2021 IEEE 41st International Conference on Distributed Computing Systems (ICDCS) | 978-1-6654-4513-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICDCS51616.2021.00082

Practical Location Privacy Attacks and Defense on Point-of-interest Aggregates

Wei Tong*, Chang Xia*, Jingyu Hua*, Qun Li†, and Sheng Zhong*
*State Key Laboratory for Novel Software Technology, Nanjing University, China
†Department of Computer Science, College of William and Mary, USA
weitong@outlook.com, {changxia656569, huajingyu2012}@gmail.com, †liqun@cs.wm.edu, zhongsheng@nju.edu.cn

***Abstract*—Location-based services have significantly affected mobile users' everyday life, and location privacy is also an essential issue in these services. In some applications (e.g., location-based recommendation, mobility analytic), the raw data is not required, and the service providers adopt aggregation to protect users' location traces. However, some works show that even these aggregation data may disclose users' location privacy when other prior knowledge is available to an adversary. We consider the location privacy problem in the presence of *Location Uniqueness*, which is a property that some geographical locations can be re-identified based on the aggregated point-of-interest (POI) information. We first study whether previous protection mechanisms are effective for defending against this novel type of attack. Then we present two practical attacks for inferring users' actual locations based on the POI aggregates. Furthermore, we propose a secure POI aggregate release mechanism that can defend against this type of re-identification attack and achieve differential privacy at the same time. We conduct extensive experiments on real-world datasets. The results show that the existing protection mechanisms cannot provide sufficient protection. The proposed enhanced attacks can significantly improve the inference performance, and the proposed protection mechanism achieves satisfactory performance.**

Keywords*—*Location privacy, POI aggregate, location uniqueness, location re-identification

I. Introduction

Nowadays, our life has been flooded by Location-based Services (LBSs), and location privacy has also been extensively studied in the past dozen years. Some LBSs do not require users' geographic locations but only leverage the knowledge of Points-of-Interest (POIs) near a user, e.g., recommendation and advertising. These systems only require the aggregation information of the POIs near a user, instead of the geographic locations of these POIs or the user's actual location, which seems privacy-friendly for users' locations in the view of previous location privacy studies that aim to protect users' geographic locations directly.

However, a recent study shows that only providing the types of POIs near a user in a city may also reveal the user's actual location[1]. They propose a notion of *location uniqueness*, which implies that many locations in a city are unique regarding the combinations of POIs around them. Based on the property of *location uniqueness*, they find that users' geographic locations can be inferred based on the nearby POIs regarding the distribution of their types and successfully show that many locations in a city have the property of *location uniqueness*. Their work reveals this vital phenomenon and shows that the property of location uniqueness can significantly affect users' location privacy. Nevertheless, there is still a gap that we need to mind to perform a practical attack based on the property of location uniqueness. Furthermore, protecting location privacy when publishing aggregate POI data in the presence of location uniqueness is also an urgent problem that has not been well studied.

In this paper, we study the practical attacks and defense for location privacy in the presence of location uniqueness. Specifically, we first consider the scenario in which the users may initiate multiple successive LBS requests and extend the concept of location uniqueness to *trajectory uniqueness* in this context. Then, we try to design a practical attack that can significantly improve the precision of the inferred locations compared with the existing re-identification attacks. We want to explore whether we can construct fine-grained attacks on users' locations by exploiting the property of location uniqueness. Our goal is to re-identify users' locations into significantly smaller areas, which allows the attacker to locate the target user practically in the real world. Furthermore, based on the studies of the practical attacks, we also investigate how to protect users' location privacy in the presence of location uniqueness without much sacrificing the performance of POI-based services.

We advance the location inference attack from three aspects: 1) we develop a re-identification attack which can infer users' location when they initiate multiple successive LBS requests and find that the success rate of the re-identification can be significantly improved when users continuously use the services by leveraging the knowledge of *trajectory uniqueness*; 2) we propose an iterative positioning scheme for location re-identification, which can significantly shrink the area where the users are in; 3) we also show that the POIs with some certain types have the property of uniqueness as well, and we resort to machine learning methods to learn these POIs even though they have been sanitized in the results for privacy-preserving consideration. This observation also reveals that some straightforward ways, e.g., merely sanitizing the POI frequency list, may not be able to protect the location privacy effectively.

An experimental study has been conducted to investigate whether previous methods like geo-indistinguishability, spatial k-cloaking, and sanitization can successfully protect the location privacy of aggregated POI data in the presence of location uniqueness attacks. The study is performed on the datasets of two representative metropolises: New York City and Beijing. Our results show that these methods can hardly mitigate the re-identification attacks or could be easily broken by more advanced attack techniques.

2575-8411/21/$31.00 ©2021 IEEE
DOI 10.1109/ICDCS51616.2021.00082

To protect the location privacy in the presence of location uniqueness, we employ the notion of differential privacy and have designed an optimization-based POI type distribution publishing mechanism that can protect the location privacy under differentially private guarantee and significantly defend against the location re-identification attacks.

The contributions of this paper can be summarized as:

- First, we revisit the concept of location uniqueness and have conducted experimental studies to evaluate the existing protection mechanisms (sanitization, geo-indistinguishability, and spatial k-cloaking) against the existing location re-identification attack. For the sanitization method, we also show that the learning-based model can easily break the protection.
- Second, we present two practical variants of the location re-identification attack. We advance the existing location re-identification attack from two aspects: extending it to a more general case where users may initiate multiple successive LBS requests and significantly improving the success rate of the attacks by leveraging the information of subsequent queries, being able to locate a specific user in a more precise area.
- Third, we have proposed a differentially private defense mechanism for releasing the POI type frequency vectors, which provides a provable privacy guarantee of the location privacy and satisfactory performance in defending against the re-identification attack.
- Fourth, extensive evaluation has been conducted on real-world data traces, which are extracted for the publicly available geo-information service OpenStreetMap [2], T-drive dataset [3], and Foursquare dataset [4]. The results show that the proposed practical attacks provide better attack performance. The results also show that our proposed differentially private mechanism can effectively defend the re-identification attacks with a reasonable cost of utility.

The rest of this paper is organized as follows. The next section presents the preliminaries, and we evaluate the existing defense mechanisms in Section III. Then present our practical variants of location re-identification attack in Section IV and present our differentially private POI aggregate release mechanism in Section V. The evaluation of the proposed attacks and defense is presented in Section VI. In Section VII, we review the related works, and we conclude this paper in Section VIII.

II. Preliminaries

A. Location Re-identification based on POI type aggregates

We consider a typical LBS architecture in which there are three types of entities: mobile users, the geo-information service provider (GSP), and LBS applications. A mobile user reports its geographic location to the geo-information service provider and gets the geographic information (*e.g.*, POIs, road networks), then it sends the geographic information to the LBS application service providers and enjoys the LBS services. The geo-information service provider stores the geographic data and shares it with the mobile users and LBS applications via a set of query interfaces. The LBS applications provide LBS services and perform various analytic based on the user-location-based geographic data.

Same as many previous works[5], we assume that the LBS applications cannot access mobile users' geographic locations directly. Instead, when a mobile user wants to use the LBS applications, it sends its location to the geo-information service provider and gets the geographic data, and then reports the geographic data aggregates to the LBS applications (*e.g.*, POI-based services). Furthermore, we assume that the geo-information service provider only provides one query interface: retrieving the POIs within a specific range of a location. The LBS architecture adopted in this paper is illustrated in Fig. 1. Note that the POI aggregates may be generated by the users or the GSP and sent to the LBS applications.

