Synthesizing Privacy Preserving Traces: Enhancing Plausibility With Social Networks

Ping Zhao, *Member, IEEE*, Hongbo Jiang, *Senior Member, IEEE*, Jie Li, Fanzi Zeng, Zhu Xiao, *Senior Member, IEEE*, Kun Xie, and Guanglin Zhang

***Abstract*— Due to the popularity of mobile computing and mobile sensing, users' traces can now be readily collected to enhance applications' performance. However, users' location privacy may be disclosed to the untrusted data aggregator that collects users' traces. Cloaking users' traces with synthetic traces is a prevalent technique to protect location privacy. But the existing work that synthesizes traces suffers from the social relationship based de-anonymization attacks. To this end, we propose W^3-$tess$ that synthesizes privacy-preserving traces via enhancing the *plausibility* of synthetic traces with social networks. The main idea of W^3-$tess$ is to credibly imitate the temporal, spatial, and social behavior of users' mobility, sample the traces that exhibit similar three-dimension mobility behavior, and synthesize traces using the sampled locations. By doing so, W^3-$tess$ can provide "*differential privacy*" on location privacy preservation. In addition, compared to the existing work, W^3-$tess$ offers several salient features. First, both location privacy preservation and data utility guarantees are theoretically provable. Second, it is applicable to most geo-data analysis tasks performed by the data aggregator. Experiments on two real-world datasets, loc-Gwalla and loc-Brightkite, have demonstrated the effectiveness and efficiency of W^3-$tess$.**

***Index Terms*— Trace privacy, trace plausibility, cloaking, differential privacy.**

I. Introduction

CURRENT popularity of mobile computing and mobile sensing have enabled the capability to collect an increasing number of users' traces, which is expected to enhance many new applications, e.g., mobility management, identifying friends, map inference, etc. [1], [2]. However, users' location privacy can be disclosed to the untrusted data aggregator which collects users' traces, thus incurring personal information disclosure, such as life style, political beliefs, etc. [3], [4].

A prevalent method to protect users' location privacy is to cloak real traces with synthetic traces consisting of synthetic locations (see Fig. 1(a)) [5], [6]. Heuristic algorithms are proposed in [7], [8] that generate synthetic traces using interpolation, circles, grids, etc. However, the synthetic traces therein cannot credibly imitate users' mobility behavior, since each user has a consistent lifestyle and meaningful mobility. As a result, these synthetic traces are filtered out by attackers, resulting in users' location privacy disclosure. The follow-up techniques [9]–[11] generate synthetic traces by modeling users' mobility behavior, e.g., following consistent movement pattern, stopping at several locations to visit attractions, etc. But these techniques only consider the temporal or spatial behavior of users' mobility, while ignoring locations' semantic features. As such, attackers can distinguish the synthetic traces from the exact ones, according to the semantic features of users' locations, incurring the location privacy disclosure. Overall, all these work are susceptible to *location inference attacks* [12], where attackers identify synthetic traces based on users' mobility behavior.

The latest research [5] (dubbed as PULE) synthesizes *plausible* traces by considering locations' semantic features, which is characterized by users' temporal and spatial behavior (i.e., when they move and where they go). However, users' mobility behavior is also shaped by their social relationships (i.e., why they move) [13], [14], as the contact graph derived from users' traces can be structurally correlated with the social relationship graph in social networks, and therefore cloaked users can be identified by these structural correlations. As a result, PULE would suffer from the *social relationship based de-anonymization attacks* [6], [15]. For example, the users denoted by red nodes in the contact-graph (cf. Fig. 1(b)) are structurally correlated with users denoted by red nodes in the social relationship graph (cf. Fig. 1(c)), and thus users in contact-graph can be identified when attackers incorporate with the social relationship graph.

To tackle the above issue, in this paper, we propose W^3-$tess$, the first work that synthesizes privacy-preserving traces via enhancing the plausibility of synthetic traces with social networks. The idea of W^3-$tess$ is simple: we first characterize the three-dimension (i.e., temporal, spatial, and social) behavior of each user's mobility, sample the traces

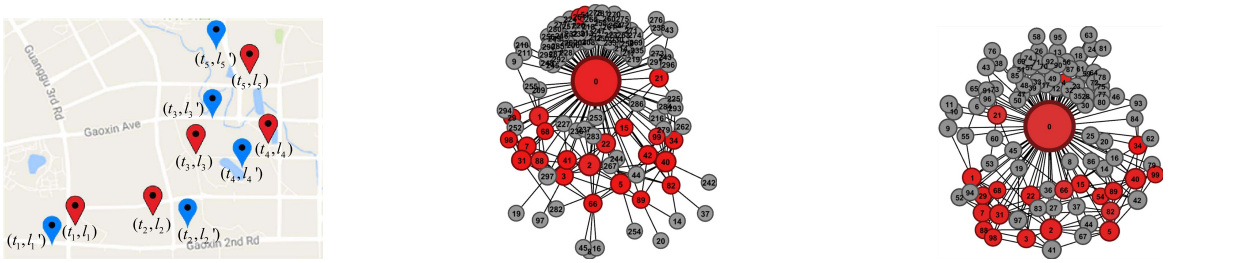
Manuscript received April 14, 2019; revised August 3, 2019; accepted October 6, 2019; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor A. Kuzmanovic. Date of publication October 28, 2019; date of current version December 17, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61572219, Grant 61732017, Grant 61502192, Grant 61671216, Grant 61471408, and Grant 61902060, in part by the Shanghai Sailing Program under Grant 19YF1402100, in part by the Fundamental Research Funds for the Central Universities under Grant 2232019D3-51, and in part by the Initial Research Funds for Young Teachers of Donghua University. *(Corresponding author: Hongbo Jiang.)*

P. Zhao is with the College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China, and also with College of Information Science and Technology, Donghua University, Shanghai 201620, China (e-mail: pingzhao2014ph@gmail.com).

H. Jiang, J. Li, F. Zeng, Z. Xiao, and K. Xie are with the College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China (e-mail: hongbojiang2004@gmail.com; jieli.csee@gmail.com; zengfanzi@hnu.edu.cn; zhxiao@hnu.edu.cn; xiekun@hnu.edu.cn).

G. Zhang is with the College of Information Science and Technology, Donghua University, Shanghai 201620, China (e-mail: glzhang1981@gmail.com).

Digital Object Identifier 10.1109/TNET.2019.2947452



(a) Synthetic locations and real locations. (b) Contact graph in loc-Gowalla dataset. (c) Relationship graph in loc-Gowalla dataset.

