MooD: MObility Data Privacy as Orphan Disease -Experimentation and Deployment Paper

Besma Khalfoun, Mohamed Maouche, Sonia Ben Mokhtar, Sara Bouchenak

MooD: *MO*bility Data Privacy as *O*rphan *D*isease

– Experimentation and Deployment Paper –

Besma Khalfoun[1,2], Mohamed Maouche[1], Sonia Ben Mokhtar[1], Sara Bouchenak[1]
firstname.lastname@insa-lyon.fr
[1]Universite de Lyon, CNRS. INSA Lyon, LIRIS, UMR5250, F69622, France
[2]Ecole Nationale Superieure d'Informatique, Algiers, Algeria

Abstract

With the increasing development of handheld devices, Location Based Services (LBSs) became very popular in facilitating users' daily life with a broad range of applications (e.g. traffic monitoring, geo-located search, geo-gaming). However, several studies have shown that the collected mobility data may reveal sensitive information about end-users such as their home and workplaces, their gender, political, religious or sexual preferences. To overcome these threats, many Location Privacy Protection Mechanisms (LPPMs) were proposed in the literature. While the existing LPPMs try to protect most of the users in mobility datasets, there is usually a subset of users who are not protected by any of the existing LPPMs. By analogy to medical research, there are orphan diseases, for which the medical community is still looking for a remedy. In this paper, we present MooD, a fine-grained multi-LPPM user-centric solution whose main objective is to find a *treatment* to mobile users' orphan disease by protecting them from re-identification attacks. Our experiments are conducted on four real world datasets. The results show that MooD outperforms its competitors, and the amount of user mobility data it is able to protect is in the range between 97.5% to 100% on the various datasets.

CCS Concepts • **Security and privacy** → *Pseudonymity, anonymity and untraceability.*

Keywords Mobility Data, User Re-identification, Location Privacy Protection Mechanism, User-Centric Protection, Data Privacy

ACM Reference Format:
Besma Khalfoun[1,2], Mohamed Maouche[1], Sonia Ben Mokhtar[1], Sara Bouchenak[1]. 2019. MooD: *MO*bility Data Privacy as *O*rphan *D*isease: – Experimentation and Deployment Paper –. In *Middleware '19: Middleware '19: 20th International Middleware Conference, December 8–13, 2019, Davis, CA, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3361525.3361542

Middleware '19, December 8–13, 2019, Davis, CA, USA

ACM ISBN 978-1-4503-7009-7/19/12…$15.00
https://doi.org/10.1145/3361525.3361542

1 Introduction

Nowadays, the proliferation of mobile devices embedding GPS chips (e.g. smartphones, tablets, smartwatches) has significantly contributed to the development of geolocated services, also named Location Based Services (LBSs). These services are useful for users' daily life as they allow them to localize nearby friends, discover their environment and request for places whenever they like and wherever they are. The downside is that huge amounts of information regarding users' locations are being gathered and stored by third party services. These services or other entities that may have access to the collected data (e.g., accidentally or through an attack) may exploit it fraudulently in order to infer and reveal sensitive information about individuals (e.g., home address, workplace, religious beliefs, sexual preferences, political orientations, social relationships). The most common threats include: (1) re-identification attacks where an anonymous mobility trace is re-associated to its originating user based on previously recorded data [11] [16] [33], (2) mobility prediction where users' next moves are anticipated [29] [15] [3], (3) extraction of user's Points of Interest (POI) (e.g., home, workplace, etc.) [35] and (4) inference of social relationships (e.g., friends, coworkers, etc.) [6] [32].

To tackle these threats, many Location Privacy Protection Mechanisms (LPPMs) were proposed in the literature. LPPMs protect user location information by relying on a wide variety of techniques such as perturbation, generalization and fake data generation [17].

To evaluate the effectiveness of LPPMs, a variety of privacy metrics are usually used and the resilience against re-identification attacks is one of them. Considering a protected mobility trace (i.e., a raw trace to which a given LPPM is applied), a re-identification attack tries to link the protected mobility trace to its owner based on past unprotected mobility data that the attacker has access to. The more an LPPM is able to protect against re-identification attacks, the better. There exist a variety of re-identification attacks in the state-of-the-art literature that differ in the way they model

and analyse user mobility (e.g., PIT-attack [16] uses Mobility Markov Chains and AP-attack [22] uses heatmaps to model users' mobility). However, when LPPMs are evaluated against re-identification attacks the focus is generally put on the protection of *the crowd*, i.e., protecting the larger proportion of users possible, and little attention is given to users that remain unprotected. Considering a set of state-of-the-art attacks and LPPMs at the disposal of a data security expert aiming at the protection of a given dataset, the question that the latter may ask is: *What should be done with mobility traces that are subject to re-identification despite the use of LPPMs?*. A straightforward, and safe solution that the expert may adopt is to delete these vulnerable mobility traces from the protected dataset. However, this solution would engender the loss of large data portions. To assess this loss, we performed an experiment in which we applied three state-of-the-art LPPMs to protect four mobility datasets.

On the protected datasets, we ran three state-of-the-art re-identification attacks and we removed from the protected dataset, the traces that were re-identified by at least one of the attacks. The results of this experiment with Geo-I [4], TRL 18 and HMC [23] (described further in Section 4.1.2 and 5) show a data loss of 42% on average and that can reach 95% in the most vulnerable dataset. The detailed results of this experiment are presented in the problem illustration (Section 2.4).

In this paper, we present MooD (*MO*bility Data Privacy as *O*rphan *D*isease), a user centric approach to enforce location privacy using multiple LPPMs aiming at the protection of *orphan* users, i.e, users that are not protected against re-identification attacks while using any of the existing LPPMs. The originality of MooD is that it combines off-the-shelf LPPMs and applies a fine-grained protection. The LPPMs' combination is realized with the application of various LPPMs on the same trace in the form of function composition, while the fine-grained protection implies the application of various LPPMs on contiguous sub-traces. MooD's mechanisms are driven by the resilience to state-of-the-art re-identification attacks and the data utility metrics set by the data security expert.

We evaluate MooD by applying it to four real life mobility datasets and comparing its performance to the application of individual and hybrid LPPMs [22]. The results of our experiments show that MooD is able to protect users' data in a range between 97.5% and 100% on the four datasets while the best competitor (HybridLPPM) protects users' data in a range between 64% and 95% on the same datasets.

The remainder of this paper is structured as follows. First, we present in Section 2 a background on user re-identification attacks and location privacy protection mechanisms and we illustrate the handled problem. Then, we describe the design principles of our system in section 3. Further, in section 4, we proceed to the experimental evaluation of our solution. Finally, we present the related work and conclude in section 5 and 6 respectively.

2 Background and Problem Statement

In this section, we describe mobility traces, re-identification attacks and LPPMs (Section 2.1, 2.2 and 2.3, respectively). Then we present an experiment we did to illustrate our problem (Section 2.4).

2.1 Mobility Traces

A mobility trace is a sequence of spatio-temporal records $r = (lat, lng, t)$ associated to a given user, where lat and lng correspond to the latitude and the longitude of GPS coordinates while t is a timestamp. To simplify mobility traces are timeseries (i.e.,$T \in (\mathbb{R}^2 \times \mathbb{R}_+)^*$).

2.2 User Re-identification Attacks

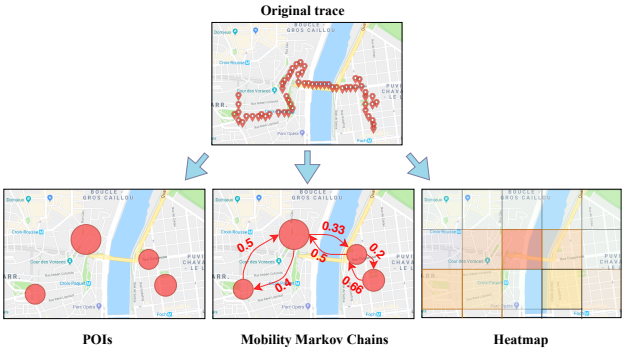


Figure 1. Models of mobility profiles

A user re-identification attack aims at associating a protected mobility trace to its originating user based on users' past mobility. Two phases are necessary to run these attacks: a training phase and an attack phase. In the training phase, the attacker collects non-obfuscated mobility history of known users from several sources and builds users' mobility profiles. Several models have been used in the literature to characterize mobility profiles of users. The most common models include Points of Interests (POIs) (i.e., the set of meaningful places where users spent time), Mobility Markov Chain (MMC) where states are POIs and edges represent the probability transition between states or HeatMaps that aggregate user mobility over time across cells. Figure 1 illustrates the above three models. Then, in the attack phase, the attacker that receives an anonymous mobility trace, tries to re-associate it to the closest user profile among the learned ones.