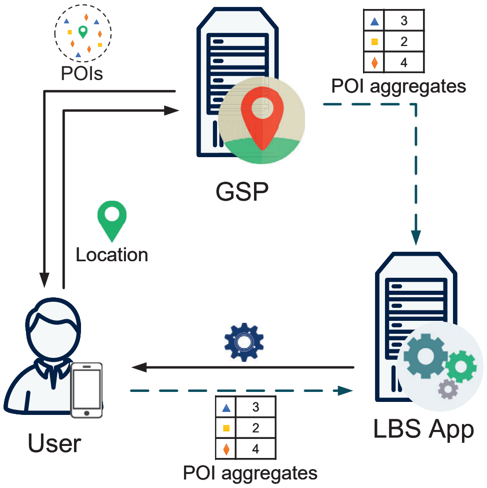


Figure 1: POI-based LBS architecture

Formally, the user reports its location l and a given query range r to a GSP; then the GSP generates the set of POIs in the queried range, denoted by $P_{l,r}$, and return it to the user. This process can be achieved by the operation:

$$P_{l,r} \leftarrow \mathsf{Query}(l, r).$$

In the context of POI type aggregates, the user or GSP do not directly reveal the actual location l or set of POIs $P_{l,r}$ to the LBS application service provider. Instead, the POI type distribution $F_{l,r} = (n_1, n_2, \ldots, n_M)$ is aggregated by users and released to the POI-based services (*e.g.*, recommendation), where n_i is the frequency of POI type t_i in the result, M is the number of different types of POIs in the city. The POI type distribution can be generated by operation:

$$F_{l,r} \leftarrow \mathsf{Freq}(l, r).$$

As it has been stated in [1], the location re-identification problem is to re-identify the location l based on the distribution $F_{l,r}$. What makes this possible is the property of location uniqueness in the city [1]: given the query range, a location could be re-identified because it has a unique combination of POIs compared to other locations in a city. Formally, the re-identification can be formulated as:

$$\Phi(l) \leftarrow \mathsf{Infer}(F_{l,r}, \mathcal{P}), \tag{1}$$

where $\mathcal{P}$ is the prior knowledge of an adversary, $\Phi(l) = \{\phi_1(l), \phi_2(l), \ldots, \phi_{|\Phi|}(l)\}$ is a set of re-identified areas where location l could be in.

B. Threat Model

We assume that the adversary is semi-honest, which means it is interested in inferring users' locations based on the informed information but does not deviate from the protocol specification.

Abilities. The adversary could be a third-party entity that uses the POI type aggregates involved in the POI-based services. We assume that the adversary can access a set of prior knowledge $\mathcal{P}$, which includes public geo-information of a city and the operation Freq to get POI type frequency of any location with required query range. Such prior knowledge can be obtained from some publicly available geo-information service providers, *e.g.*, OpenStreetMap [2]. Besides, same to [1], we also assume that the adversary can obtain: 1) the user's identification corresponding to the reported $F_{l,r}$, which make it possible for the adversary to link the re-identified location to a particular user; 2) user's query range r. These two types of information are essential information for any location-based services and are usually included in the meta-data with queries.

Goals. The adversary tries to re-identify a user's location based on $F_{l,r}$ and the prior knowledge by implementing an inference mentioned in (1). Ideally, $|\Phi|$ should be 1, and the size of the only element $\phi^*(l)$ in the set should be as small as possible. Particularly, same to [1], we define the case where $|\Phi| = 1$ as a successful attack, and $|\Phi| \neq 1$ means that the attack fails. Therefore, we adopt two metrics to evaluate the inference method Infer: 1) success rate of attacks, which equals to the ratio between the number of successful attacks to the number of all attacks; 2) when an attack is successful, the area of $\phi^*(l)$ is used to measure the precision of the inference.

C. Privacy Model

Differential privacy (DP) has become an very important standard for data privacy protection in recent years. For the sake of completeness, below we first review the definition of DP [6].

Definition 1: A randomized mechanism $\mathcal{M} : \mathcal{X}^d \rightarrow \mathcal{Y}$ is (ϵ, δ)-differentially private if and only if any two neighboring datasets $\mathbb{D}_1$, $\mathbb{D}_2 \in \mathcal{X}^d$, and all $\mathcal{S} \subset \mathcal{Y}$,

$$\Pr[\mathcal{M}(\mathbb{D}_1) \in \mathcal{S}] \leq \exp(\epsilon) \Pr[\mathcal{M}(\mathbb{D}_2) \in \mathcal{S}] + \delta, \tag{2}$$

where ϵ and δ are privacy parameters.

Then, we review the Gaussian mechanism, which we will adopt as a component in our private defense mechanism.

Definition 2: Guassian mechanism adds a noise $\mathcal{N}(0, \sigma^2)$ to $f(\mathbb{D})$, where f is a function with sensitivity Δ. If

$$\sigma \geq \sqrt{2\ln(1.25/\delta)}\Delta/\epsilon, \tag{3}$$

then, the mechanism achieves (ϵ, δ)-differential privacy.

Lemma 3 (Post-processing[6]): Let $\mathcal{M} : \mathcal{X}^d \rightarrow \mathcal{Y}$ be a randomized mechanism satisfying (ϵ, δ)-differential privacy. Let $\mathcal{A} : \mathcal{Y} \rightarrow \mathcal{Y}'$ be an arbitrary deterministic or randomized mechanism. If $\mathcal{M}' : \mathcal{X}^d \rightarrow \mathcal{Y}'$ is sequential apply of $\mathcal{M}$ and $\mathcal{A}$, then $\mathcal{M}'$ is (ϵ, δ)-differentially private.

D. Region Re-identification

For the sake of completeness, we review the location re-identification method in [1] in this part. Specifically, the attack runs by the following steps:

❶ Counting the overall POI frequency in the entire city, denoted by F;

❷ Sorting $F_{l,r}$ by F, and denoted by t_l the most infrequent POI type in F which satisfies $n_l > 0$;

❸ Finding all POIs with type t_l in the city, and denoted by P_{t_l} the resulted set of POIs;

❹ Pruning the set of locations P_{t_l} with following rule:
   - ○ For each $p_{t_l} \in P_{t_l}$, get
   $$F_{p_{t_l},2r} \leftarrow \mathsf{Freq}(p_{t_l}, 2r);$$
   - ○ For $i = 1, 2, \ldots, M$, if exist any i such that $F_{p_{t_l},2r}[i] < F_{l,r}[i]$, remove p_{t_l} from the candidate set P_{t_l}.

❺ After the above pruning process, if there is only remaining one location $p^*_{t_l}$ in the set P_{t_l}, the location l has the property of uniqueness. The adversary can infer that location l is in the range of $p^*_{t_l}$ with radius r.

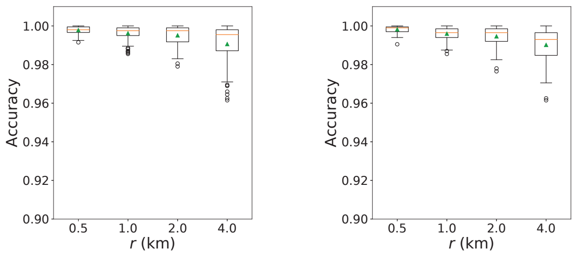
Their method is based on the property that the circle that is centered at l with radius r is completed covered by the circle centered at p_{t_l} with radius $2r$ if p_{t_l} is a POI in the distance r of location l. By using this method, the adversary can re-identify a location by POI type distribution with no false negative, but the success rate is affected due to the gap between $F_{p_{t_l},2r}$ and $F_{l,r}$. Also, the adversary can only determine that location l is in the range with distance r of $p^*_{t_l}$, which means the size of $\phi^*(l)$ is πr^2, which seems to be an infeasible range for attacks on location privacy. For convenience, we refer to this attack as *region re-identification* or *Cao et al.'s attack* in the following parts of this paper.