Fig. 1. (a) Cloaking a user's trace with a synthetic trace, where the red symbols refer to the user's trace, and the blue symbols refer to the synthetic trace. (b) An example of the contact graph in loc-Gowalla dataset, where nodes refer to users, numbers are users' IDs, and the edges indicate contacts. (c) An example of relationship graph in loc-Gowalla dataset, where the edges indicate friendships. The red nodes in (b) are structurally correlated with these red nodes in (c).

that exhibit similar three-dimension mobility behavior with the user, and then synthesize traces for fake users[1] using the sampled locations. However, we need to address several challenges:

(1) It is nontrivial to model users' social mobility behavior, since the influence of friends on users' mobility are dynamic. Our method is to investigate the temporal-spatial dynamic of friends' influence on users' mobility, based on which a three-dimension mobility model and a dynamic strategy are proposed accordingly.

(2) It is not easy to cloak users' traces for three-dimension mobility behavior, as synthetic traces in one dimension can be easily re-identified from another, e.g., traces cloaked in temporal dimension can be identified in spatial and social dimensions. Therefore, we propose to partition traces in temporal, spatial, and social dimensions, and synthesize traces using the traces that exhibit similar three-dimension mobility behavior.

(3) It is difficult to guarantee trace data utility, since cloaking users' traces with synthetic traces definitely deteriorates the data utility of users' traces. To this end, we propose to sample the traces that exhibit similar three-dimension behavior to achieve specially designed differential privacy so that the statistical feature of users' traces and synthetic traces is statistically the same as that of users' traces.

Compared with the existing work, W^3-$tess$ offers several salient features. First, both location privacy preservation and trace data utility guarantee (a.k.a QoS) are theoretically guaranteed in W^3-$tess$ (cf. Theorems 3 and 4), while most if not all of existing work lacks theoretical guarantee for QoS. Second, W^3-$tess$ is applicable to most geo-data analysis tasks as long as they are composable (cf. Corollary 5), while existing work can only preserve several features of traces gathered in specific geo-data analysis tasks.

The remainder of this paper is organized as follows. Section II introduces some preliminary knowledge. Section III describes the design of W^3-$tess$ in detail, followed by some theoretical discussions in Section IV. Section V evaluates the performance of W^3-$tess$, and Section VI reviews the related studies. Finally, Section VII concludes the paper.

[1]We define that a specific synthetic trace corresponds to a fake user.

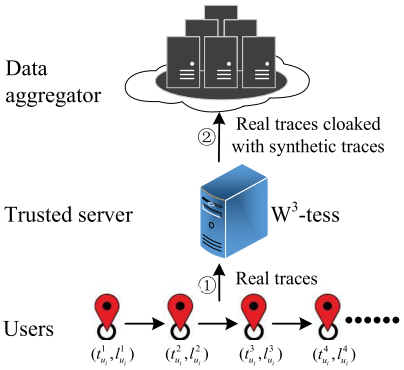


Fig. 2. The scenario considered in W^3-$tess$.

II. Preliminary

A. Scenario

We consider the scenario shown in Fig. 2. Users who take part in data aggregation, first send their traces to the trusted server. Then, the trusted server performs the proposed mechanism, W^3-$tess$, to generate synthetic traces, and cloaks users' traces by injecting these synthetic traces into the database of users' traces. Thereafter, the trusted server sends the cloaked traces (including users' traces and the synthetic ones) to the data aggregator. Upon receiving these cloaked traces, the data aggregator executes geo-data analysis tasks, e.g., searching for top-k frequently visited locations, updating map, etc., to support new applications.

Similar to existing work [6], [16]–[18], we consider the data aggregator is semi-honest. It strictly executes geo-data analysis tasks, but it may share these traces to, e.g., advertisers, illegal organizations, etc., for commercial interests. These advertisers, illegal organizations, etc., may be attackers that attempt to filter out the synthetic traces through launching, e.g., social relationship based de-anonymization attacks [6], [15], [19] or inference attacks [12], [20], to identify users' traces. If users' traces are identified by the attackers, it will incur more personal information disclosure, such as life style, political beliefs, etc. Therefore, it is desirable to enhance the plausibility of synthetic traces to avoid the synthetic traces from being identified by the attackers.

Note that users' locations are periodically updated in many APPs, and W^3-$tess$ does not have any constraints on

users' traces. Specifically, for example, Didi Taxi records and updates each user's location every $2 \sim 5$ minutes. Goole Map, Gaode Map, Baidu map, etc., also periodically updates users' locations when users use these applications. The recorded locations indicate that where the users are at a specific time. Furthermore, we have to emphasize that this paper focus on generating synthetic traces, given users' traces consisting of locations, and that how to record locations is not the main contribution of this manuscript. Moreover, no mater what the traces users take, W^3-$tess$ generates synthetic traces to protect users' location privacy, upon receiving users' traces.

B. Differential Privacy

ε-Differential privacy [21] is a method to prevent information disclosure when two input datasets $\mathcal{D}$ and $\mathcal{D}_{-t}$ differ only in one tuple t (i.e., neighboring databases). The standard interpretation of the notion, neighboring databases, is that any single tuple t is removed from the input $\mathcal{D}$ [22], [23]. An alternative is that $\mathcal{D}$ and $\mathcal{D}_{-t}$ include the same number of tuples, and all tuples in $\mathcal{D}$ and $\mathcal{D}_{-t}$ except for one tuple t are same [24]. We adopt the first kind of interpretation in our work. (δ, ε)-Differential privacy is a relaxation of the ε-differential privacy. Its formal definition is:

Definition 1: A randomized algorithm Al satisfies (ε, δ)-differential privacy *if for any dataset $\mathcal{D}$, any tuple $t \in \mathcal{D}$, and any $S \in Range(Al)$, the following holds with probability at least $(1-\delta)$ [25]:*

$$e^{-\varepsilon} \leq \frac{Pr[Al(\mathcal{D}) = S]}{Pr[Al(\mathcal{D}_{-t}) = S]} \leq e^{\varepsilon}, \tag{1}$$

where $Range(Al)$ is the output range of Al; $\mathcal{D}_{-t}$ is the dataset where a tuple t is removed from $\mathcal{D}$; $\varepsilon \in (0, 1)$ and $\delta \in (0, 1)$.

The following composition property of differential privacy describes the privacy preservation that a sequence of mechanisms $\mathcal{M}_i$ provide.

Theorem 1: Assume each of mechanisms $\mathcal{M}_i$, $i = (1, \ldots, r)$, provides ε_i-differential privacy, and $\mathcal{M}$ performs $\mathcal{M}_i$, $i = (1, \ldots, r)$ with independent randomness. Then $\mathcal{M}$ satisfies $\sum_{i=1}^{r} \varepsilon_i$-differential privacy [26].

Theorem 1 implies that if W^3-$tess$ meets the (ε, δ)-differential privacy in each of the three dimensions, i.e., temporal, spatial, and social dimensions, W^3-$tess$ will provide (ε, δ)-differential privacy on location privacy preservation (cf. Theorem 3).