More formally, considering an anonymous mobility trace T, a set of past mobility traces of known users $\mathbb{H} = \{H_1, H_2, \ldots\}$ where $H_i \in (\mathbb{R}^2 \times \mathbb{R}_+)^*$ and a set of users U, a re-identification attack $\mathcal{A}$ is defined in Equation 1.

$$\begin{array}{rcccc} \mathcal{A} & : & (\mathbb{R}^2 \times \mathbb{R}_+)^* & \rightarrow & \mathbb{U} \\ & & T & \mapsto & \mathcal{A}(T, \mathbb{H}) = u_a \end{array} \quad (1)$$

2.3 Location Privacy Protection Mechanisms

In order to mitigate location privacy threats, Location Privacy Protection Mechanisms (LPPMs) have been introduced in the literature. LPPMs operate modifications on raw mobility data in order to offer end-users a set of privacy guarantees. More formally, a protection mechanism $\mathcal{L}$ is defined in Equation 2, it takes as input a mobility trace T and a set of parameters Y and produces an obfuscated version of the mobility trace as an output.

$$\begin{array}{rcccc} \mathcal{L} & : & (\mathbb{R}^2 \times \mathbb{R}_+)^* & \rightarrow & (\mathbb{R}^2 \times \mathbb{R}_+)^* \\ & & T & \mapsto & \mathcal{L}(\mathrm{Y}, T) = T' \end{array} \quad (2)$$

LPPMs differ in the way they alter the original mobility data and in the guarantees they offer to end-users. These guarantees can be theoretical (e.g., k-anonymity [31], differential privacy [12]) or practical (e.g., the resilience to known attacks).

2.4 Problem Illustration

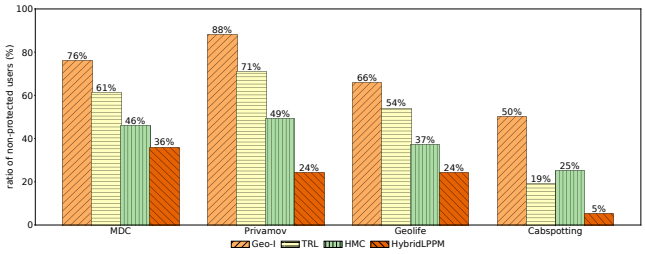


Figure 2. Ratio of non-protected users with state-of-the-art LPPMs and Hybrid LPPM on four real datasets

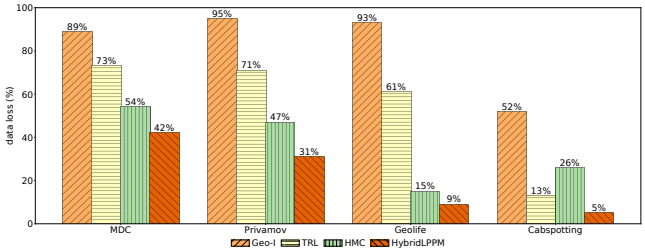


Figure 3. Ratio of Data Loss with state-of-the-art LPPMs and Hybrid LPPM on four real datasets

Consider a data security expert that has to protect a given mobility dataset before its publication. The security expert has access to a set of LPPMs and a set of user re-identification attacks found in the literature. In order to assess the effectiveness of the LPPMs in front of the attacks, the expert may decide to run the re-identification attacks on the protected dataset and choose the LPPM that better protects her original dataset. We performed such an experiment on four real mobility datasets protected using three state-of-the-art LPPMs (i.e., Geo-I 4, TRL [18] and HMC [23]) and a hybrid solution proposed in [22] on which we ran three state-of-the-art attacks (i.e., POI-Attack [27], PIT-Attack [16] and AP-Attack [22]). The details of the used LPPMs and attacks are presented in Section 4. The results of this experiment are depicted in Figure 2. These results show, on each dataset, the number of users for whom at least one of the attacks was able to disclose their identities. From these results, we can see that on all the datasets, there are several users, from 19% to 88% that are not protected in front of re-identification attacks despite the use of single LPPMs.

The question that a data security expert may ask in this situation is *what should be done with these vulnerable portions of the respective datasets?*. A safe answer would be to delete these parts of the datasets in order to prevent eventual user re-identifications that an attacker may perform on the published data. However, this may generate a massive data loss that ranges from 13% to 95% of the overall datasets, as depicted in Figure 3.

A closer look to the protected datasets shows that LPPMs perform differently from one user to another. Hence, a second step considered was to move to a user-centric approach where the hybridLPPM proposed in [22], is applied to each user of the considered datasets. The latter selects an LPPM among a set of LPPMs that resists to re-identification attacks (if any) with the best utility metric, i.e, a spatial-temporal distortion is computed [23]. Column HybridLPPM of Figure 2 shows the ratio of non-protected users on the four datasets for which the best LPPM was chosen (i.e., an LPPM that protects against all the three considered attacks with the lowest spatio-temporal distortion). This result shows that despite the use of an hybrid LPPM for protecting mobility datasets, there is still a large portion of users that are vulnerable to re-identification attacks, i.e. 5% to 36%. Consequently, the generated data loss, as depicted in Figure 3 and that varies between 5% and 42% on the four datasets is still high.

The objective of this paper is thus to design a novel methodology that combines off-the-shelf LPPMs to protect a given mobility dataset in front of a set of user re-identification attacks while minimizing the eventual data loss. In this way, we protect the *crowd* as has been done in the literature and in addition provide other tools to protect *orphan* users.

3 MooD Design Principles

In this section, we present MooD (MObility Data Privacy as Orphan Disease), a system that aims at protecting users that are not protected against re-identification attacks when using a single LPPM. In the following, we start by describing

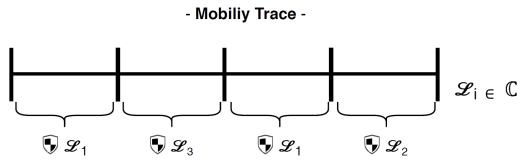


Figure 4. Fine Grained Protection

the system model. Then, we present an overview of MooD followed by a description of its components.

3.1 System Model

Let $\mathbb{U} = \{U_1, U_2, ..., U_N\}$ be the set of users in the system. Each user is represented by two mobility traces, T_{u_i} the one she wants to share and H_{u_i} a past mobility trace used to control the risk of user re-identification.

Definition of a Composition of LPPMs:
A composition of $p \in \mathbb{N}$ protection mechanisms $\{\mathcal{L}_{i_1}, \mathcal{L}_{i_2}, \ldots, \mathcal{L}_{i_p}\}$ a subset of all available LPPMs in $\mathbb{L}$ noted $C_p(\mathcal{L}_{i_k})$ is the application of p LPPMs sequentially and gradually on a mobility trace. As described in Equation 3, it means that we start by applying the first LPPM $\mathcal{L}_{i_1}$. The resulting data is used as an entry for the second LPPM $\mathcal{L}_{i_2}$ and so on. The order of the LPPMs is important since it is similar to a composition of functions[1].

$$\begin{aligned} C_p(\mathcal{L}_{i_k})(T) &= \mathcal{L}_{i_p} \circ \mathcal{L}_{i_{p-1}} \circ \ldots \circ \mathcal{L}_{i_2} \circ \mathcal{L}_{i_1}(T) \\ &= \mathcal{L}_{i_p}(\mathcal{L}_{i_{p-1}}(\ldots \mathcal{L}_{i_1}(T))) \end{aligned} \quad (3)$$

From a set $\mathbb{L}$ of LPPMs, the set of all possible composition is noted $\mathbb{C}$ with $|\mathbb{C}| = \sum_{i=1}^{n} \frac{n!}{(n-i)!}$ where $n = |\mathbb{L}|$

Definition of a Fine-Grained Protection : The fine-grained protection splits the mobility trace into multiple sub-traces and protects each sub-trace independently with different LPPMs (from $\mathbb{L}$ or $\mathbb{C}$) as illustrated in figure 4. The objective of splitting traces is to separate descriminative mobility patterns. To this end, several techniques can be used, e.g., splitting traces according to time, distance or inter-POIs.