E. POI Datasets

The POI datasets are extracted from a publicly available geo-information service, OpenStreetMap [2]. We choose New York City and Beijing as the targets of our analysis. Beijing dataset contains 10,249 POIs with 177 different types; New York City (NYC) dataset contains 30,056 POIs with 272 different types.

III. Evaluating the Existing Protection against Region Re-identificaiton

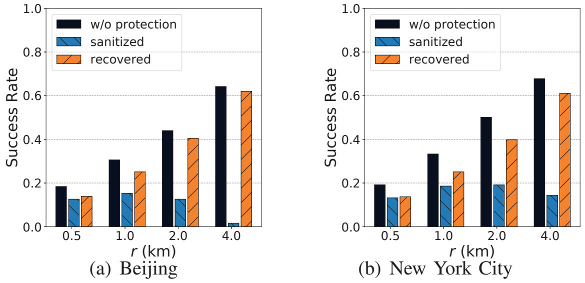
We now measure the region re-identification attack against three protection mechanisms: sanitization and geo-indistinguishability, and spatial k-cloaking.



(a) Beijing

(b) New York City

Figure 2: The accuracy of prediction models.



(a) Beijing

(b) New York City

Figure 3: Performance of the sanitization.

A. *Sanitization*

A straightforward solution that seems can be applied to protect location privacy in the presence of location uniqueness is to sanitize the frequencies of the POI types, especially for those infrequent POI types. Below we describe a sanitization strategy for the protection, which removes the information of frequencies of POI types that are infrequent in the city. Our results show that the aggressive sanitization strategy can significantly reduce the success rate of the region re-identification attack in some cases, but we propose a learning-based inference method to show that the defense can be easily compromised.

Defense strategy. Based on the overall POI frequency in the entire city, F, the sanitizer chooses a set of POI types T^S which satisfies that any POI type $t_i \in T^S$, $F[i] <= S$. When trying to report the POI type distribution $F_{l,r}$ of the location l in range r, the user sets $F_{l,r}[i] = 0$ if $t_i \in T^S$. When implementing the strategy, we adopt a very aggressive sanitization behavior, *i.e.*, sanitizing 138 (90, resp.) POI types whose frequencies are no more than 10 in New York City (Beijing, resp.).

Prediction against sanitization. We assume that the adversary knows whether a specific POI type is sanitized or not sanitized. For example, the attacker may collect the historically reported frequencies for inferring such information. For each sanitized POI type t^S, we train a prediction model based on the reported frequencies of POI types. Formally, the prediction model is formulated as

$$\mathsf{Pred}(\mathbf{x}^{-S}) \rightarrow n^S.$$

where $\mathbf{x}^{-S} = (n_1 n_2 \ldots n_{|T^{-S}|})$ is the feature vector of prediction sample, in which n_i is the corresponding frequency of POI type t_i. We should clarify that t_i is a type in the set T^{-S}, which is the set of POI types that are not sanitized; n^S is the target of the prediction model, which is the frequency of sanitized POI type t^S.

We adopt the support vector machine (SVM) classification [7] with radial basis function (RBF) kernel as an installation of the prediction model. Our experiments are implemented by using Scikit-learn machine learning package [8]. In the training process, random locations are generated in the corresponding city, and by adopting Freq operation, the POI type distributions are generated from these locations. We compose a training dataset with 10,000 samples and a validation dataset with 2,000 samples for training based on the generated POI type distributions. All samples in the prediction model are normalized by being centered to mean and scaled with unit standard deviation.

Results. Fig. 2 shows the classifiers' performance for different query range (r). In this set of experiments, we evaluate the defense strategy with user locations that are randomly generated in corresponding cities. We can observe that for both Beijing and New York City, in the cases of typical query ranges with 0.5km, 1.0km, 2.0km, and 4.0km, the average validation accuracy of trained classifiers for all targets is larger than 95%. Specifically, for the Beijing, the means of accuracies are 0.998 (±0.002), 0.996 (±0.004), 0.995 (0.005), and 0.991 (±0.010) for the above four query ranges, respectively. For New York City, the means of accuracies are 0.998 (±0.002), 0.996 (±0.003), 0.995 (0.005), and 0.990 (±0.008) for the above four query ranges, respectively.

Fig. 3 shows that the sanitization can mitigate a major part of the attacks, and reduces the success rate from 0.184, 0.306, 0.440, and 0.642 to 0.126, 0.153, 0.126, and 0.016, respectively. For New York City, the success rates decrease from 0.192, 0.333, 0.501, and 0.678 to less than 0.2 for four cases, respectively, when the defense is applied. However, we observe that the prediction models can recover the sanitized types, and achieve an almost success rate compared with the original attacks without protection.

B. *Geo-indistinguishability*

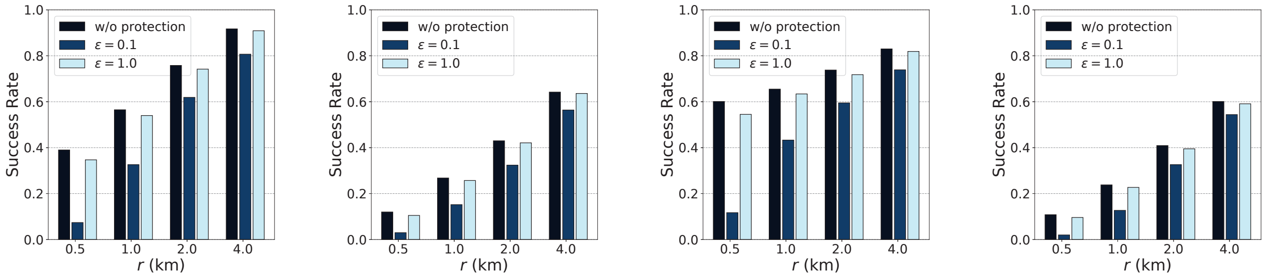
Geo-indistinguishability [9] is a variant of differential privacy, which provides provable guarantees of location privacy. Its main idea is to bound the difference between distributions of observations that are produced by two close locations by probabilistic perturbation. Formally, a mechanism M is geo-indistinguishable if and only if for any l, l' which satisfy $\mathsf{dist}(l, l') \leq R$:

$$|\ln \frac{M(l)}{M(l')}| \leq \epsilon R. \tag{4}$$

Planar Laplacian [9] is a canonical way to achieve geo-indistinguishability, which runs in the following way: given user's location l and the privacy parameter ϵ, for any other location l' in the considered area, the mechanism chooses l' as the reported location by the following probability:

$$M_\epsilon(l)(l') = \frac{\epsilon^2}{2\pi} \exp^{\epsilon \times \mathsf{dist}(l,l')}. \tag{5}$$

Results. In this set of experiments, we evaluate the defense strategy with four datasets:(a) T-drive [3] user locations in



(a) Bejing: T-drive (b) Bejing: Random (c) NYC: Foursquare (d) NYC: Random

Figure 4: Performance of Planar Laplacian

Beijing; (b) randomly generated user locations in Beijing; (c) Foursquare check-ins [4] in NYC; (d) randomly generated user locations in NYC. In our experiments, we also note that the unit of distance is set to 100 meters, which will affect the privacy level and the utility of the perturbation given the specific privacy parameter. For each dataset, 1000 locations are randomly selected for the experiments.