In addition, (ε, δ)-differential privacy can guarantee that the output is statistically the same when any single tuple in $\mathcal{D}$ is removed. However, it cannot be directly applied to W^3-$tess$, as we aim to guarantee trace data's statistical characteristics when any k ($k \geq 1$) tuples are removed. To this end, we propose the (k, ε, δ)-differential privacy to guarantee trace data utility as follows.

Definition 2: A randomized algorithm Al meets (k, ε, δ)-differential privacy *if for any dataset $\mathcal{D}$, any tuple $D_i \in \mathcal{D}$, and any $S \in Range(Al)$*

$$e^{-\varepsilon} \leq \frac{Pr[Al(\mathcal{D}) = S]}{Pr[Al(\mathcal{D}') = S)]} \leq e^{\varepsilon}, \tag{2}$$

where input datasets $\mathcal{D} = (D_1, D_2, \ldots, D_n)^T$, $\mathcal{D}' = (D'_1, D'_2, \ldots, D'_n)^T$, D_i and D'_i are tuples in datasets $\mathcal{D}'$ and $\mathcal{D}$; for each D_i, D_j, D'_i, D'_j, $i < j$, $D_i \neq D'_i$, and $D_j \neq D'_j$, it holds $j - i + 1 \leq k$; D'_i (resp. D'_j) is obtained by removing or adding a row in D_i (resp. D_j); $Al(D_i) = S[i]$ (resp. $Al(D_j) = S[j]$).

Compared to (ε, δ)-differential privacy, (k, ε, δ)-differential privacy can guarantee that the output is statistically the same (i.e., the error is bounded by ε) with the probability $(1 - \delta)$ when k tuples are removed from or added to $\mathcal{D}$.

For example, assume the trace of a specific user u_1 is $D_1 = \{l_{11}, l_{12}, \ldots\}$. Assume the user u_1's trace D_1 is cloaked with another $(k-1)$ synthetic traces D'_2, D'_3, $\ldots$, D'_k, and each $D'_i = \{l'_{11}, l'_{12}, \ldots\}$ $(i = (2, 3, \ldots, k))$. In such a case, the input dataset $\mathcal{D}' = (D'_1, D'_2, \ldots, D'_k)^T = (D_1, D'_2, \ldots, D'_k)^T$, and $\mathcal{D} = (D_1, D_2, \ldots, D_k)^T = (D_1, \Theta, \ldots, \Theta)^T$ ($\Theta = \{O, O, \ldots\}$, O is an zero vector). $D_2, \ldots, D_k$ are obtained by removing the row in D'_2, D'_3, $\ldots$, D'_k. According to Definition 2, the trace data utility of the user u_1's trace is guaranteed when it is cloaked with these synthetic traces.

III. Design of W^3-$tess$

The design of W^3-$tess$ follows two steps: modeling mobility behavior and synthesizing traces, which are introduced in detail in this section.

A. Modeling Mobility Behavior

Users' mobility exhibits strong periodicity and seemingly random jumps, which is motivated by the existing work [13], [27]–[29]. It is validated in studies [13], [27]–[29] that users periodically move back and forth among home, workplaces, and vocation places (hereafter *centers*), which is mainly constrained by the time and space. For example, people alternate between home and workplaces throughout certain periods of the day on weekdays. In addition, users also seemingly randomly travel to meet friends, especially at evening and weekends, which is mainly constrained by their social relationships (i.e., influential friends). Therefore, we propose to map users' traces into three dimensions (i.e., temporal, spatial, and social dimensions). We refer to the locations in random jumps as "*social locations*", and the locations in periodic movement as "*non-social locations*".

Denote the trace of a specific user u_i as $\mathcal{L}_{u_i} = \{l^1_{u_i}, l^2_{u_i}, \ldots, l^j_{u_i}, \ldots, l^n_{u_i}\}$ when the user u_i posts his location at time t_n, where $l^j_{u_i} = (x^j_{u_i}, y^j_{u_i})$ is u_i's location at time t_j, $x^j_{u_i}$ and $y^j_{u_i}$ are latitude and longitude coordinates, and n is the total number of u_i's locations when he posts his location at t_n. We first define the three-dimension mobility model of u_i at time t_n as follows:

$$\mathcal{M}_{t_n} = (\Gamma_{t_n,s}, \Gamma_{t_n,sp}, \Gamma_{t_n,so}), \tag{3}$$

where $\mathcal{M}_{t_n}$ is the three-dimension mobility model, $\Gamma_{t_n,s}$ is the social behavior, $\Gamma_{t_n,sp}$ is the spatial behavior, $\Gamma_{t_n,so}$ is the temporal behavior, and $\Gamma_{t_n\overline{s}} = (\Gamma_{t_n,sp}, \Gamma_{t_n,so})$ is the non-social behavior.

1) Modeling Social Behavior: Next, we focus on modeling social behavior and non-social behavior. We first define $\mathcal{J}(\cdot)$ to distinguish the social locations where users are geographically in contact with their friends:

$$\mathcal{J}(l^j_{u_i}) = \sum_{i'=1}^{n_f} \sum_{j'=1}^{n_{i'}} D_{\alpha_d}(l^j_{u_i}, l^{j'}_{u_{i'}}) T_{\alpha_t}(t^j_{u_i}, t^{j'}_{u_{i'}}), \tag{4}$$

where $D_{\alpha_d}(\cdot) = 1$ when the Euclidean distance $\| \cdot \| \leq \alpha_d$, otherwise $D_{\alpha_d}(\cdot) = 0$; $T_{\alpha_t}(\cdot) = 1$ when modulus $| \cdot | \leq \alpha_t$, otherwise $T_{\alpha_t}(\cdot) = 0$; α_d and α_t are the spatial and temporal distance specified by the trusted server; $u_{i'}$ is one of u_i's friends; n_f is the number of u_i's friends; $n_{i'}$ is the total number of $u_{i'}$'s locations when u_i posts his location at time t_n; $l^{j'}_{u_{i'}}$ is the location of $u_{i'}$ at time $t_{j'}$. When $\mathcal{J}(l^j_{u_i}) > 0$, it means $l^j_{u_i}$ is a social location; Otherwise, it is a non-social location. This idea is motivated by the existing work [6], [30]. Study [6] introduced the observation that friends are in contact with each other when they are within certain spatio-temporal distance. Here, we first formalize this observation (cf. Eq. (4)). Work [30] inferred friendship according to the number of co-locations. In contrast, we decide whether friends are visiting the same place.

However, it is non-trivial to model users' social behavior as the influence of friends on users' social locations varies with time and space. To address this problem, we propose a dynamic strategy to update the most influential friends whenever users post their locations.