Definition of an Orphan User: A user U is an orphan user with respect to a set of LPPMs $\mathbb{L}$, a set of re-identification attack $\mathbb{A}$ and a background knowledge $\mathbb{H}$, if she satisfies the property described:

$$\forall \mathcal{L}_j \in \mathbb{L}, \exists \mathcal{A}_k \in \mathbb{A}, \mathcal{A}_k(\mathcal{L}_j(T_U), \mathbb{H}) = U \quad (4)$$

Definition of a Protected User with Single-LPPM: A user U is said to be protected by a single-LPPM if she satisfies the property: which states that there exists at least one LPPM in the set of considered LPPMs $\mathbb{L}$ that makes all the considered attacks in $\mathbb{A}$ fail at re-identifying the user.

$$\exists \mathcal{L}_j \in \mathbb{L}, \forall \mathcal{A}_k \in \mathbb{A}, \mathcal{A}_k(\mathcal{L}_j(T_U)) \neq U \quad (5)$$

[1] To simplify the notations we omit the parameters of each LPPM and past mobility used of each attack.

Definition of Protected User with Multi-LPPM: A user U is said to be protected by multi-LPPM if she satisfies the property:

$$\exists C_j \in \mathbb{C}, \forall \mathcal{A}_k \in \mathbb{A}, \mathcal{A}_k(C_j(T_U)) \neq U \quad (6)$$

Definition of Data Loss: we define the data loss over a dataset $\mathbb{D} = \{T_1, T_2, \ldots, T_N\}$, with the set of LPPMs Λ against the set of re-identification attacks $\mathbb{A}$ as the ratio of data size (counted by records) of non-protected mobility traces in $\mathbb{D}$. In other words, it is the amount of data remaining after every non-protected mobility trace of the dataset has been erased to avoid user re-identification. As described in Equation 7 (with $|\mathbb{D}|_r$ computes the number of records in $\mathbb{D}$).

$$\begin{aligned} &data\_loss(\mathbb{D}, \Lambda, \mathbb{A}) = \frac{|\mathbb{D}_{NP}|_r}{|\mathbb{D}|_r} \\ &\mathbb{D}_{NP} = \{T \in \mathbb{D} \mid \forall \mathcal{L} \in \Lambda, \exists \mathcal{A} \in \mathbb{A}, \mathcal{A}(\mathcal{L}(T_U)) = U\} \end{aligned} \quad (7)$$

3.2 Overview of MooD

MooD is a fine-grained multi-LPPM user-centric approach. Its main objective is to protect the mobility trace of all users and in particular *orphan* users who are not protected by any single LPPM. The architecture of MooD is depicted in Figure 5 and its behaviour is described in Algorithm 1. MooD takes as inputs: the mobility trace of a user, denoted T, a set of LPPMs denoted $\mathbb{L}$ of cardinality n, a set of user re-identification attacks denoted $\mathbb{A}$ of cardinality m and a utility metric $\mathcal{M}$. It returns obfuscated mobility data as an entire mobility trace T' or as multiple sub-traces $\{T'_1, T'_2, \ldots\}$. It has three main components, the first component *Multi-LPPM Composition Search* aims at finding a multi-LPPM composition for orphan users, i.e. users who are not protected by a single LPPM against re-identification attacks. The second component *Fine-Grained data protection* manages mobility traces for which the first component was not able to find a protecting composition of LPPMs and uses fine-grained protection. In this case, the latter splits the original trace into a set of sub-traces and sends each one back to the first component as depicted in Figure 5. Finally, in the last component *Best LPPM Selection*, only the protected mobility trace (i.e. using single-LPPM or multi-LPPM protection) against all the attacks with the best utility value is retained.

3.3 Multi-LPPM Composition Search

The Multi-LPPM Composition Search is the main component in our system. It takes as input the mobility trace of a user T, the set of LPPMs $\mathbb{L}$ and the set of all considered re-identification attacks $\mathbb{A}$. First (lines 4- 12), we start by applying LPPMs independently to search if there exists an LPPM that can protect the mobility trace.

Then (lines 15- 23), we apply all possible combinations of the considered LPPMs in an incremental and exhaustive

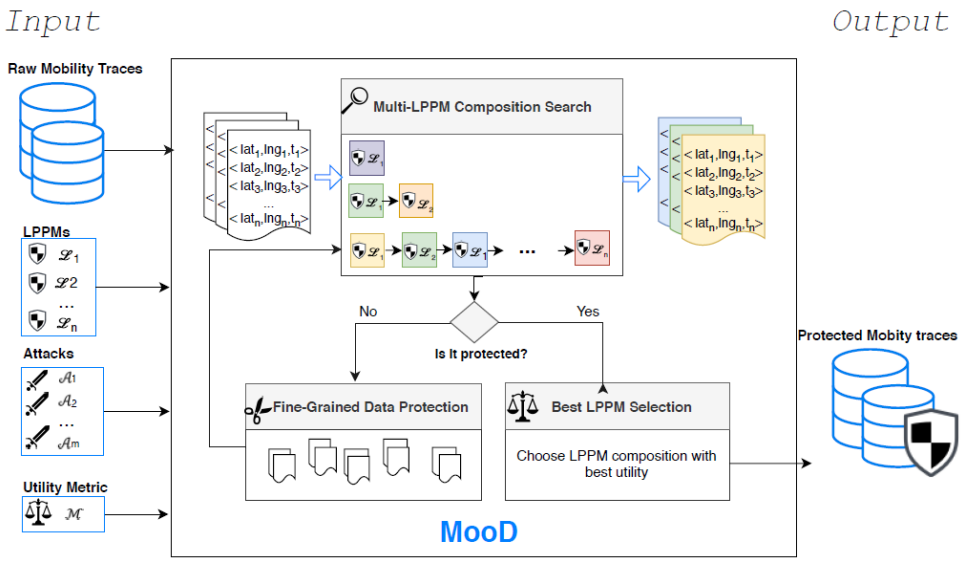


Figure 5. MooD Architecture

manner so that the output mobility trace of the current LPPM becomes the input mobility trace of the next LPPM. For $n = |\mathbb{L}| = 3$, the number of different compositions is $|\mathbb{C} = 15|$ (given by the expression 3). After that, once a mobility trace T is transformed by each composition of LPPMs separately, all the re-identification attacks $\{\mathcal{A}_k\}_{i=1..m}$ are launched in order to evaluate the resilience of each composition of LPPMs and keep only ones that prevent from re-identification (if any).

If all re-identification attacks fail in re-associating the obfuscated mobility trace T' to its originating user, the privacy protection process is done and the user's mobility trace is protected by MooD. In this case, the Best LPPM Selection component chooses the best LPPM composition based on a utility metric (section 3.5). However, if at least one re-identification attack succeeds, it means that the user is still vulnerable. In this case, the mobility trace of the user is undertaken by the next component.