Fig. 4 shows the performance of *Planar Laplacian* for defending against location re-identification. We can find that when the privacy budget is larger (*e.g.*, $\epsilon = 1.0$), the mechanism can barely mitigate the inference attack. When we set $\epsilon = 0.1$ and $r = 0.5$, 1.0, 2.0, and 4.0, respectively, the *Planar Laplacian* can mitigates about 81.01%, 42.30%, 18.34%, and 12.00% of attacks for T-drive dataset in Beijing, about 75.00%, 43.28%, 24.65%, and 12.15% for random locations in Beijing, about 80.53%, 33.89%, 19.38%, and 10.96% for Foursquare dataset in NYC, and about 81.48%, 46.64%, 20.29%, and 9.48% for random locations in NYC. The defense can mitigate most of the attacks when the query range is small, but the performance is limited when the query range is large.

C. Spatial k-cloaking

Spatial k-cloaking is a type of location privacy protection mechanism, which aims to hide a location into a larger area containing the requester and at least k-1 other users. In our evaluation, we have adopted the adaptive-interval cloaking algorithm [10] as the protection scheme. For the sake of completeness, we briefly review the adaptive-interval cloaking algorithm below:

❶ The algorithm first sets the whole city area as the initial current area for cloaking.

❷ It partitions the current area into four non-overlapping sub-regions with equal size, and test whether the sub-region which contains the targeted location satisfies the k-anonymous property, *i.e.*, there are at least k users in this sub-region.

❸ If the sub-region satisfies the k-anonymous property, it repeats ❷ and ❸; otherwise, it chooses the generated region in the last iteration as the cloaking area.

Results. In this set of experiments, we evaluate the defense strategy with four datasets that are the same as we have adopted in the Section III-B. We assume that there are 10,000 users who are uniformly distributed all over the city for each city. Fig. 5 shows the performance of spatial k-cloaking for defending against location re-identification. We can find that the success rate decreases with k increasing, but its performance is still not satisfactory when k is sufficiently large (*e.g.*, $k = 50$).

D. Takeaways

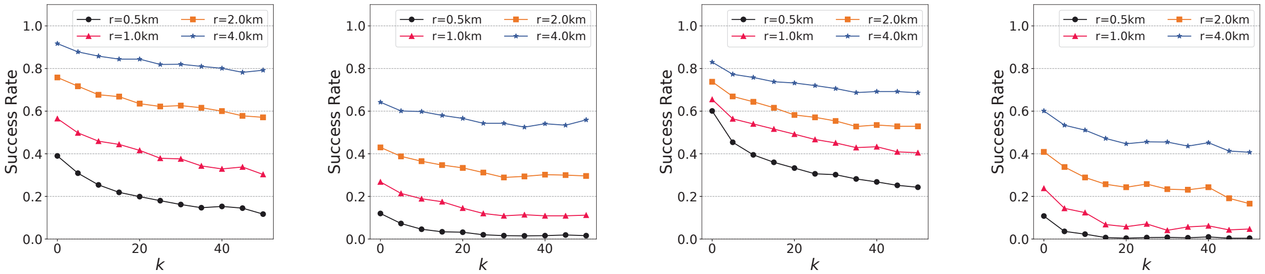
We have the following three major observations from the above experiments study:

- Location-level protection (*e.g.*, Geo-indistinguishability, Spatial k-cloaking) achieves better performance when the query range is small. From Fig. 4 and Fig. 5, we can find that when the query range is small, both the Geo-indistinguishability and the Spatial k-cloaking can reduce the success rate of the attacks more significantly compared with the cases where the query ranges are larger.
- Frequency-level protection (*e.g.*, sanitization) can provide better protection when the query range is large. From Fig. 3, we can find that the sanitization can significantly reduce the success rate when the query range is large. Unfortunately, it cannot provide sufficient protection when the query range is low or more powerful attacks exist.
- The attack could be more powerful when the user traces in real-world applications. From these three sets of experimental studies, we can observe that the re-identification attack can achieve higher success rates for the real-world data traces.

IV. Understanding the Location Uniqueness via Practical attacks

An important direction for location privacy research has been pointed out by Cao et al.'s work [1] by introducing the concept of *location uniqueness*. They also provide a feasible method for re-identifying regions that may contain the target user based on POI type distribution. However, as we have mentioned above, their approach is mainly used for exploring the existence of location uniqueness, and an adversary who is interested in users' location privacy may need more powerful tools for launching the attacks.

For the practical attacks, we identify two major goals: 1) the re-identified location should be more precise, which



(a) Bejing: T-drive (b) Bejing: Random (c) NYC: Foursquare (d) NYC: Random

Figure 5: Performance of spatial k-cloaking.

means the adversary can determine the user's location in a sufficiently small area; 2) the success rate should be further improved, which means the adversary has a high probability to determine the user's location in only one area successfully. In this section, we introduce two practical variants of the region re-identification based on POI type distribution to pursue the above two goals, respectively.

A. *Fine-grained Attack*

After applying the Cao et al.'s attack, an adversary can re-identify those locations with the property of location uniqueness and narrow each successfully re-identified location in a circle with radius r, but cannot determine where the locations exactly are. We find that the basic re-identification method uses the relationship between $F_{l,r}$ (*i.e.*, POI type distribution around actual location l with radius r) and $F_{p_{t_l},2r}$ (*i.e.*, POI type distribution around found POI p_{t_l} with radius $2r$) to prune the candidate set of re-identified POIs. Their method only uses the POI type distribution information of POIs with the most infrequent type. Nevertheless, we find that other POIs with other types can also be exploited to locate the user.

The basic idea of the proposed fine-grained inference method is to shrinkage the area the user is in by iteratively applying the candidate pruning strategy for other types of POIs, and find POIs in $P_{l_q^*,2r}$ that are also in $P_{l,r}$. After finding these POIs, the adversary can further locate l because l is definitely within r of the selected POIs. Specifically, we present the following scheme to find a significantly smaller area l should be in, which consists of three steps:

- The first step is to re-identify the location l by Cao et al.'s region re-identification method, which can infer that location l is in the range of $p_{t_l}^*$ with radius r. We refer to the found POI $p_{t_l}^*$ as the major anchor for the location inference.
- Once the anchor POI $p_{t_l}^*$ is found, we can further improve the accuracy of the re-identification of location l by leveraging other types of POIs in the surrounding area. Though the queried POIs $P_{l,r}$ based on location l are unknown to the attacker, it can obtain the set of POIs $P_{p_{t_l}^*,2r}$, which is a superset of $P_{l,r}$. Based on this knowledge, the attacker can carefully filter the points in $P_{p_{t_l}^*,2r}$, and find some auxiliary anchors to position the location l. An algorithm to find these auxiliary anchors is presented in Algorithm 1.
- After generating the set of auxiliary anchors, which are all in the range of r of the location l, and thus the location l can be positioned into a very fine-grained area by computing the feasible area that satisfied the requirements of these anchors.