Denote $\mathcal{C}^1_{u_i}, \ldots, \mathcal{C}^m_{u_i}$, as m centers related to the user u_i. m is the number of centers. Studies [31], [32] introduced the notion, centers, and intuitively regarded home as users' mobility center without any theoretical explanation. This paper formalizes and generalizes the definition of centers in the following. Denote u_i's social locations by $\mathcal{L}_{u_i,s} = \{l^1_{u_i,s}, l^2_{u_i,s}, \ldots, l^j_{u_i,s}, \ldots, l^{n_s}_{u_i,s}\}$. We first define the following $\mathcal{C}_{t^j_{u_i}}$ as the movement center of u_i at time t_j, or,

$$\mathcal{C}_{t_j} = \arg \min_{\tau:1\leq\tau\leq m} \{\| l^j_{u_i,s}, \mathcal{C}^\tau_{u_i} \|\}. \tag{5}$$

So the movement center of u_i at time $t_{n_s,s}$ is $\mathcal{C}_{t_{n_s,s}}$. Denote $u_{i'}$'s influence on u_i's social locations by $\mathcal{I}(u_i, u_{i'})$, and $\mathcal{I}(u_i, u_{i'}) = \{\mathcal{I}_{t_{1,s}}(u_i, u_{i'}), \mathcal{I}_{t_{2,s}}(u_i, u_{i'}), \ldots, \mathcal{I}_{t_{n_s,s}}(u_i, u_{i'})\}$.

First, in terms of spatial dynamic, users are more likely to visit their friends living near their movement center [31], e.g., home, workplace, or vocation places. For example, on the basis of real world datasets from Gowalla and Brightkite, [31] discovered the probability that a user visits his friends within $1km$ from his home is ten times larger than the probability that he visits his friends who are $1000km$ away. In addition, users are more likely to visit their friends around their locations. In summary, we formalize the *spatial influence* $\mathcal{SI}_{t_{n_s,s}}$ at time $t_{n_s,s}$ which varies with space as follows:

$$\mathcal{SI}_{t_{n_s,s}}(u_i, u_{i'}) = \pi_1 e^{-\pi_2 \frac{\| l^{n_s}_{u_i,s}, \mathcal{C}^1_{u_{i'}} \|}{\| \mathcal{C}^1_{u_{i'}}, \mathcal{C}_{t_{n_s,s}} \|}}, \tag{6}$$

where π_1, π_2 are parameters, and $\pi_2 > 0$; $\pi_1 > 0$ when $\| \mathcal{C}^1_{u_{i'}}, \mathcal{C}_{t_{n_s,s}} \| \leq \alpha_d$, otherwise, $\pi_1 = 0$; $\mathcal{C}_{u_{i'},1}$ is the home of $u_{i'}$. The spatial influence implies the movement of user u_i is more likely to be affected by his friends living near his movement center. Moreover, friends around u_i's location $l^{n_s}_{u_i,s}$ exhibit stronger influence on u_i's mobility.

Second, friends' influence varies with time. Specifically, the friends checking in the same place with u_i at the same time exhibit the strongest influence on u_i's mobility. In addition, u_i tends to visit places his friends have visited within a α_t time interval. Thus, we define the *temporal influence* $\mathcal{TI}_{t_{n_s,s}}(u_i, u_{i'})$ at time $t_{n_s,s}$ to characterize the temporal dynamics of $u_{i'}$'s influence,

$$\mathcal{TI}_{t_{n_s,s}}(u_i, u_{i'}) = \sum_{j'=1}^{n_{i'}} \mathcal{J}'(j', j), \tag{7}$$

where $\mathcal{J}'(j', j) = \pi_3 e^{-ds_{\alpha_d}(l^j_{u_i,s}, l^{j'}_{u_{i'}}) ti_{\alpha_t}(t^j_{u_i,s}, t^{j'}_{u_{i'}})}$, $ds_{\alpha_d}(\cdot) = \| \cdot \|$ when $\| \cdot \| \leq \alpha_d$, otherwise $ds_{\alpha_d}(\cdot) = 0$; $ti_{\alpha_t}(\cdot) = | \cdot |$ when $| \cdot | \leq \alpha_t$, otherwise $ti_{\alpha_t}(\cdot) = 0$; $\pi_3 > 0$ when $ds_{\alpha_d}(\cdot) ti_{\alpha_t}(\cdot) > 0$, otherwise $\pi_3 = 0$.

Based on the definition of the spatial and temporal influence of $u_{i'}$, we define $u_{i'}$'s influence at time $t_{n_s,s}$ as follows:

$$\mathcal{I}_{t_{n_s,s}}(u_i, u_{i'}) = \omega_s \mathcal{SI}_{t_{n_s,s}}(u_i, u_{i'}) + \omega_t \mathcal{TI}_{t_{n_s,s}}(u_i, u_{i'}), \tag{8}$$

where ω_s and ω_t are parameters that are independent of the friend $u_{i'}$ and the time $t_{n_s,s}$, and specified by the trusted server.

Therefore, we characterise u_i's social behavior at time $t_{n_s,s}$ with the influence of his friends, $\Gamma_{t_{n_s,s}} = \{\mathcal{I}_{t_{n_s,s}}(u_i, u_1), \ldots, \mathcal{I}_{t_{n_s,s}}(u_i, u_{n_f})\}$, where $u_1, \ldots, u_{n_f}$ are u_i's friends.

2) Modeling Non-Social Behavior: We model the non-social locations based upon the notion that periodic movement always occurs around some centers, e.g., home, workplace [13]. For example, users move around their workplaces during the working hours, and around their homes in evening.

We first model the temporal behavior as follows. Denote the user u_i's non-social locations $\mathcal{L}_{u_i,\overline{s}} = \{l^1_{u_i,\overline{s}}, l^2_{u_i,\overline{s}}, \ldots, l^{n_{\overline{s}}}_{u_i,\overline{s}}\}$. When the user u_i is moving around a specific center $\mathcal{C}^j_{u_i}$ (computed according to Eq. (5)) at time $t_{n_{\overline{s}},\overline{s}}$, we denote the state of u_i's movement at this time as $\mathcal{S}(t_{n_{\overline{s}},\overline{s}}) = \mathcal{C}^j_{u_i}$. The corresponding probability which varies with time is

$$Pr[\mathcal{S}(t_{n_{\overline{s}},\overline{s}}) = \mathcal{C}^j_{u_i}] = \frac{1}{1 + \sum_{\substack{\tau=1 \\ \tau\neq j}}^{m} \frac{\gamma_{\mathcal{C}^\tau_{u_i}}}{\gamma_{\mathcal{C}^j_{u_i}}} e^{-(\frac{\pi}{12})^2 [\frac{(t-\lambda_{\mathcal{C}^j_{u_i}})^2}{2\chi^2_{\mathcal{C}^j_{u_i}}} - \frac{(t-\lambda_{\mathcal{C}^\tau_{u_i}})^2}{2\chi^2_{\mathcal{C}^\tau_{u_i}}}]}}, \tag{9}$$

where $\gamma_{\mathcal{C}^j_{u_i}}$ and $\gamma_{\mathcal{C}^\tau_{u_i}}$ are the proportions of non-social locations centered at $\mathcal{C}^j_{u_i}$ and $\mathcal{C}^\tau_{u_i}$, respectively. $\lambda_{\mathcal{C}^j_{u_i}}$ and $\chi_{\mathcal{C}^j_{u_i}}$ are the average and variance of the visiting time of the non-social locations which are centered at $\mathcal{C}^j_{u_i}$. Note that Eq. (9) is a generalization of the temporal behavior defined in [13].