3.4 Fine-Grained Data Protection

The Fine-Grained Data Protection is a complementary component in our system (line 28- 34). It is launched when the user's mobility trace is protected by neither a single LPPM nor a composition of LPPMs. The idea we adopt is to split the original trace into a set of sub-traces and try to protect each sub-trace separately. For that purpose, several techniques can be used or combined for splitting the original trace such as the fixed time slices where we split the trace after a fixed time duration (e.g. every hour) or the fixed distance slices (e.g. every 1 km). In our work, we opt for the fixed time slice. The assumption behind going towards fine-grained protection is that short mobility traces may contain less discriminative information than larger ones. Therefore, re-identification attacks which are based on profiling user mobility will be less successful at re-identifying users because the discriminative mobility patterns collected from the mobility trace of the user are separated. This way, a user can still participate in the published dataset or in the crowd-sensing campaign but only with multiple protected sub-traces that seem to come from different users. In practice, MooD cuts the trace in half according to time and recursively calls for MooD (line 32) with new user IDs (line 34). When the length of a mobility trace is shorter than δ, the protection process for this trace is stopped and the corresponding mobility records are erased from the published dataset or not sent to the crowdsensing server. The main role of the parameter δ is to stop the recursive split of traces. Moreover, in real use cases, the value of δ can be chosen according to the type of analysis the data will go through. For instance, for traffic congestion analysis (or count queries in general) there is no particular limit since the length of each sub-trace is not important to count the presence of users in particular places. But, if the application needs to study human mobility habits, a more reasonable value of δ is likely to be more than 24 hours.

3.5 Best LPPM Selection

It is important to protect data while maintaining high utility of the resulting trace. For that pupose, the Best LPPM Selection component was added to MOOD. Its main role is to choose T', a protected version of a mobility trace with one among all the resilient LPPMs or multi-LPPMs against re-identification attacks, while retaining a high value of utility. To this end, a utility metric $\mathcal{M}$ is considered. Basically, a utility metric measures the distortion of obfuscated data

Algorithm 1 MooD algorithm.

function $\mathcal{M}ood(T_U, \mathbb{A}, \mathbb{L}, \mathbb{C}, \mathcal{M}, \delta)$
  $distortion \leftarrow \infty$
  $out \leftarrow \emptyset$
  for $\mathcal{L}_j$ **in** $\mathbb{L}$ **do** ▷ Single-LPPM protection
    $T' \leftarrow \mathcal{L}_j(T_U)$
    $k \leftarrow 1$
    $limit \leftarrow |\mathbb{A}|$
    while $\mathcal{A}_k(T') \neq U$ **and** $k \leqslant limit$ **do**
      $k \leftarrow k + 1$
    end while
    if $k > limit$ **then** $out \leftarrow out \cup \{T'\}$
  end for
  if $out \neq \emptyset$ **then**
    return $\{\arg\min_{T' \in out}(\mathcal{STD}(T_U, T'))[0]\}$
  else ▷ Composition of multi-LPPMs
    for C_j **in** $\mathbb{C} - \mathbb{L}$ **do**
      $T' \leftarrow C_j(T)$
      $k \leftarrow 1$
      while $\mathcal{A}_k(T') \neq U$ & $k \leqslant limit$ **do**
        $k \leftarrow k + 1$
      end while
      if $k > limit$ **then** $out \leftarrow out \cup \{T'\}$
    end for
  end if
  if $out \neq \emptyset$ **then**
    return $\{\arg\max_{T' \in out}(\mathcal{M}(T_U, T'))[0]\}$
  else if $length(T_U) \geq \delta$ **then**
    $S \leftarrow Split\_in\_half(T_U)$
    ▷ Fine-Grained protection
    $out \leftarrow \emptyset$
    for T_i **in** S **do**
      $out \leftarrow out \cup \mathcal{M}ood(T_i, \mathbb{A}, \mathbb{L}, \mathbb{C}, \delta)$
    end for
    return $renew\_Ids(out)$
  else
    return $\emptyset$
  end if
end function

in comparison with the original data. The lower the distortion the better the quality of the resulting data. In our paper, we measured the utility using the Spatial-Temporal Distortion metric (STD) [23]. As defined in Equation 8, the spatio-temporal distortion $\mathcal{STD}$ is the average distance between each record of T' and its temporal projection into T. The temporal projection of the record $x = (lat_x, lon_x, t_x)$ in T' is the expected position r_e in T at time t. Specifically, we search for $r_i = (lat_i, lon_i, t_i)$ and $r_{i+1} = (lat_{i+1}, lon_{i+1}, t_{i+1})$ in T such as $t_i \leq t_x \leq t_{i+1}$, then we compute r_e the interpolation with the ratio $(t_x - t_i)/(t_{i+1} - t_i)$.

$$\mathcal{STD}(T, T') = \frac{1}{|T'|} \sum_{x \in T'} d_{temporal\_projection}(x, T) \quad (8)$$

4 Experimental Evaluation

In this section, we evaluate the effect of MooD on the protection against re-identification attacks. In Section 4.1, we present the experimental setup and the configuration of the considered LPPMs and attacks. To better understand the impact of the composition of LPPMs and the fine-grained protection in the protection of *orphan* users, we evaluate these parts separately. Specifically, we evaluate MooD's composition effect against single and multiple attacks in Sections 4.3 & 4.4. Then, for the remaining unprotected users, we analyze the effect of MooD's fine-grained protection in Section 4.5.

4.1 Experimental Setup

All the experiments were carried out in a computer running an Ubuntu 16.04 LTS OS with 5GB of RAM and 3 cores of 1.8Ghz each. The different considered LPPMs and attacks were taken from an open-source library [26] or the authors' own source code.

4.1.1 User Re-identification Attack Configuration

The three chosen re-identification attacks in this paper are: AP-attack, POI-attack, and PIT-attack described below.

POI-Attack is introduced in [27] by Primault and others, where Points Of Interest (POIs for short) are used to represent the user's mobility profile. In the training phase, a clustering algorithm [36] is run on raw mobility traces to extract POIs and build each user's mobility profile. After that, in the re-identification phase, when an adversary entity receives an anonymous mobility trace T'_x, the attack builds its mobility profile. Then, it measures the similarity between the anonymous mobility profile and all the known mobility profiles previously built. The similarity is based on the geographical distance between POIs. Finally, only the mobility profile which minimizes the distance is selected.

PIT-Attack is introduced in [16] by Gambs et al., where the Mobility Markov Chain (MMC) model is used to describe users' mobility behaviours. The states are POIs ordered by the number of records inside them and the edges represent the probability of transition between each pair of POIs. In the training phase, the MMC model of each individual is built based on the previously recorded mobility data. After that, in the re-identification phase, when the attacker has at his disposal an anonymous mobility trace, it constructs its MMC model and measures the distance between the latter and the background knowledge. Two information are considered: the geographical distance between POIs based on Euclidian distance and the weight of POIs. The weight of a POI is computed using the proportion of points contained

inside a given POI. The authors of this article proposed many distance metrics to compare MMCs, the most effective one is *stats-prox* distance, a combination of two distances: the stationary distance and the proximity distance [16]. Finally, the MMC model which minimizes the *stats-prox* distance with the anonymous MMC is selected.

AP-Attack is introduced in [22] by Maouche and al, where the concept of heatmaps is used to describe the mobility user profile. To this end, a map is divided into several cells of equal size. The frequency of a visited cell is defined by the number of user's records inside the cell. In this way, it is easy to distinguish between extremely to poorly frequented regions by individuals. To measure the distance between an anonymous profile and the profile of known users, the *Topsoe Divergence* metric is used [13].

The attacks are parameterized as follows: AP-attack has a cell size parameter fixed at 800 meters which is the default value in [22]. *POI*-attack and *PIT*-attack require a diameter of the clustering area to extract points of interest (POIs), and a duration of time spent at a POI. These values were respectively set to 200 meters and 1 hour as done in [22].

4.1.2 LPPM Configuration

To evaluate MooD, we select three representative LPPMs: (1) Geo-I [4], (2) Trilateration [18] (TRL) and (3) HMC [23]. Each LPPM belongs to a class of protection methods. We chose (1) Geo-I, which belongs to data perturbation based mechanisms. It obfuscates mobility traces by adding a laplacian noise to each record of the mobility trace. (2) TRL based mechanism is a different way to generate dummies in online services. It is based on trilateration: when a user launches an LSS (i.e. Location Searching Service) query, looking for a restaurant or a gas station, the algorithm chooses randomly 3 assisted locations $l1$, $l2$ and $l3$ in a range of r from the real location l of the user. The service provider looks for nearby places according to the three assisted locations and sends the result to the user. Finally, the user gets an accurate result by intersecting the result of the three sent locations using trilateration. (3) Heat Map Confusion (HMC) is a combination of data perturbation technique and dummies generation. In HMC, the mobility trace of each user is represented as a heatmap. Then, the algorithm alters the given heatmap by making it look similar to the one of another user. The objective of such an approach is to preserve a certain level of data utility and confuse an attacker that tries to re-identify users'mobility traces. Finally, HMC transforms back each obfuscated heatmap into a set of mobility traces.