Algorithm 1: Iteratively Shrink the Region.

```
Input: $F_{l,r}$: the frequency vector of POI types;
       $t_l$: most infrequent POI type;
       $p_{t_l}^*$: correponding POI for re-identifying $l$;
       $\mathrm{MAX_{aux}}$: maximum size of set of anchors.
Output: $Aux$: set of POI for positioning $l$
$Aux \leftarrow \emptyset$;
$P_{l_q^*,2r} \leftarrow \mathsf{Query}(l_q^*, 2r)$;
$F_{l_q^*,2r} \leftarrow \mathsf{Freq}(l_q^*, 2r)$;
$F_{\mathrm{diff}} \leftarrow F_{l_q^*,2r} - F_{l,r}$;
Sort $F_{\mathrm{diff}}$ based on the corresponding frequencies of
  POI types.
foreach $t_i \in F_{\mathrm{diff}}$ do
    if $F_{\mathrm{diff}}[t_i] = 0$ then
        $Aux \leftarrow Aux \cup \{p \in P_{l_q^*,2r} | p.\mathrm{type} = t_i\}$
    else
        foreach $p \in P_{l_q^*,2r}$ and $p.\mathrm{type} = t_i$ do
            $flag \leftarrow True$;
            $F_{p,2r} \leftarrow \mathrm{Freq(p, 2r)}$;
            foreach $t_p, v_p \in P(p, 2r)$ do
                if $v_p < p(l,r)[t_p]$ then
                    $flag \leftarrow False$;
            if $flag$ then
                $Aux \leftarrow Aux \cup \{p\}$
    if $|Aux| >= \mathrm{MAX_{aux}}$ then
        break;
```

In Algorithm 1, we first compute the difference between POI type distributions of the actual location and major anchor and get a differential vector F_{diff}. The algorithm traverses all types of POI based on the sorted type in F_{diff}. This operation is adopted to speed up the iterative shrinkage process because the algorithm can first consider the types that need fewer efforts to prune the POIs. For example, if for a type t_i such that $F_{\mathrm{diff}}[i] = 0$, then all POIs with type t_i in $P_{l_q^*,2r}$ are in $P_{l,r}$. Therefore, we can directly use these POIs with type t_i to shrink the target area without additional efforts.

B. Attack with Trajectory Uniqueness

When users are using location-based services, they often query the service multiple times. Several successive queries may further reveal users' location based on the inference on the aggregated POI frequencies. We call this property as trajectory uniqueness and demonstrate that it can be leveraged for location re-identification with a better success rate.

For the cases that the adversary can leverage multiple releases of POI type frequencies, the location re-identification problem is extended to the following form:

$$\{\Phi(l_1), \Phi(l_2), \ldots\} \leftarrow \mathsf{Infer}(\{F_{l_1,r}, F_{l_2,r}, \ldots\}, \mathcal{P}). \quad (6)$$

By repeatedly applying the single location version of re-identification attack, the adversary can get a series of inference candidates: $\{\hat{\Phi}(l_1), \hat{\Phi}(l_2), \ldots\}$. Our goal is to figure out which subset contains the areas that the user is possible in for a given candidate set $\hat{\Phi}(l_t)$. An ideal case is that the adversary has the knowledge about the distance between two locations, *i.e.*, $\mathsf{dist}(\phi^*(l_t), \phi^*(l_{t+1}))$. Thus, the adversary can filter pair of candidate areas in the candidate sets and find the possible pair of locations based on their distance. However, this assumption seems unrealistic for most cases, and we need a practical way to estimate the distance between two successive locations.

We consider the distance estimation as a regression problem. We find that the crucial part of this extended re-identification problem is that the prior knowledge $\mathcal{P}$ is also extended. The adversary also captures the duration between two successive releases. Therefore, we try to build a regression model mainly based on the duration and other auxiliary information to predict the distance between two locations of corresponding releases. Specifically, we construct the feature vector with the following information:

- The duration between two successive releases: $\mathsf{time}(l_t, l_{t+1})$;
- The $L1$-distance between $F_{l_t,r}$ and $F_{l_{t+1},r}$;
- In which hour of a day the first POI type frequency is released, and which day of a week for this release. These two types of information are encoded by one hot encoding in the feature vector.

Based on the constructed feature vectors, we adopt support vector regression [11] that is provided Scikit-learn machine learning package [8] to train the regressor.

V. Optimization-based Defense with Differential Privacy

In this section, we will describe a differentially private mechanism to protect users' locations against the re-identification attacks in the sharing of POI frequencies.

By revisiting the Cao et al.'s attack [1], we can find that the region re-identification attack succeeds so long as the adversaries can locate the targeted users in areas with radius r. Instead of finding the target user's exact location, the attack tries to figure out to which POI the target is close. Under this setting, it is hard to achieve good defense performance with the location-level methods, *i.e.*, only perturbing an actual location to a noise location seems not a right choice, and the evaluation of geo-indistinguishability also supports this argument. On the other hand, the aggregate-level methods can provide effective protection to some extent, but it is vulnerable to advanced attacks when the adversary obtains other background information, as it has been shown in the evaluation of the sanitization. Besides, the aggregate-level protection may yield the POI type frequencies with poor utility because it will remove some essential information from the reported aggregate if we take an aggressive defense strategy.

We resort to aggregate-level protection, which could perturb the POI type frequencies of users and make two improvements over the naïve sanitization method in Section III-A:

- The naïve sanitization method does not take the utility into account, but it is crucial for services that need the POI type frequencies. We show that the proposed defense can provide the comparable utility of the perturbed frequencies.
- It has been shown that the naïve sanitization is vulnerable to advanced attacks with auxiliary information. The proposed defense provides a plausible guarantee of the perturbed frequencies against the auxiliary information-based attacks by introducing the notion of differential privacy.

A. Non-private Formulation

We first formulate the perturbation objective in a non-private way. Given the original POI type frequency vector $F_{l,r}$, we adopt the following optimization to find a proper release $\tilde{F}$:

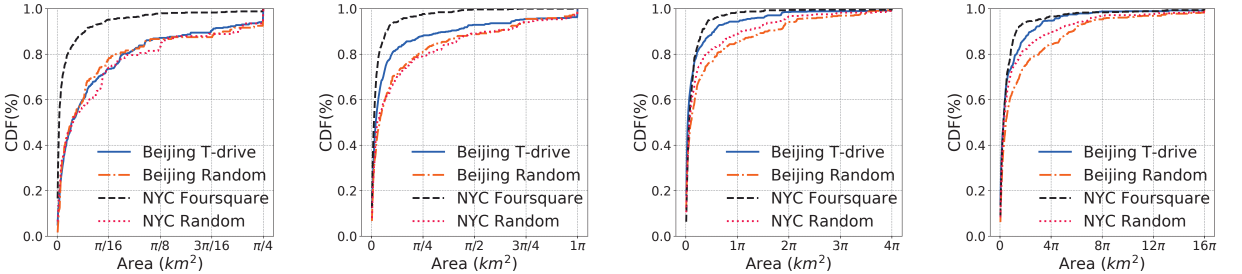
$$\begin{aligned} &\max_{\tilde{F}} \sum_{i=1}^{M} \frac{1}{R(i)} |\tilde{F}_i - F_{l,r}[i]|, \\ &s.t. \;\; \frac{1}{M} \sum_{i=1}^{M} \frac{1}{F_{l,r}[i]+1} |\tilde{F}_i - F_{l,r}[i]| \leq \beta, \\ &\qquad \tilde{F}_i \in \mathbb{N}^+, i = 1, 2, \ldots, M. \end{aligned} \quad (7)$$

We want to maximize the weighted perturbation to the released frequencies while constraining the total distortion to the frequencies under a certain level, β. In the above formulation, $R(i)$ is the infrequent rank of each POI type (the most infrequent POI type ranks 1, and so forth).

B. Differentially Private Release

The indistinguishability provided by the definition of DP guarantees that the released POI type frequency vector is insensitive to each POI type's frequency in the original frequency vector. To further illustrate the guarantee provided by DP, we specify the neighboring datasets in the release of POI type frequencies. We refer to a pair of datasets $\mathbb{D}_1, \mathbb{D}_2 \in \mathcal{X}^d$ as neighbors if they are two POI frequency vectors and $\mathbb{D}_2$ can be obtained from $\mathbb{D}_1$ by only modifying one dimension of POI type frequency.