Motivated by mobility centers introduced in [13], [32], we model the spatial behavior, i.e., the distribution of non-social locations that are centered at $\mathcal{C}^j_{u_i}$ using the

m-dimensional Gaussian distribution. Formally, we have

$$Pr[l^{n_{\overline{s}}}_{u_i,\overline{s}}] = \sum_{j=1}^{m} Pr[l^{n_{\overline{s}}}_{u_i,\overline{s}} | \mathcal{S}(t_{n_{\overline{s}},\overline{s}}) = \mathcal{C}^j_{u_i}] \times Pr[\mathcal{S}(t_{n_{\overline{s}},\overline{s}}) = \mathcal{C}^j_{u_i}] = \sum_{j=1}^{m} \mathcal{N}(\mathcal{U}_{\mathcal{C}^j_{u_i}}, \Sigma_{\mathcal{C}^j_{u_i}}) Pr[\mathcal{S}(t_{n_{\overline{s}},\overline{s}}) = \mathcal{C}^j_{u_i}], \quad (10)$$

where $\Sigma_{\mathcal{C}^j_{u_i}}$ is the covariance matrix of locations that are centered at $\mathcal{C}^j_{u_i}$, and $\mathcal{U}_{\mathcal{C}^j_{u_i}}$ is the mean of these locations. Therefore, we get the non-social behavior $\Gamma_{t_{n_{\overline{s}},\overline{s}},\overline{s}} = Pr[l^{n_{\overline{s}}}_{u_i,\overline{s}}]$.

Given the three-dimension mobility model, next we propose to synthesize traces to protect users' location privacy and guarantee trace data utility.

B. Synthesizing Traces

To protect user u_i's location privacy, the synthetic traces cloaked with users' trace should exhibit similar social, temporal, and spatial behavior as u_i's trace. Unfortunately, location cloaking suffers from the curse of dimensions [16], [33]. That is to say, a specific user's trace cloaked in social dimension can be identified when attackers incorporate the other two attributes in temporal and spatial dimensions. Thus we propose to partition three-dimension traces. Specifically, we group users' mobility behavior into three different dimensions (i.e., temporal, spatial, and social dimensions), and synthesize traces that exhibit similar behavior in each dimension.

Furthermore, to guarantee the trace data utility, synthetic traces should exhibit similar statistical features as u_i's trace in a specific geo-data analysis task, e.g., searching for top-k_t frequently visited locations, etc. Differential privacy is expected to solve this problem. But existing work relies on output perturbation to satisfy differential privacy while the output of the geo-data analysis task cannot be perturbed by users. Moreover, different statistical features of traces are collected in various geo-data analysis tasks. To this end, we propose $(\mathcal{F}, k, \rho)$ sampling ($\mathcal{F}$ is a specific geo-data analysis task, k is u_i's privacy parameter, and $0 < \rho < 1$) to provide an alternative approach (i.e., input perturbation) to achieve specially designed differential privacy. As a result, the statistical feature (analyzed in $\mathcal{F}$) of u_i's trace cloaked with synthetic traces is statistically the same as that of the user's trace (to be proved in Section IV).

In summary, the main idea of synthesizing traces is to sample from the so called seed locations in each of the three dimensions, and synthesize locations using the sampled locations that exhibit similar statistical feature (analyzed in $\mathcal{F}$) as u_i's trace.

First, to synthesize plausible social locations, we propose to use the locations of u_i's influential friends as the social seed locations, since influential friends exhibit similar social behavior with u_i. We first rank u_i's friends by the influence $\mathcal{I}_{t^{n_s}_{u_i,s}}(u_i, u_{i'})$ (cf. Section III-A.1), and select N_f ($N_f > k$) most influential friends. The social locations of these most influential friends are regarded as the social seed locations.

Second, to synthesize plausible non-social locations, we propose to select the traces of other users who exhibit similar temporal and spatial behavior as u_i. The locations in N_f most similar traces are treated as the non-social seed locations. To this end, we use the Kullback-Leibler Divergence (KL) to define the mobility similarity in temporal and spatial dimensions:

$$tr(\Sigma^{-1}_{\mathcal{C}^{j'}_{u_{i'}}} \Sigma_{\mathcal{C}^j_{u_i}}) + (\mathcal{U}_{\mathcal{C}^{j'}_{u_{i'}}} - \mathcal{U}_{\mathcal{C}^j_{u_i}})^T \times \Sigma^{-1}_{\mathcal{C}^{j'}_{u_{i'}}} (\mathcal{U}_{\mathcal{C}^{j'}_{u_{i'}}} - \mathcal{U}_{\mathcal{C}^j_{u_i}}) - \ln(\frac{\det \Sigma_{\mathcal{C}^j_{u_i}}}{\det \Sigma_{\mathcal{C}^{j'}_{u_{i'}}}}). \quad (11)$$

It quantifies the similarity of user u_i and a specific user $u_{i'}$ in temporal and spatial dimensions, and moreover u_i exhibits more similar mobility features with the user $u_{i'}$ when KL is small. Finally, these selected users' locations are regard as the non-social seed locations.

Denote the synthetic traces at time $t^{n-1}_{u_i}$ as $\{\mathcal{L}^1_{t^{n-1}_{u_i}}, \mathcal{L}^2_{t^{n-1}_{u_i}}, \dots, \mathcal{L}^j_{t^{n-1}_{u_i}}, \dots, \mathcal{L}^k_{t^{n-1}_{u_i}}, \dots\}$. And a synthetic trace $\mathcal{L}^j_{t^{n-1}_{u_i}} = \{\mathcal{L}^j_{t^{n-1}_{u_i},s}, \mathcal{L}^j_{t^{n-1}_{u_i},\overline{s}}\}$, $(j = 1, 2, \dots, k, \dots)$, where $\mathcal{L}^j_{t^{n-1}_{u_i},s}$ (resp. $\mathcal{L}^j_{t^{n-1}_{u_i},\overline{s}}$) is the set of social locations (resp. non-social) in synthetic trace $\mathcal{L}^j_{t^{n-1}_{u_i}}$. To guarantee the trace data utility, the synthetic traces at time $t^n_{u_i}$ should be statistically similar to that of users' traces. Specifically, we first select location $l_{u_i,s}$ (resp. $l_{u_i,\overline{s}}$) from the social (resp. non-social) seed locations that meet the following constraints: there exists at least one synthetic trace $\mathcal{L}^j_{t^{n-1}_{u_i}}$ $(j = (1, 2, \dots, k, \dots))$ so that