Each LPPM has its own configuration parameters. These parameters have an impact on the balance between privacy and data utility. In these experiments, we chose medium values of parameters because the objective of our study is not to find the best configuration as previous works focused on ([9], [25]) but to show that it is possible with a reasonable configuration and a relevant combination of LPPMs to reach an adequate trade-off between privacy and utility. Specifically, Geo-I has ε as a privacy parameter, which tunes the amount of noise added to the mobility data, (the lower the epsilon the higher the protection). We have fixed the value of this parameter to $0,01$ which corresponds to a medium privacy level. TRL has a radius r from the real user's position where the fake locations are generated. The latter was set to 1 km. Finally, as HMC is based on heatmaps, the cell size of the heatmap is a parameter of this technique and it was set to 800 meters which concords with the used value in the original paper [23].

Moreover, we compared our solution to the HybridLPPM algorithm proposed in [22] with slight variations. Briefly, we selected the aforementioned LPPMs with the same parameter values. Then, we ordered them according to the degree of data distortion they generate after obfuscation: (HMC → Geo-I → TRL). Finally, we opt for the LPPM which degrades the least the user's mobility data while protecting it, based on the defined order.

4.2 Mobility Datasets

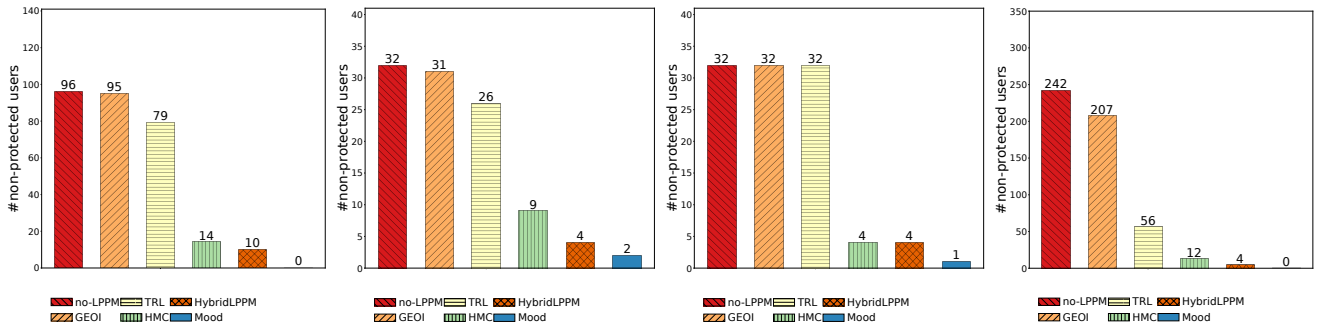
In our experiments, we used four real mobility datasets with a summary depicted on table 1. These datasets are : (1) MDC [19] that contains the mobility of users in the city of Geneva; (2) Privamov [8] that contains the mobility of users in the city of Lyon; (3) Cabspotting [24] that contains the mobility of cab drivers in the city of San Francisco and (4) Geolife [34] that contains the mobility of users in the city of Beijing.

In our experiments, we considered the 30 most active successive days of each dataset. After that, we split the mobility trace of each user chronologically into a period of 15 days used as a training set (i.e. background knowledge) and the remaining 15 days used as a testing set. Only active users during those periods were considered.

In the experiments we conducted, a mobility trace that is still re-identified after the multi-LPPM composition search is split into sub-traces of 24 hours period before applying the recursive splitting algorithm presented previously. We choose chunks of 24 hours to simulate the scenario of a crowdsensing application where users send their data daily. Besides, we set the value of δ to 4 hours in MooD' algorithm.

Table 1. Description of datasets

Name	Cabspotting	Geolife	MDC	PrivaMov
# users	531	41	141	41
location	San Francisco	Beijing	Geneva	Lyon
# records	11 179 014	1 468 989	904 282	948 965



(a) MDC **(b)** Privamov **(c)** Geolife **(d)** Cabspotting

Figure 6. Resilience to one attack – MooD vs. competitors

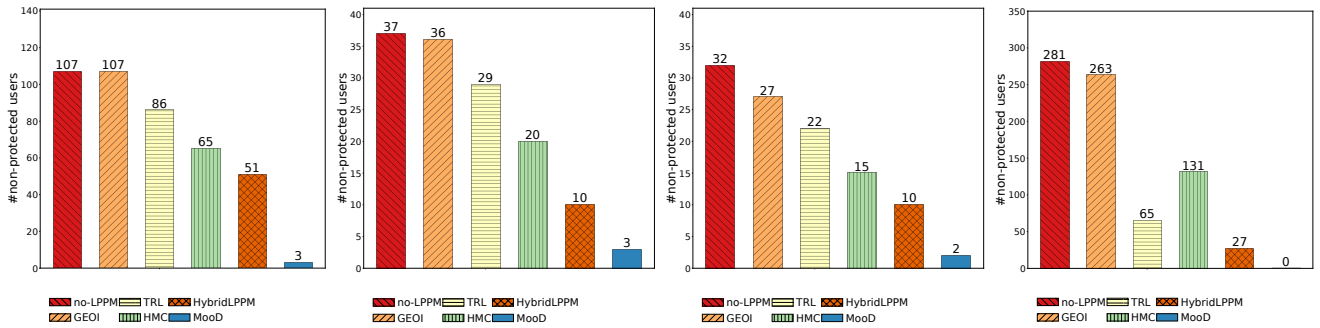
4.3 Resilience of MooD's Multi-LPPM Composition to a Single Re-identification Attack

As a first step, we want to showcase the problem of orphan users when a single attack is used by the data analyst. We consider a set of state-of-the-art LPPMs (n = 3) and we select AP-attack as - the most powerful attack currently known in the literature - in order to evaluate the robustness of the generated obfuscated data against a re-identification attack. We compared the result of MooD with the existing LPPMs applied as single LPPMs on the four datasets.

The results depicted in Figure 6, show that in the MDC dataset, 96 out of 141 users are re-identified when no LPPM is applied, which means that 45 users are naturally insensitive to AP-attack. Additionally, 95, 79 and 14 users are re-identified while applying Geo-I, TRL, and HMC respectively. Whereas these numbers are lower (i.e 10 users) when HybridLPPM is used. All of those users are protected when using MooD. In the PrivaMov dataset, 32 out of 41 users are exposed to re-identification threat. 31, 26 and 9 users are re-identified when Geo-I, TRL and HMC are applied as a single LPPM respectively. Whereas in the case of HybridLPPM, only 4 users remain unprotected, Similarly, in the Geolife dataset, 32 out of 41 users identities are revealed when no LPPM is applied and AP-Attack is launched. Then 4 users are still non-protected by neither a single LPPM nor a HybridLPPM, whereas with MooD'composition of LPPMs, only one user is still vulnerable against AP-attack. Finally, in the Cabspotting, nearly half of the dataset is naturally protected against AP-Attack (242 out of 536), this is due to the homogeneity of cab drivers moving patterns. After applying a single LPPM, almost all the remaining unprotected users became protected except 4 users, for which the application of MooD with its multi-LPPM composition succeeds in protecting them.