The main idea is to generate a privacy-preserving alternative to $F_{l,r}$ in Eq. (7), such that we can find a proper release $\tilde{F}^*$ which can defend against the re-identification attack and achieve differential privacy at the same time.



(a) $r = 0.5$km (b) $r = 1.0$km (c) $r = 2.0$km (d) $r = 4.0$km

Figure 6: Performance of the fine-grained attack: the CDF of search area. The search area of Cao et al's attack is always $\pi r^2 \mathrm{km}^2$.

Specifically, the defense mechanism consists of the following steps:

- It first adopts the spatial k-cloaking mechanism [10] to generate the a group of dummy locations as we have reviewed in Section III-C, and the generated k locations (including l) are denoted by $d_1, d_2, \ldots, d_k$;
- For locations $d_1, d_2, \ldots, d_k$, get their POI type frequencies $F_{d_1,r}, F_{d_2,r}, \ldots, F_{d_k,r}$, and compute the mean with noise of these frequencies by applying Guassian mechanism for $i = 1, 2, \ldots, M$:

$$F^*_{\mathcal{D},r}[i] = (\sum_{j=1}^{k} F_{d_j,r}[i] + \mathcal{N}(0, \sigma^2))/k, \qquad (8)$$

  where the variance σ is set to $\Delta\sqrt{2\ln(1.25/\delta)}/\epsilon$ according to Definition 2. ϵ and δ are privacy parameters.
- It replaces $F_{l,r}$ in Eq. (7) by $F^*_{\mathcal{D},r}$, and then optimizes the following problem and gets $\tilde{F}^*$:

$$\max_{\tilde{F}^*} \sum_{i=1}^{M} \frac{1}{R(i)} |\tilde{F}^*_i - F^*_{\mathcal{D},r}[i]|,$$
$$s.t. \;\; \frac{1}{M} \sum_{i=1}^{M} \frac{1}{F^*_{\mathcal{D},r}[i] + 1} |\tilde{F}^*_i - F^*_{\mathcal{D},r}[i]| \leq \beta, \qquad (9)$$
$$\tilde{F}^*_i \in \mathbb{N}^+, \;\; i = 1, 2, \ldots, M.$$

Theorem 4: The above defense mechanism achieves (ϵ, δ)-differential privacy.

Proof: First we analyze the sensitivity of sum of the POI type frequencies. Consider a pair of neighboring databases $F_{d_1,r}, F_{d_2,r}, \ldots, F_{d_j,r}, \ldots, F_{d_k,r}$ and $F_{d_1,r}, F_{d_2,r}, \ldots, F'_{d_j,r}, \ldots, F_{d_k,r}$, which differ in one POI frequency vector at one dimension. For any dimension i, $\sum_{j=1}^{k} F_{d_j,r}[i]$ will change at most $\max_d F_{d,r}[i]$, such that we can set the sensitivity at this dimension as $\max_d F_{d,r}[i]$.

Then we show that the publish of $F^*_{\mathcal{D},r}[i]$ is differentially private. We set the variance $\sigma = \Delta\sqrt{2\ln(1.25/\delta)}/\epsilon$. By Definition 2, we have that Eq. (8) achieves (ϵ, δ)-differential privacy. In the optimization (Eq. (9)), we do not need to access the original POI frequency vector. The proposed defense mechanism is a sequentially apply of Eq. (8) and Eq. (9). By Lemma 3, it is (ϵ, δ)-differentially private. ■

VI. Experimental Evaluation

We have implemented the proposed methods and evaluated them based on real-world user data traces. Specifically, we have carried out two sets of experiments:

- One set of experiments is on the performance of the attacks. Our results show that the proposed attack needs less than 25% of the search area compared with the existing attack in most cases. The attack leveraging trajectory uniqueness can increase the attack's success rate up to about 20% when $r = 0.5$.
- The other set of experiments is on the performance of the differentially private defense. Our results show that the proposed defense can mitigate the location re-identification attacks to less than 20% success rate in most settings while well preserving the utility of the POI aggregates.

A. Settings

The evaluation of fine-grained attack is conducted on four datasets: :(a) T-drive [3] user locations in Beijing, which contains trajectory data of 10,357 taxis in Beijing. We extract the trajectories which are within the given area of the city. (b) randomly generated user locations in Beijing; (c) Foursquare check-ins [4] in New York City, which contains 227,428 check-ins from 824 users; (d) randomly generated user locations in New York City. The evaluation of trajectory uniqueness is carried out on trajectories that are extracted from T-drive dataset. The evaluation of the proposed defense mechanism is performed on T-drive dataset and Foursquare NYC dataset. For each set of experiments, we randomly choose 1,000 locations or segments from the datasets for evaluation.

When evaluating the utility of the defense mechanism, we adopt the Top-K function as the target application and use the Jaccard Index [12] to measure the similarity between original POI type frequencies and protected POI type frequencies. Specifically, for the original POI type frequency vector $F_{l,r}$ and the protected POI type frequency vector $\tilde{F}^*$, we find the set of K types with highest frequencies in the vectors and denote them by $Top_K(F_{l,r})$ and $Top_K(\tilde{F}^*)$, respectively. Then we use the Jaccard Index to measure the utility of the protection mechanism:

$$|Top_K(F_{l,r}) \cap Top_K(\tilde{F}^*)| / |Top_K(F_{l,r}) \cup Top_K(\tilde{F}^*)|.$$

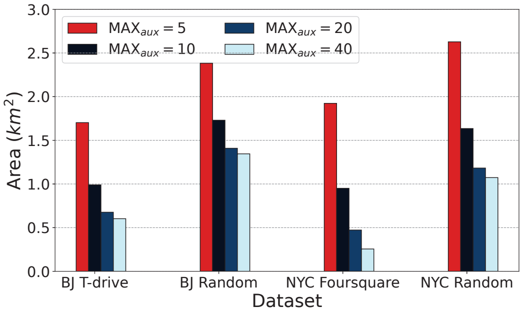


Figure 7: Search area with changing the number of auxiliary anchors ($r = 2.0\mathrm{km}$; the search area of Cao et al's attack is always $4\pi\mathrm{km}^2$ in this setting).

B. The Attacks

Fig. 6 shows the performance of fine-grained attack that we have proposed in Section IV-A when $\mathrm{MAX_{aux}} = 20$. We can find that this attack dramatically reduces the area's size that needs to search for the user's actual location. In about 80% cases, the proposed attack can reduce the search area to no more than a quarter of the search area required by Cao et al's attack. Furthermore, we can find that with the query range increasing, the fine-grained attack performs better on the search area reduction.

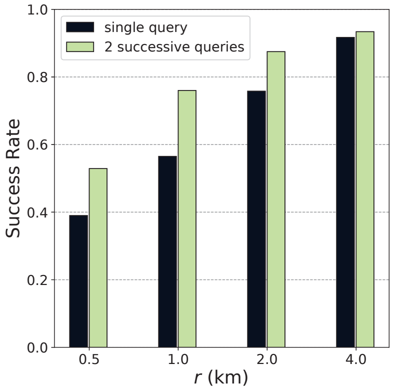
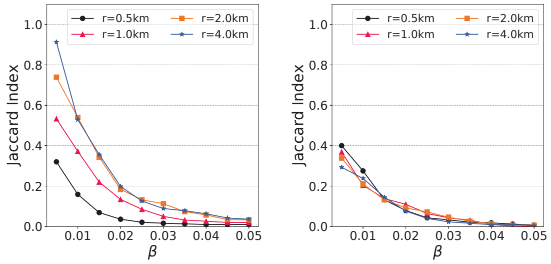


Figure 8: Exploiting the power of two successive queries.