$$\mathcal{F}(\mathcal{L}_{u_i,s}) = \mathcal{F}(\mathcal{L}^j_{t^{n-1}_{u_i},s} \cup \{l_{u_i,s}\}) \pm \Delta f,$$
$$\mathcal{F}(\mathcal{L}_{u_i,\overline{s}}) = \mathcal{F}(\mathcal{L}^j_{t^{n-1}_{u_i},\overline{s}} \cup \{l_{u_i,\overline{s}}\}) \pm \Delta f, \quad (12)$$

where Δf is the predefined threshold by the trusted server. Denote the set of these locations $l_{u_i,s}$ (resp. $l_{u_i,\overline{s}}$) as $\mathcal{L}'_{u_i,s}$ (resp. $\mathcal{L}'_{u_i,p}$). Thereafter, we sample from $\mathcal{L}'_{u_i,s}$ and $\mathcal{L}'_{u_i,\overline{s}}$ with the probability ρ. The sampled locations that are geographically consistent are regarded as synthetic locations. Denote the corresponding synthetic traces at time $t^n_{u_i}$ as $\{\mathcal{L}^1_{t^n_{u_i}}, \mathcal{L}^2_{t^n_{u_i}}, \dots, \mathcal{L}^j_{t^n_{u_i}}, \dots, \mathcal{L}^k_{t^n_{u_i}}, \dots\}$.

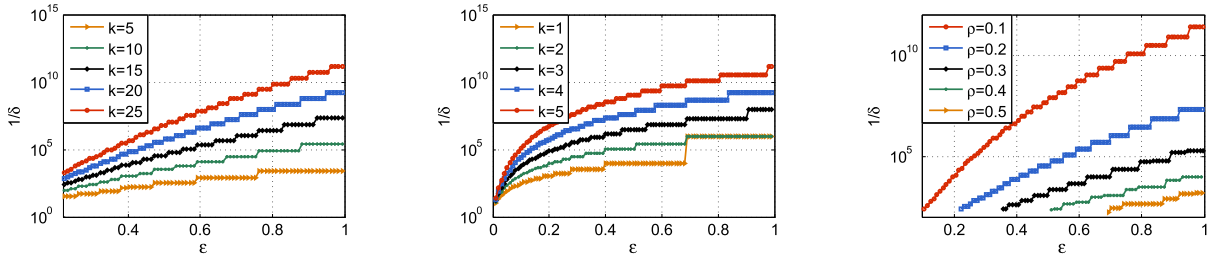
Lastly, to protect the location privacy of users whose locations are selected to synthesize traces (hereafter *participating users*), the statistical features of synthesize traces should be different from these participating users so that attackers cannot infer these participating users' locations by observing the synthesize traces. In particular, we select no less than k synthetic traces which meet the following constraints as the final output:

$$\mid \mathcal{F}(\mathcal{L}_{u_\tau}) - \mathcal{F}(\mathcal{L}^j_{t^n_{u_i}}) \mid \geq \Delta f', \quad (13)$$

where $\mathcal{L}_{u_\tau}$ is the trace of the user u_τ whose location is selected to generate j^{th} synthetic trace at time $t^n_{u_i}$ (namely, u_τ is a participating user).

IV. Theoretical Analysis

In this part, we theoretically analyze the privacy preservation and trace data utility guarantee that W^3-$tess$ provides. To begin with, we introduce an theorem in the existing work [24], which is shown as follows.



(a) $\rho = 0.2$. (b) $\rho = 0.01$ (c) $k = 15$.

Fig. 3. Location privacy preservation varies with parameters ε, δ, ρ, and k, where δ and ε mean that the error of three dimension mobility behavior of synthetic traces and users' traces is within ε with probability $(1-\delta)$, ρ is the sampling probability, and k is the number of synthetic traces.

Theorem 2: The version of randomized response described above is $(\ln 3; 0)$*-differentially private.*

It implies that any a kind of the randomized response meets differential privacy. Furthermore, it provides an alternative approach (without injecting noise) to achieve differential privacy. Motivated by Theorem 2, we try to theoretically prove that W^3-$tess$ meets differential privacy, as it integrates the $(\mathcal{F}, k, \rho)$ sampling.

In addition, according to Definitions 1 and 2, both (ε, δ)-differential privacy and (k, ε, δ)-differential privacy can quantify the privacy preservation and trace data utility guarantee. Specifically, a smaller ε indicates the outputs of the algorithm Al on these two neighboring datasets are more statistically similar. Moreover, a smaller δ means the outputs on these two neighboring datasets are statistically similar with a larger probability. Therefore, on one hand, smaller ε and δ strengthen the trace data utility as the outputs are more statistically similar with a larger probability. On the other hand, smaller ε and δ improve the privacy preservation, since it is more difficult for attackers to infer the participation or absence of a specific tuple in the dataset.

In summary, we will theoretically prove that W^3-$tess$ meets (ε, δ)-differential privacy, to quantify the privacy preservation. Furthermore, we will theoretically prove that W^3-$tess$ meets (k, ε, δ)-differential privacy, to quantify the trace data utility, since (k, ε, δ)-differential privacy can guarantee that the output is statistically the same when less than k tuples are removed from or added to the dataset.

A. Privacy Preservation Analysis

W^3-$tess$ protects a specific user's location privacy via cloaking user's trace with more than k synthetic traces that exhibit similar three dimension mobility behavior. In the following, we take one step further to theoretically prove the privacy preservation that W^3-$tess$ provides using (ε, δ)-differential privacy. We have the following results:

Theorem 3: W^3*-tess provides* (ε, δ)*-differential privacy for any a specific user's trace, with* $0 < \rho < 1$, $\varepsilon \geq -ln(1-\rho)$, $\delta = \max_{n: n \geq (\frac{k}{\varrho} - 1)} \sum_{i > \varrho n}^{n} F(i, n, \rho)$, $\varrho = \frac{(e^{\varepsilon} - 1 + \rho)}{e^{\varepsilon}}$, *and* $F(i, n, \rho) = \prod_{i'=0}^{i-1} \frac{n-i'}{i-i'} \rho^i (1-\rho)^{n-i}$.

Proof: See Appendix A. ■

We next focus on investigating the impact of the four parameters δ, ε, ρ, and k on the privacy preservation, according to Theorem 3. First we set $k \in [5, 25]$, $\rho = 0.2$, and $n = 10000$. It can be seen from Fig. 3(a), a larger k results in a smaller δ. For example, δ decreases by 10^{-2} at $\varepsilon = 1$ when increasing k from 20 to 25. In addition, a larger k leads to a smaller ε, given the parameter δ. In summary, increasing k can enhance the privacy preservation. Second, we vary the parameter k within $[1, 5]$ with $\rho = 0.01$, as shown in Fig. 3(b). As we can see, sampling with probability ρ strengthens location privacy preservation, as $\delta < \rho$ when $k = 1$. Furthermore, when $k \geq 2$, location anonymization enhances location privacy preservation, with δ significantly decreasing and $\delta \ll \rho$. Lastly, we set ρ within $[0.1, 0.5]$ and $k = 15$ to investigate the impact of ρ on location privacy preservation, which is shown in Fig. 3(c), where δ and privacy budget ε rapidly decrease with decreasing ρ. Thus, decreasing ρ can strengthen location privacy preservation.