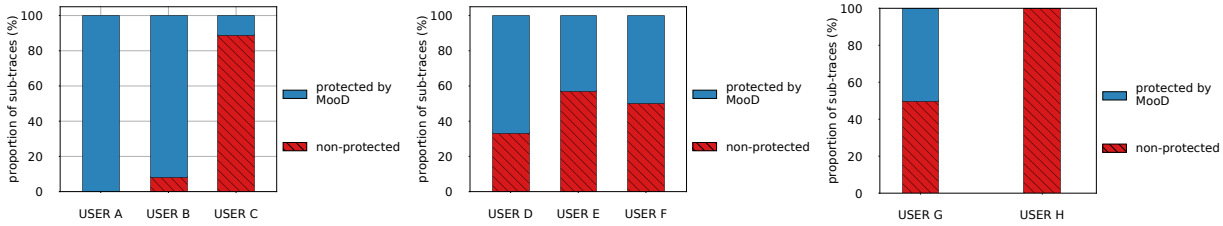
4.4 Resilience of MooD's Multi-LPPM Composition to Multiple Re-identification Attacks

In this experiment, we use MooD with a stronger virtual attacker. It uses multiple re-identification attacks (m = 3) to assess whether the protected users are uncovered by at least one of the attacks. This is possible because MooD knows the ground truth about the real identity of the users. The corresponding results, as shown in Figure 7a, in the MDC dataset, 107 out of 141 users are re-identified when no LPPM is applied. This means that 34 users are naturally protected without the application of LPPMs. Thereafter, 107, 86 and 65 users are non-protected against at least one-attack among the considered ones when Geo-I, TRL and HMC are applied respectively. Then, 51 out of 141 users are still re-identified when HybridLPPM is applied to mobility traces. Whereas only 3 users remain non-protected with the Multi-LPPM combination Search of MooD. In the PrivaMov Dataset, as depicted in Figure 7b, 37 out of 41 users are vulnerable against re-identification attacks when no LPPM is used to protect data. Then, 36 users are re-identified when Geo-I mechanism is applied with medium privacy. The latter is not robust against re-identification attacks. The only way to make it resilient to the considered attacks is to increase its level of privacy (i.e. reduce the value of ϵ) at the expense of data quality. Moreover, 29 and 20 users are non-protected when TRL and HMC are applied, respectively. Whereas these numbers decrease to 10 users when the HybridLPPM is considered. Finally, only 3 users remain non-protected while the multi-LPPM composition in MooD is used. Similarly, in the Geolife dataset, as shown in Figure 7c, 32 users out of 41 users are unprotected against at least one among all the attacks. Then, the number of re-identified users decreases slightly to 28, 23 and 15 users when Geo-I, TRL and HMC are applied separately. Furthermore, the application of HybridLPPM generated 10 unprotected users and finally, only 2 users are still vulnerable against one or more re-identification attacks. Finally, in Cabspotting dataset, as illustrated in Figure7d, more than half of the whole users are re-identified in case of no LPPM. After that, when Geo-I, HMC, and TRL are applied as a Single LPPM, 263, 131 and 65 users are re-identified respectively. Then, the number of re-identified users declines to 27 users with HybridLPPM. Lastly, while considering the multi-LPPM composition of MooD, no users left unprotected. It means



(a) MDC **(b)** Privamov **(c)** Geolife **(d)** Cabspotting

Figure 7. Resilience to multiple attacks – MooD vs. competitors



(a) MDC **(b)** Privamov **(c)** Geolife

Figure 8. Fine-grained protection with MooD

that we can protect the whole dataset from all the considered attacks.

4.5 Evaluation of Fine-Grained Data Protection

As few users remain vulnerable to re-identification attacks, we zoom on this category of users and apply the fine-grained data protection. We start by splitting their mobility traces into multiple sub-traces of 24 h period (as explained in section 3.4). Then, each sub-trace feeds MooD's multi-LPPM composition component in order to protect each sub-trace independently.

In the MDC Dataset, three users are not protected with the composition of LPPMs, denoted $\{A, B, C\}$. We split their mobility traces into 24 h sub-traces. Then, we apply MooD on the resulting sub-traces. Figure 8 shows that there are 68% of protected sub-traces with MooD and 32% of sub-traces remain unprotected with the multi-LPPM composition. Therefore, we can see that user A became protected. User B almost protected (92% of his sub-traces are protected), whereas User C is still non-protected (only 11% of his sub-traces are protected). Thus, the granularity of the considered traces has an impact on privacy protection. With the privamov dataset, three users were left non protected when only the multi-LPPM composition is used, denoted $\{D, E, F\}$. Similarly, we split their mobility traces into sub-traces of 24 h period. Then, we apply MooD's multi-LPPM composition on each sub-trace. The results show that 67% of the sub-traces of user D are protected with MooD, 43% of the sub-traces of user E are protected with MooD and 50% of the sub-traces of user F are protected by MooD too. Therefore, almost the sub-traces are partially protected. Finally, on the geolife dataset, only two users $\{G, H\}$ were not protected by MooD's multi-LPPM composition. Then, after splitting their mobility traces, we obtain 4 sub-traces (i.e. 2 sub-traces for each user), the results show that only one sub-trace is protected by MooD.

4.6 Impact of MooD on Data Utility and Data Loss

It is important to evaluate the effectiveness of MooD in terms of data utility and data loss. In this paper, as discussed in Section 3.5, data utility is measured using the spatio-temporal distortion metric [23] (see Section 3.5) and the data loss is computed as defined in Section 3.1 After that, we compare MooD to state-of-the art LPPMs, previously described (see Section 4.1.2) by considering four different limits of the spatial-temporal distortion: low (i.e. < 500m), medium (i.e. < 1000m), high (i.e. < 5000m) and extremely high (i.e. $\geqslant$ 5000m). Over all the datasets, as depicted in Figure 9, the result shows that for all the protected users (i.e. 754 users in the four datasets), 53.47% have a high utility using MooD (i.e. < 500m) compared its competitors, i.e. 38%, 12%, 45% and 49% with Geo-I, TRL, HMC and HybridLPPM, respectively. Moreover, with medium utility (i.e. < 1000m), MooD outperforms its competitors with a ratio of 78% as to Geo-I, TRL, HMC

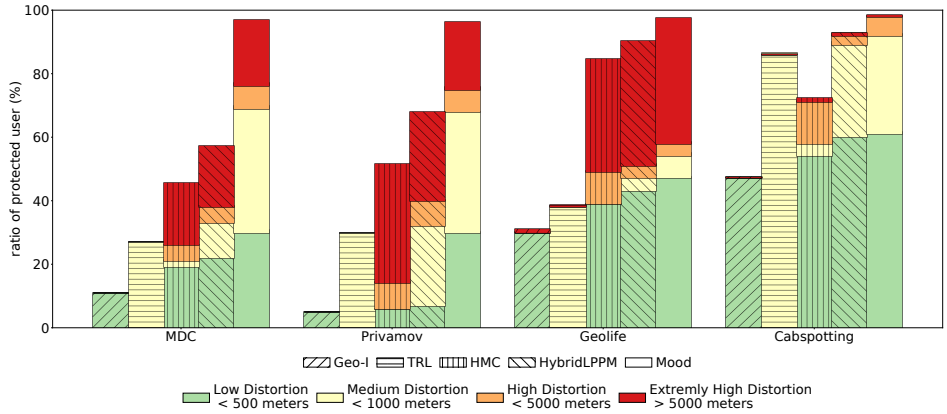
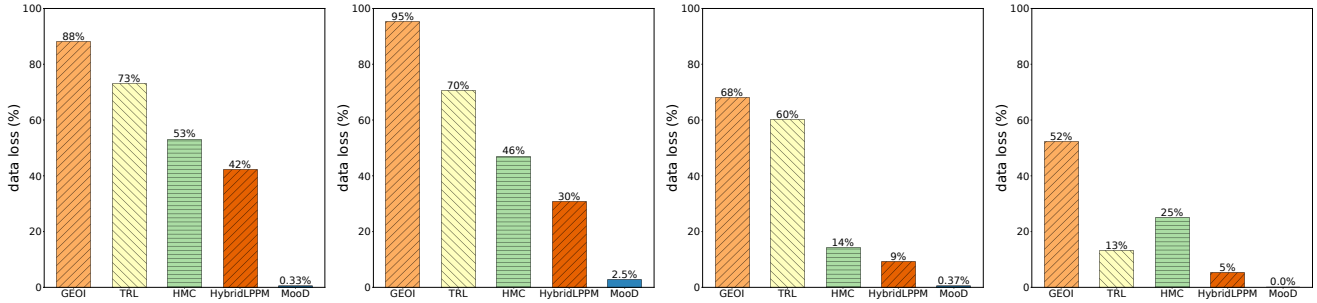


Figure 9. Utility of data protected with MooD vs. Competitors



(a) MDC **(b)** Privamov **(c)** Geolife **(d)** Cabspotting

Figure 10. Ratio of Data Loss MooD vs. Competitors.

and HybridLPPM with 38%, 70%, 48% and 74%. This means that MooD can provide a good balance between privacy and data utility compared to its competitors.

