. Pang, and R. Xue, "Constructing and comparing user mobility profiles for location-based services," in *Proc. ACM Symposium on Applied Computing (SAC)*. ACM, 2013, pp. 264–269.

[9] ——, "Constructing and comparing user mobility profiles," *ACM Transactions on the Web*, vol. 8, no. 4, p. 21, 2014.

[10] X. Chen, R. Lv, X. Ma, and J. Pang, "Measuring user similarity with trajectory patterns: Principles and new metrics," in *Proc. 16th Asia-Pacific Web Conference (APWeb)*, ser. LNCS, vol. 8709. Springer, 2014, pp. 437–448.

[11] M. Ye, D. Shou, W.-C. Lee, P. Yin, and K. Janowicz, "On the semantic annotation of places in location-based social networks," in *Proc. 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. ACM, 2011, pp. 520–528.

[12] A. Khoshgozaran, C. Shahabi, and H. Shirani-Mehr, "Location privacy: going beyond K-anonymity, cloaking and anonymizers," *Knowledge and Information Systems*, vol. 26, no. 3, pp. 435–465, 2011.

[13] J. T. Meyerowitz and R. R. Choudhury, "Hiding stars with fireworks: location privacy through camouflage," in *Proc. 15th Annual International Conference on Mobile Computing and Networking (MOBICOM)*. ACM, 2009, pp. 345–356.

[14] P. Golle and K. Partridge, "On the anonymity of home/work location pairs," in *Proc. 7th International Conference on Pervasive Computing (Pervasive)*, ser. LNCS, vol. 5538. Springer, 2009, pp. 390–397.

[15] Y. D. Mulder, G. Danezis, L. Batina, and B. Preneel, "Identification via location-profiling in GSM networks," in *Proc. 7th ACM Workshop on Privacy in the Electronic Society (WPES)*. ACM, 2008, pp. 23–32.

[16] R. Shokri, G. Theodorakopoulos, G. Danezis, J.-P. Hubaux, and J.-Y. L. Boudec, "Quantifying location privacy: The case of sporadic location exposure," in *Proc. 11th International Symposium on Privacy Enhancing Technologies (PETS)*, ser. LNCS, vol. 6794. Springer, 2011, pp. 57–76.

[17] G. Acar, M. Juarez, N. Nikiforakis, C. Diaz, S. Grses, F. Piessens, and B. Preneel, "Optimal sporadic location privacy preserving systems in presence of bandwidth constraints," in *Proc. 12th ACM Workshop on Privacy in the Electronic Society (WPES)*. ACM, 2013, pp. 167–178.

[18] R. Shokri, G. Theodorakopoulos, C. Troncoso, J.-P. Hubaux, and J.-Y. L. Boudec, "Protecting location privacy: optimal strategy against localization attacks," in *Proc. 19th ACM Conference on Computer and Communications Security (CCS)*. ACM, 2012, pp. 617–627.

[19] Y. Zheng, L. Wang, R. Zhang, X. Xie, and W.-Y. Ma, "GeoLife: Managing and understanding your past life over maps," in *Proc. 9th International Conference on Mobile Data Management (MDM)*. IEEE CS, 2008, pp. 211–212.

[20] X. Chen, P. Kordy, R. Lv, and J. Pang, "MinUS: Mining user similarity with trajectory patterns," in *Proc. 7th European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD)*, ser. LNCS, vol. 8726. Springer, 2014, pp. 436–439.

[21] F. Chierichetti, R. Kumar, P. Raghavan, and T. Sarlós, "Are web users really Markovian?" in *Proc. 21st World Wide Web Conference (WWW)*. ACM, 2012, pp. 609–618.

[22] X. Chen, A. Mizera, and J. Pang, "Quantifying location privacy revisited: Preliminary report," University of Luxembourg, Tech. Rep., 2014, available at http://satoss.uni.lu/jun/papers/locpriv.pdf.