B. Trace Data Utility Guarantee Analysis

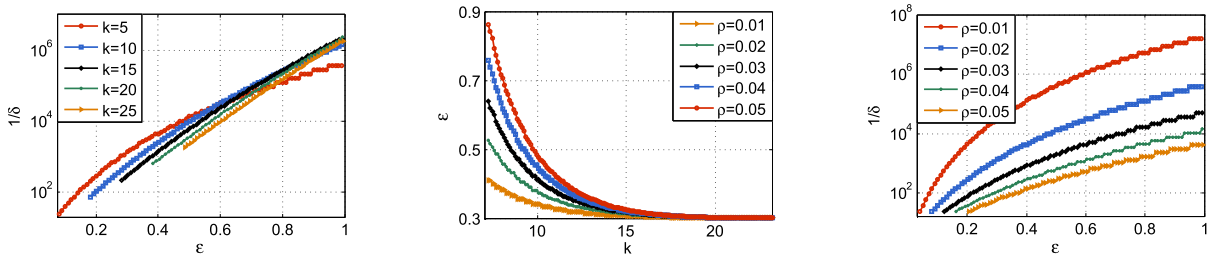
The synthetic traces deteriorate the trace data utility for user u_i, since u_i is cloaked with the synthetic traces. To guarantee the trace data utility, the statistical feature (analyzed in $\mathcal{F}$) of u_i's trace $\mathcal{L}_{u_i}$ should be statistically the same when no less than k synthetic traces are added to $\mathcal{L}_{u_i}$. According to Definition 2, (k, ε, δ)-differential privacy can guarantee that the output is statistically the same (i.e., the error is bounded by ε) with the probability $(1-\delta)$ when k tuples are removed from or added to the input D. Therefore, in the following, we bound the trace data utility guarantee using the two parameters ε and δ.

Theorem 4: W^3*-tess guarantees the trace data utility of any a specific user* u_i *with* (k, ε, δ)*-differential privacy in the geo-data analysis task* $\mathcal{F}$*, where* $0 < \rho < 1$, $\varepsilon \geq -ln(1-\rho)^{(k-1)}$, $\delta = \max_{n: n \geq (\frac{k}{\varrho} - 1)} \sum_{i > \varrho n}^{n} F(i, n, \rho)$, *and* $\varrho = \frac{1}{n} \arg\min_i [f(i) \geq e^{\varepsilon}]$,

$$f(i) = \frac{n(n-1)\cdots[n-(k-1)+1]}{(n-i)(n-i-1)\cdots[n-i+1-(k-1)]}(1-\rho)^{k-1}.$$

Proof: See Appendix B. ■

Let us present some explanations of Theorem 4. First, we set parameter $k \in [5, 25]$, and $\rho = 0.02$. As parameter ε is bounded by k and ρ, ε is bounded within different ranges as shown in Fig. 4(a). We can see that δ and ε decrease with



(a) $\rho = 0.02$. (b) $\delta < 10^{-6}$. (c) $k = 5$.

Fig. 4. Trace-data utility varies with parameters ε, δ, ρ, and k, where δ and ε mean that the error of statistical feature of synthetic traces and users' traces is within ε with probability $(1-\delta)$, ρ is the sampling probability, and k is the number of synthetic traces.

decreasing k, indicating the improvement of the trace data utility. In addition, δ decreases with the increasing ε. This shows the trade-off between ε and δ in terms of the data utility guarantee. Next, we focus on investigating the relationship between ε and k when δ is set less than 10^{-6} (cf. Fig. 4(b)). As shown in Fig. 4(b), given the parameter δ, a larger ε will lead to a smaller k. Lastly, we study the impact of ρ on the data utility guarantee, which is shown in Fig. 4(c). It shows that δ and ε increase with the increasing ρ. So decreasing sampling probability ρ strengthens the data utility guarantee.

It is important to note that Theorem 4 holds when the geo-data analysis task $\mathcal{F}$ is composable. Next, we use the following corollary to explain this in detail.

Corollary 5: W^3-tess can guarantee trace data utility with (k, ε, δ)-differential privacy in all geo-data analysis tasks $\mathcal{F}$ as long as they are composable, i.e., $\mathcal{F}(\Omega) = \sum_{\tau=1}^{N} \mathcal{F}(\mathcal{L}_{u_\tau})$, or $\mathcal{F}(\Omega) = \sum_{\tau=1}^{N} \mathbb{F}_\tau(\mathcal{L}_{u_\tau})$, where Ω is the set of all users' traces when u_i posts his location at time $t_{u_i}^n$; N is the number of all users; $\mathcal{L}_{u_\tau}$ is the trace of user u_τ; $\mathbb{F}_\tau$ is one or several subtasks on u_τ's trace.

For the first case $\mathcal{F}(\Omega) = \sum_{\tau=1}^{N} \mathcal{F}(\mathcal{L}_{u_\tau})$, the synthetic trace should exhibit similar statistical feature $\mathcal{F}(\mathcal{L}_{u_i})$ with u_i's trace. In contrast, the synthetic trace should exhibit similar statistical feature $\mathbb{F}_\tau(\mathcal{L}_{u_i})$ with u_i's trace when $\mathcal{F}(\Omega) = \sum_{\tau=1}^{N} \mathbb{F}_\tau(\mathcal{L}_{u_\tau})$. Overall, if $\mathcal{F}$ is not composable, W^3-tess cannot guarantee the trace data utility. In addition, since the composable geo-data analysis task $\mathcal{F}$ is involved in $(\mathcal{F}, k, \rho)$ sampling, W^3-tess is applicable to various geo-data analysis tasks even different statistical features of traces are collected in each geo-data analysis task.

V. Performance Evaluation

In this section, we evaluate the location privacy preservation and trace data utility guarantee of W^3-tess, using two real-world datasets, loc-Gwalla and loc-Brightkite [34].

A. Datasets and Setup

Dataset loc-Gwalla consists of $196,591$ nodes (i.e., users), $950,327$ edges (i.e., friendships) and 6.4 million check-ins (i.e., locations) from Feb. 2009 to Oct. 2010. Dataset loc-Brightkite records 4.5 million check-ins of $58,228$ nodes from Apr. 2008 to Oct. 2010 and there are 214078 edges among users. The average degree and graph density in loc-Gwalla and loc-Brightkite are 9.7, $4.92E$-5, and 7.5, $1.32E$-5. For each month, we delete the users who posted their locations less than 100 times.