Depending on the degree of the distorted data, we can imagine several scenarios of data publishing and crowdsensing applications using our system. For instance, measuring the level of noise in a city when the distortion is low, like for instance [21] where users participate with their mobile phones equipped with a GPS chip to measure the amount of noise in their vicinity. For medium distortion, our system can be used in an application that measures the level of pollution in a specific area. Finally, for high distortion, an application could be related to weather forecasting where the spatial precision of the protected data is not sensitive as in the previous scenarios.

In addition, we compared the data loss generated by MooD and the state of the art LPPMs. We found that a data loss between 14% and 95% is caused by the application of a single LPPM (i.e. Geo-I, TRL and HMC). Furthermore, when HybridLPPM is used the generated data loss is in a range between 5% and 42%. Whereas with MooD, a data loss between 0% and 2.5% is generated which is a negligible amount of data compared to its competitors.

5 Related Work

Several directions of investigation were taken by researchers to mitigate location privacy threats. Various Location Privacy Protection Mechanisms have been proposed [17], [7], [30], [5]. Some of them rely on formal privacy guarantees such as k-anonymity [31] or differential privacy [12], some of them target sensitive information and try to hide them (e.g., POIs…) [28] and some of them aim at neutralizing specific attacks or preserving specific applications.

NeverWalkAlone [1] and its extension W4M (Wait for Me) [2] are examples of LPPMs which enforce k-anonymity. They guarantee that at each instant, there are at least k users walking inside a cylindrical volume. Another way of achieving k-anonymity is to generate fake data called dummies in order to make the real location of a target user unrecognized. For instance, Trilateration (TRL) [18] randomly chooses 3 assisted locations $l1$, $l2$ and $l3$ in a range of r from the real location l of the user and send them to the service provider instead of the real position l. The latter is mainly used in LSS services (Location searching services) where trilateration is used to get the exact distance between the user's location and the requested places from each assisted location, which

makes this LPPM particularly suited for this type of applications. As for differential privacy, Geo-indistinguishability (Geo-I) introduced by Andres et al. in [4] applies it in the context of location privacy. It bounds the probability of two points to be reported positions of the same real position within a given radius r. Thus, a user can quantify the level of privacy according to their preferences using a privacy parameter ϵ (the lower the ϵ, the higher the privacy level). The authors suggested a way to achieve ϵ-Geo-I in practice by adding Laplacian spatial noise to each GPS coordinate. Additionally, the same authors proposed an extension of ϵ-Geo-I in [10], where contextual information about users' environments are used to calibrate the amount of noise applied on mobility data. Moreover, HeatMap Confusion (HMC) [23], which uses a combination of data perturbation and dummies generation, is an LPPM that is designed to counter particular type of attacks, i.e. re-identification attacks. In the latter, the entire mobility of a user is represented in the form of a heatmap. Each user heatmap is altered so that it looks similar to the one of another user. The heatmap is finally transformed back to a mobility trace using traces of multiple users.

Furthermore, Some authors found that the application of the same LPPM on a whole dataset is not fair regarding the location privacy of users. Some may be overprotected, in the sense that the data utility is decreased further than it is needed and others may remain unprotected [22]. This is why, some works proposed a user-centric approach where each individual is protected according to his mobility trace, his characteristics and his preferences in term of privacy and utility. For example, SmartMask [20] is an example designed to automatically learn users' privacy preferences under different contexts (e.g. location semantic, frequency of visits, duration of visiting a location, time period). Once the privacy level is determined, different techniques ranging from simple operations of obfuscation to cloaking strategies are used depending on the required privacy level. LP-guardian is also a user-centric solution [14] implemented on Android users' smartphones where a decision tree is used to choose the adequate action to perform against different threats. These works are closed to MooD but it was not possible to compare with them because they require richer datasets and not only timestamped mobility traces.

Finally, another possible direction is to play on the configuration of LPPMs in order to protect the users' mobility data. Thus, instead of considering an LPPM with the same configuration, the idea is to tailor the LPPM configuration for each user according to his behavior over the time. To this end, some authors exploited optimization algorithms to find the best configuration for a given LPPM which ensure the trade-off between privacy and data utility objectives. For instance, ALP (Adaptive Location Privacy) [25] a framework which enables an automatic configuration of the LPPM parameters using simulated annealing. Also, PULP [9] is another system which automatically configures LPPMs according to users' objectives in term of privacy and utility. MooD is complementary to those configuration frameworks. In other words, instead of setting LPPMs with a default configuration of parameters, we can integrate one of those framework to provide a tailored configuration for each LPPM or Multi-LPPM according to each user in our system.

6 Conclusion

In this paper, we presented MooD a user-centric fine-grained protection system based on the composition of multiple LPPMs. Its main objective is to protect the minority of orphan users who are not protected by any single LPPM and thus reducing the data loss in a published dataset or in a crowd-sensing campaign. The experiments conducted on four real mobility datasets show that the proposed system is resilient to multiple re-identification attacks and can achieve a high level of privacy protection while maintaining a high utility level.

However, as MooD is based on brute force search for multi-LPPM composition, it is a time-consuming approach. That is why we aim in our future work at optimizing the search by exploring new heuristics and advanced ML techniques.

Furthermore, MooD can be extended by using state-of-the-art LPPMs, attacks and utility metrics. As future work, since the amount of mobility data is continuously growing, the training set of the re-identification attacks can be periodically updated, in order to better feed our system and have a dynamic protection that evolves with the possible evolutions of the user behaviour. Another open direction is to test more sophisticated techniques for the fine-grained data protection component. For instance, a mobility trace can be split by inter-POIs or according to time gaps in mobility traces.

Acknowledgments

This work benefited from the support of the French National Research Agency (ANR), through the SIBIL-Lab project (ANR-17-LCV2-0014), and the PRIMaTE project (ANR-17-CE25-0017).

References

[1] Osman Abul, Francesco Bonchi, and Mirco Nanni. 2008. Never Walk Alone: Uncertainty for Anonymity in Moving Objects Databases. In *Proceedings of the 24th International Conference on Data Engineering, ICDE 2008, April 7-12, 2008, Cancún, Mexico*. 376–385. https://doi.org/10.1109/ICDE.2008.4497446

[2] Osman Abul, Francesco Bonchi, and Mirco Nanni. 2010. Anonymization of moving objects databases by clustering and perturbation. *Inf. Syst.* 35, 8 (2010), 884–910. https://doi.org/10.1016/j.is.2010.05.003

[3] Berker Agir, Kévin Huguenin, Urs Hengartner, and Jean-Pierre Hubaux. 2016. On the Privacy Implications of Location Semantics. *PoPETs* 2016, 4 (2016), 165–183. https://doi.org/10.1515/popets-2016-0034

[4] Miguel E. Andrés, Nicolás Emilio Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. 2013. Geo-indistinguishability: differential privacy for location-based systems. In *2013 ACM SIGSAC Conference on Computer and Communications Security, CCS'13, Berlin, Germany, November 4-8, 2013*. 901–914. https://doi.org/10.1145/2508859.2516735

[5] Omer Barak, Gabriella Cohen, and Eran Toch. 2016. Anonymizing mobility data using semantic cloaking. *Pervasive and Mobile Computing* 28 (2016), 102–112. https://doi.org/10.1016/j.pmcj.2015.10.013

[6] Igor Bilogrevic, Kévin Huguenin, Murtuza Jadliwala, Florent Lopez, Jean-Pierre Hubaux, Philip Ginzboorg, and Valtteri Niemi. 2013. Inferring social ties in academic networks using short-range wireless communications. In *Proceedings of the 12th ACM workshop on Workshop on privacy in the electronic society*. ACM, 179–188.