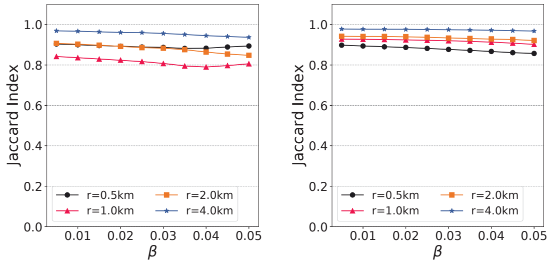
In Fig. 7, we can see that, with the number of auxiliary anchors increasing, the attack achieves better performance for all the four datasets. On average, for these four datasets, the fine-grained attack can reduce the size of the search area from $1.70\mathrm{km}^2$ to $0.60\mathrm{km}^2$, $2.38\mathrm{km}^2$ to $1.35\mathrm{km}^2$, $1.92\mathrm{km}^2$ to $0.26\mathrm{km}^2$, and $2.63\mathrm{km}^2$ to $1.07\mathrm{km}^2$, respectively, when the number of auxiliary anchors increases from 5 to 40. We can also find that the reduction brought by more auxiliary anchors decreases with the number of auxiliary anchors increasing. Therefore, it may not be the best choice to use all the auxiliary anchors because time cost will increase when more auxiliary anchors involve. In our experiments, let $\mathrm{MAX_{aux}} = 20$ could be a reasonable choice. We note that the search area of Cao et al's attack is always about $16.56\mathrm{km}^2$ when $r = 2\mathrm{km}$.



(a) BJ: T-drive

(b) NYC: Foursquare

Figure 9: The performance of the non-private defense mechanism (a lower success rate means better defense performance).



(a) BJ: T-drive

(b) NYC: Foursquare

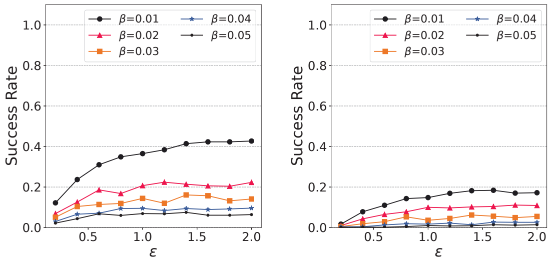
Figure 10: Jaccard index achieved by the non-private defense mechanism.

Fig. 8 shows the performance of the re-identification attack when two successive releases are leverages in Beijing with T-drive datasets. In our experiments, we extract the points in the trajectories satisfying the requirements: 1) the released POI type frequencies are changed because the adversary can be aware that if the POI type frequency is the same as the previous release, this release is useless; 2) the duration of two successive releases is less than 10 minutes, because when the duration is large, the user may start another new session of using the location-based services. We can observe from the results that the success rate is improved by using the knowledge of two successive queries. For $r = 0.5\mathrm{km}$, $1.0\mathrm{km}$, $2.0\mathrm{km}$, and $4.0\mathrm{km}$, the enhanced attack has 0.203, 0.146, 0.09, 0.001 gains on success rate, respectively. We can find that the gain is minimal when $r = 4.0\mathrm{km}$ because the performance of location re-identification is good enough with a large query range.

C. The Defense

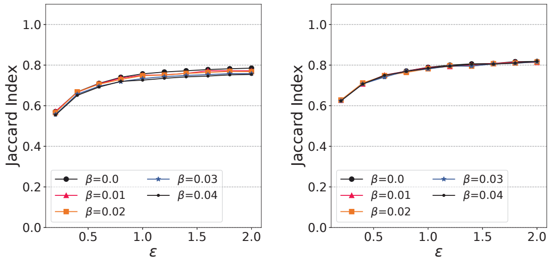
In this set of experiments, we adopt Top-10 as the target application. Fig. 9 and Fig. 10 shows the defense performance and the utility achieved by the non-private defense that is formulated in Eq. (7). We change the parameter β, which is used to balance the utility and defense performance in the formulation. We can find that with the larger β, the mechanism performs better on the defense while the utility only decreases slightly.

Fig. 11 and Fig. 12 shows the defense performance and the utility achieved by the differentially private defense mechanism that we have proposed in Section V-B. We set the spatial cloaking parameter $k = 20$, privacy parameter $\delta = 0.2$, and change ϵ from 0.2 to 2.0. We can observe that for various choices of β, the defense performence gets worse and the utility increases when the privacy budget increases, and the utility is merely affected by the parameter β.



(a) BJ: T-drive (b) NYC: Foursquare

Figure 11: The performance of the differentially private defense mechanism ($r = 2.0$km; a lower success rate means better defense performance).



(a) BJ: T-drive (b) NYC: Foursquare

Figure 12: Jaccard index achieved by the differentially private defense mechanism ($r = 2.0$km).

VII. Related Works

The related works of this paper fall into two categories: POI-based data analysis and location privacy.

A. POI-based Analysis and Applications

POI data have been widely used in the applications of spatial-based analysis and recommendations. Some works, *e.g.*, [13], [14] have been done for identifying the place with special meanings by leveraging POI data. Nishida et al. [14] propose a probabilistic identification method for personalized check-in in LBSs by analyzing users‘ past visited POIs. In [13], and a clustering-based algorithm is proposed for analyzing the attractive areas by using crowdsourced data.

POI-based recommendation also has been extensively studied, *e.g.*, [15], [16], [17], [18], [19]. In [17], [18], the authors study the problem of time-aware POI recommendation to recommend POIs for a user to visit at a given time. Bin et al. [15] propose a personalized recommendation framework by leveraging users' multi-aspect behavior and preferences. POI data have also been used for human behavior analysis, *e.g.*, [20], [21], [22], [23], [19], [24] Liu et al. [19] developed a systematic POI demand modeling framework to model POI demands by exploiting the daily needs of people identified from their large-scale mobility data. [24] revealed the collective intelligence of the spatial choices expressed in the mobility patterns of the people that live in a city.

B. Location privacy

A lot of research has been carried out on protecting location privacy, *e.g.*, [25], [26], [27], [28], [29], [30], [31]. Previous works on location privacy protection mainly focus on protecting users' physical locations. The most related work to our work is Cao et al. [1], which finds that even the actual locations are not revealed, the adversary can still re-identify users' location by the aggregated POI type distribution. They observe the phenomenon of location uniqueness, which is ubiquitous in many metropolises. A computationally efficient location re-identification method is also proposed by [1]. However, as we have mentioned above, their attack may not apply to launch practical attacks.

POI type frequency can be viewed as a type of location aggregate data. Therefore, studies on aggregate data privacy are also related to our work. The aggregate data are often considered a way to hinder the exposure of individuals' data [32]. However, a recent work [33] shows that an adversary with some prior knowledge can exploit aggregate information to improve his knowledge or even localize specific individuals that are part of the aggregates. Xu et al. [34] even extracted users' "trajectories" from aggregate mobility data without prior knowledge. In [35], a generic methodology is proposed for studying membership privacy in aggregated location data.

VIII. Conclusion

In this paper, we conducted an in-depth study of the location privacy problem in the presence of location uniqueness. We have also conducted a study to evaluate whether the existing protection methods can adequately defend against the location re-identification attack. Then results show that methods like sanitization, geo-indistinguishability, and spatial k-cloaking can hardly provide adequate location privacy protection in the presence of location uniqueness. Based on the existing location re-identification method, we present two practical variants, which achieve higher precision of locating a user and better re-identification performance, respectively. Furthermore, we propose a differentially private POI type frequency release mechanism, and the evaluation shows that it can provide adequate location privacy protection with acceptable utility loss.