[7] Igor Bilogrevic, Kévin Huguenin, Stefan Mihaila, Reza Shokri, and Jean-Pierre Hubaux. 2015. Predicting Users' Motivations behind Location Check-Ins and Utility Implications of Privacy Protection Mechanisms. In *22nd Annual Network and Distributed System Security Symposium, NDSS 2015, San Diego, California, USA, February 8-11, 2015*. https://doi.org/10.14722/ndss.2015.23032

[8] Antoine Boutet, Sonia Ben Mokhtar, and Vincent Primault. 2016. *Uniqueness Assessment of Human Mobility on Multi-Sensor Datasets*. Ph.D. Dissertation. LIRIS UMR CNRS 5205.

[9] Sophie Cerf, Vincent Primault, Antoine Boutet, Sonia Ben Mokhtar, Robert Birke, Sara Bouchenak, Lydia Y. Chen, Nicolas Marchand, and Bogdan Robu. 2017. PULP: Achieving Privacy and Utility Trade-Off in User Mobility Data. In *36th IEEE Symposium on Reliable Distributed Systems, SRDS 2017, Hong Kong, Hong Kong, September 26-29, 2017*. 164–173. https://doi.org/10.1109/SRDS.2017.25

[10] Konstantinos Chatzikokolakis, Catuscia Palamidessi, and Marco Stronati. 2015. Constructing elastic distinguishability metrics for location privacy. *PoPETs* 2015, 2 (2015), 156–170. http://www.degruyter.com/view/j/popets.2015.2015.issue-2/popets-2015-0023/popets-2015-0023.xml

[11] Yoni De Mulder, George Danezis, Lejla Batina, and Bart Preneel. 2008. Identification via location-profiling in GSM networks. In *Proceedings of the 7th ACM workshop on Privacy in the electronic society*. ACM, 23–32.

[12] Cynthia Dwork. 2011. Differential privacy. *Encyclopedia of Cryptography and Security* (2011), 338–340.

[13] Dominik Maria Endres and Johannes E Schindelin. 2003. A new metric for probability distributions. *IEEE Transactions on Information theory* (2003).

[14] Kassem Fawaz and Kang G. Shin. 2014. Location Privacy Protection for Smartphone Users. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, November 3-7, 2014*. 239–250. https://doi.org/10.1145/2660267.2660270

[15] Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. 2012. Next place prediction using mobility markov chains. In *Proceedings of the First Workshop on Measurement, Privacy, and Mobility*. ACM, 3.

[16] Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. 2014. De-anonymization attack on geolocated data. *J. Comput. System Sci.* 80, 8 (2014), 1597–1614.

[17] Ruchika Gupta and Udai Pratap Rao. 2017. An Exploration to Location Based Service and Its Privacy Preserving Techniques: A Survey. *Wireless Personal Communications* 96, 2 (2017), 1973–2007. https://doi.org/10.1007/s11277-017-4284-2

[18] Yan Huang, Zhipeng Cai, and Anu G. Bourgeois. 2018. Search locations safely and accurately: A location privacy protection algorithm with accurate service. *J. Network and Computer Applications* 103 (2018), 146–156. https://doi.org/10.1016/j.jnca.2017.12.002

[19] J K Laurila, Daniel Gatica-Perez, I Aad, Blom J., Olivier Bornet, Trinh-Minh-Tri Do, O Dousse, J Eberle, and M Miettinen. 2012. The Mobile Data Challenge: Big Data for Mobile Computing Research. In *Pervasive Computing*.

[20] Huaxin Li, Haojin Zhu, Suguo Du, Xiaohui Liang, and Xuemin Sherman Shen. 2018. Privacy leakage of location sharing in mobile social networks: Attacks and defense. *IEEE Transactions on Dependable and Secure Computing* 15, 4 (2018), 646–660.

[21] Nicolas Maisonneuve, Matthias Stevens, Maria E. Niessen, and Luc Steels. 2009. NoiseTube: Measuring and mapping noise pollution with mobile phones. In *Information Technologies in Environmental Engineering, Proceedings of the 4th International ICSC Symposium, ITEE 2009, Thessaloniki, Greece, May 28-29, 2009*. 215–228. https://doi.org/10.1007/978-3-540-88351-7_16

[22] Mohamed Maouche, Sonia Ben Mokhtar, and Sara Bouchenak. 2017. AP-Attack: A Novel User Re-identification Attack On Mobility Datasets. In *Proceedings of the 14th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, Melbourne, Australia, November 7-10, 2017*. 48–57. https://doi.org/10.1145/3144457.3144494

[23] Mohamed Maouche, Sonia Ben Mokhtar, and Sara Bouchenak. 2018. HMC. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–25.

[24] Michal Piorkowski, Natasa Sarafijanovic-Djukic, and Matthias Grossglauser. 2009. CRAWDAD data set epfl/mobility (v. 2009-02-24).

[25] Vincent Primault, Antoine Boutet, Sonia Ben Mokhtar, and Lionel Brunie. 2016. Adaptive Location Privacy with ALP. In *35th IEEE Symposium on Reliable Distributed Systems, SRDS 2016, Budapest, Hungary, September 26-29, 2016*. 269–278. https://doi.org/10.1109/SRDS.2016.044

[26] Vincent Primault, Mohamed Maouche, Antoine Boutet, Sonia Ben Mokhtar, Sara Bouchenak, and Lionel Brunie. 2018. ACCIO: How to Make Location Privacy Experimentation Open and Easy. In *38th IEEE International Conference on Distributed Computing Systems, ICDCS 2018, Vienna, Austria, July 2-6, 2018*. 896–906. https://doi.org/10.1109/ICDCS.2018.00091

[27] Vincent Primault, Sonia Ben Mokhtar, Cédric Lauradoux, and Lionel Brunie. 2014. Differentially private location privacy in practice. *arXiv preprint arXiv:1410.7744* (2014).

[28] Vincent Primault, Sonia Ben Mokhtar, Cédric Lauradoux, and Lionel Brunie. 2015. Time Distortion Anonymization for the Publication of Mobility Data with High Utility. *CoRR* abs/1507.00443 (2015). arXiv:1507.00443 http://arxiv.org/abs/1507.00443

[29] Adam Sadilek and John Krumm. 2012. Far Out: Predicting Long-Term Human Mobility. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, July 22-26, 2012, Toronto, Ontario, Canada*. http://www.aaai.org/ocs/index.php/AAAI/AAAI12/paper/view/4845

[30] Pravin Shankar, Vinod Ganapathy, and Liviu Iftode. 2009. Privately querying location-based services with SybilQuery. In *UbiComp 2009: Ubiquitous Computing, 11th International Conference, UbiComp 2009, Orlando, Florida, USA, September 30 - October 3, 2009, Proceedings*. 31–40. https://doi.org/10.1145/1620545.1620550

[31] Latanya Sweeney. 2002. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10, 05 (2002), 557–570.

[32] Dashun Wang, Dino Pedreschi, Chaoming Song, Fosca Giannotti, and Albert-Laszlo Barabasi. 2011. Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. Acm, 1100–1108.

[33] Rong Wang, Min Zhang, Dengguo Feng, Yanyan Fu, and Zhenyu Chen. 2015. A de-anonymization attack on geo-located data considering spatio-temporal influences. In *International Conference on Information and Communications Security*. Springer, 478–484.

[34] Yu Zheng, Xing Xie, Wei-Ying Ma, et al. 2010. Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull.* 33, 2 (2010), 32–39.

[35] Changqing Zhou, Dan Frankowski, Pamela Ludford, Shashi Shekhar, and Loren Terveen. 2004. Discovering personal gazetteers: an interactive clustering approach. In *Proceedings of the 12th annual ACM international workshop on Geographic information systems*. ACM, 266–273.

[36] Changqing Zhou, Dan Frankowski, Pamela J. Ludford, Shashi Shekhar, and Loren G. Terveen. 2004. Discovering personal gazetteers: an interactive clustering approach. In *12th ACM International Workshop on Geographic Information Systems, ACM-GIS 2004, November 12-13, 2004, Washington, DC, USA, Proceedings*. 266–273. https://doi.org/10.1145/1032222.1032261