Acknowledgment

The authors would like to thank all the reviewers for their helpful comments. This work was supported in part by National Key R&D Program of China under Grant 2020YFB1005900, NSFC-62002159, NSFC-61872176, NSFC-61972195, and the Leading-edge Technology Program of Jiangsu NSF (No. BK20202001). Q. Li was supported in part by US NSF Grant CNS-1816399 and the Commonwealth Cyber Initiative. S. Zhong and J. Hua are the corresponding authors.

References

[1] H. Cao, J. Feng, Y. Li, and V. Kostakos, "Uniqueness in the city: Urban morphology and location privacy," *IMWUT*, vol. 2, no. 2, pp. 62:1–62:20, 2018. [Online]. Available: http://doi.acm.org/10.1145/3214265

[2] O. Contributors, "Openstreetmap," *URL www. openstreetmap. org*, 2012.

[3] J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, and Y. Huang, "T-drive: driving directions based on taxi trajectories," in *18th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems, ACM-GIS 2010, November 3-5, 2010, San Jose, CA, USA, Proceedings*, D. Agrawal, P. Zhang, A. E. Abbadi, and M. F. Mokbel, Eds. ACM, 2010, pp. 99–108. [Online]. Available: http://doi.acm.org/10.1145/1869790.1869807

[4] D. Yang, D. Zhang, V. W. Zheng, and Z. Yu, "Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 1, pp. 129–142, 2015.

[5] D. Yu, Y. Li, F. Xu, P. Zhang, and V. Kostakos, "Smartphone app usage prediction using points of interest," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 4, pp. 174:1–174:21, 2017. [Online]. Available: https://doi.org/10.1145/3161413

[6] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3-4, pp. 211–407, 2014. [Online]. Available: https://doi.org/10.1561/0400000042

[7] Y.-W. Chang, C.-J. Hsieh, K.-W. Chang, M. Ringgaard, and C.-J. Lin, "Training and testing low-degree polynomial data mappings via linear svm," *Journal of Machine Learning Research*, vol. 11, no. Apr, pp. 1471–1490, 2010.

[8] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[9] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: differential privacy for location-based systems," in *ACM Conference on Computer and Communications Security*. ACM, 2013, pp. 901–914.

[10] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *MobiSys*. USENIX, 2003.

[11] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.

[12] Wikipedia contributors, "Jaccard index — Wikipedia, the free encyclopedia," https://en.wikipedia.org/w/index.php?title=Jaccard_index&oldid=965083523, 2020, [Online; accessed 16-August-2020].

[13] S. Kisilevich, F. Mansmann, and D. Keim, "P-dbscan: a density based clustering algorithm for exploration and analysis of attractive areas using collections of geo-tagged photos," in *Proceedings of the 1st international conference and exhibition on computing for geospatial research & application*. ACM, 2010, p. 38.

[14] K. Nishida, H. Toda, T. Kurashima, and Y. Suhara, "Probabilistic identification of visited point-of-interest for personalized automatic check-in," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2014, pp. 631–642.

[15] B. Liu, Y. Fu, Z. Yao, and H. Xiong, "Learning geographical preferences for point-of-interest recommendation," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2013, pp. 1043–1051.

[16] B. Liu and H. Xiong, "Point-of-interest recommendation in location based social networks with topic and location awareness," in *Proceedings of the 2013 SIAM International Conference on Data Mining*. SIAM, 2013, pp. 396–404.

[17] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann, "Time-aware point-of-interest recommendation," in *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2013, pp. 363–372.

[18] Q. Yuan, G. Cong, and A. Sun, "Graph-based point-of-interest recommendation with geographical and temporal influences," in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM, 2014, pp. 659–668.

[19] Y. Liu, C. Liu, X. Lu, M. Teng, H. Zhu, and H. Xiong, "Point-of-interest demand modeling with human mobility patterns," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2017, pp. 947–955.

[20] D. Yu, Y. Li, F. Xu, P. Zhang, and V. Kostakos, "Smartphone app usage prediction using points of interest," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, p. 174, 2018.

[21] G. Yu, J. Yuan, and Z. Liu, "Predicting human activities using spatio-temporal structure of interest points," in *Proceedings of the 20th ACM international conference on Multimedia*. ACM, 2012, pp. 1049–1052.

[22] Z. Yu, H. Xu, Z. Yang, and B. Guo, "Personalized travel package with multi-point-of-interest recommendation based on crowdsourced user footprints," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 1, pp. 151–158, 2016.

[23] N. J. Yuan, Y. Zheng, X. Xie, Y. Wang, K. Zheng, and H. Xiong, "Discovering urban functional zones using latent activity trajectories," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 3, pp. 712–725, 2015.

[24] S. Park, M. Bourqui, and E. Frias-Martinez, "Mobinsight: understanding urban mobility with crowd-powered neighborhood characterizations," in *Data Mining Workshops (ICDMW), 2016 IEEE 16th International Conference on*. IEEE, 2016, pp. 1312–1315.

[25] P. Golle and K. Partridge, "On the anonymity of home/work location pairs," in *International Conference on Pervasive Computing*. Springer, 2009, pp. 390–397.

[26] W. Tong, J. Hua, and S. Zhong, "A jointly differentially private scheduling protocol for ridesharing services," *IEEE Trans. Information Forensics and Security*, vol. 12, no. 10, pp. 2444–2456, 2017. [Online]. Available: https://doi.org/10.1109/TIFS.2017.2707334

[27] J. Hua, W. Tong, F. Xu, and S. Zhong, "A geo-indistinguishable location perturbation mechanism for location-based services supporting frequent queries," *IEEE Trans. Information Forensics and Security*, vol. 13, no. 5, pp. 1155–1168, 2018. [Online]. Available: https://doi.org/10.1109/TIFS.2017.2779402

[28] H. Shen, M. Zhang, H. Wang, F. Guo, and W. Susilo, "A lightweight privacy-preserving fair meeting location determination scheme," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 3083–3093, 2020.

[29] L. Wang, D. Zhang, D. Yang, B. Y. Lim, X. Han, and X. Ma, "Sparse mobile crowdsensing with differential and distortion location privacy," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 2735–2749, 2020.

[30] W. Jin, M. Xiao, M. Li, and L. Guo, "If you do not care about it, sell it: Trading location privacy in mobile crowd sensing," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, 2019, pp. 1045–1053.

[31] S. Narain, A. Ranganathan, and G. Noubir, "Security of gps/ins based on-road location tracking systems," in *2019 IEEE Symposium on Security and Privacy (SP)*, 2019, pp. 587–601.

[32] R. A. Popa, A. J. Blumberg, H. Balakrishnan, and F. H. Li, "Privacy and accountability for location-based aggregate statistics," in *Proceedings of the 18th ACM conference on Computer and communications security*. ACM, 2011, pp. 653–666.

[33] A. Pyrgelis, C. Troncoso, and E. De Cristofaro, "What does the crowd say about you? evaluating aggregation-based location privacy," *Proceedings on Privacy Enhancing Technologies*, vol. 2017, no. 4, pp. 156–176, 2017.

[34] F. Xu, Z. Tu, Y. Li, P. Zhang, X. Fu, and D. Jin, "Trajectory recovery from ash: User privacy is not preserved in aggregated mobility data," in *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2017, pp. 1241–1250.

[35] A. Pyrgelis, C. Troncoso, and E. D. Cristofaro, "Knock knock, who's there? membership inference on aggregate location data," *CoRR*, vol. abs/1708.06145, 2017. [Online]. Available: http://arxiv.org/abs/1708.06145