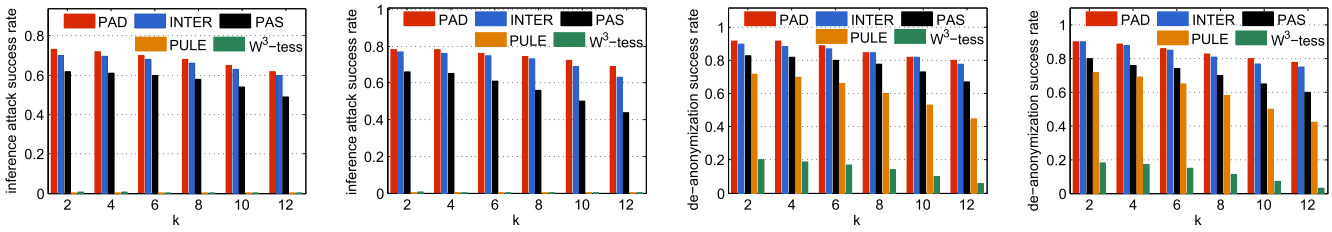
To evaluate W^3-tess, we compare it with the latest work [5] PULE and three heuristic algorithms dubbed as: PAD [7], INTER [8], PAS [11]. PAD and INTER generate traces based on a virtual grid and interpolation strategies respectively; PAS generates fake users pausing at some locations; PULE considers the semantic features of users' locations. In addition, we focus on the following metrics for performance evaluation.

Location Privacy Preservation Metrics: The related work, heuristing algorithms are susceptible to inference attacks, and the latest work suffers from the social relationship based de-anonymization attacks. So, we consider the above two kind of state-of-the art attacks: inference attacks [12] and social relationship based de-anonymization attacks [6], to validate the performance of W^3-tess against such two kind of attacks. Specifically, in inference attacks [12], attackers try to identify each user's trace from the anonymous traces consisting of users' traces and synthetic traces generated by W^3-tess, via finding the most likely assignment of users to obfuscated traces and maximizing the probability of all users using a joint assignment algorithm. In contrast, in social relationship based de-anonymization attacks [6], attackers first construct the contact graph according to users' locations, and then search for the optimal mapping between the contact graph and the social network graph using the method, Distance Vector, aiming to distinguish users' traces from the synthetic traces.

Furthermore, to quantify the location privacy preservation, we use two metrics, inference attack success rate $\overline{P}_{in}$, and social relationship based de-anonymization attack success rate $\overline{P}_{so}$. Both $\overline{P}_{in}$ and $\overline{P}_{so}$ are proportion of users whose locations (i.e., traces) can be identified by attackers.

Trace Data Utility Guarantee Metrics: As different statistical features of traces are collected in different geo-data analysis tasks, we quantify the trace data utility guarantee according to the specific geo-data analysis task.

In continuous location based services (LBS), users' locations are continuously sent to the LBS server along with the synthetic locations. Upon receiving the results sent by LBS server, users can obtain the accurate results through filtering



(a) $\overline{P}_{in}$ in loc-Gwalla. (b) $\overline{P}_{in}$ in loc-Brightkite. (c) $\overline{P}_{so}$ in loc-Gwalla. (d) $\overline{P}_{so}$ in loc-Brightkite.

Fig. 5. Location privacy preservation in both datasets loc-Gwalla and loc-Brightkite varying with the number of synthetic traces corresponding to each user's trace k. (a) (b) Inference attack success rate $\overline{P}_{in}$ and (c) (d) social relationship based de-anonymization attack success rate $\overline{P}_{so}$ in both loc-Gwalla and loc-Brightkite.

out the results corresponding to synthetic locations using the existing indexing schemes, e.g., air indexing in work [35]. Therefore, no data utility loss in continuous LBS.

In addition, we consider the following geo-data analysis tasks to quantify the trace data utility: (1) *Top-κ point of interest (POI) extraction*: searching for top-κ frequently visited locations. We compute the distribution of visits among top-κ frequently visited locations. (2) *Map inference*: updating map using users' traces. We investigate the KL-divergence KL of the distribution of visits among locations both in synthetic traces and real traces. (3) *Modeling users' mobility*: studying the mobility behavior. Thus, we study the mobility similarity sim (cf. Eq. (11)) of synthetic traces and real traces. (4) *Friendship inference*: inferring friendship from the geographic coincidences. So we investigate the probability of friendship deduced from synthetic traces and real traces as the existing work [36]. Note that the deduced friendship is compared to the social relationship in social network (i.e., Gwalla and Brightkite). (5) *Influential friends inference*: inferring influential friends according to users' traces. We define the probability of error as $1-(\Pi_s \cap \Pi_r)/no_f$, where Π_s and Π_r are the sets of top-no_f influential friends inferred from synthetic traces and real traces respectively [37].

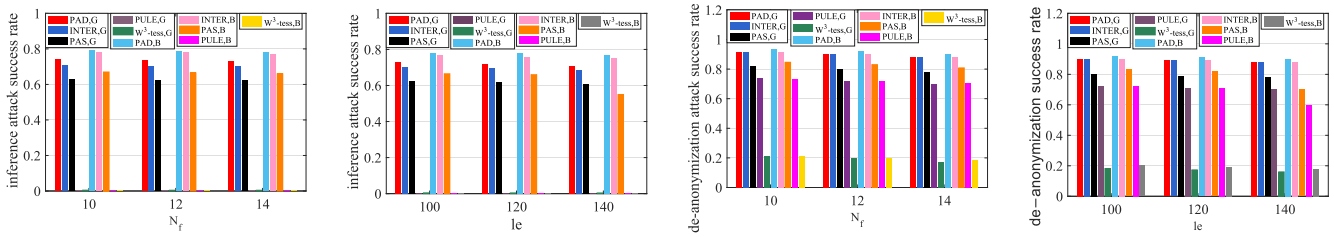
Computation Cost Metric: In addition, we also investigate the computation cost to evaluate the scalability of the proposed W^3-$tess$. Specifically, in W^3-$tess$, the computation cost refers to the computation operations of the generation of plausible traces. In PAD [7], we investigate the computation overhead of the circle-based dummy generation, and in INTER [8], we consider the computation cost of the interpolation strategies. Moreover, in PAS [11] and PULE [5], generation of fake traces dominates the whole computation cost. What's more, we consider the impact of the number of synthetic traces corresponding to each user's trace k on the computation cost, since the number of synthetic traces significantly affects the computation overhead.

Other default parameters are set as follows: $\alpha_d = 25$km; $\alpha_t = 2000$s; $\pi_1 = \pi_2 = 0.5$, $\pi_3 = 1$; $\omega_s = 0.3$, $\omega_t = 0.7$; $\rho = 0.2$; the number of synthetic traces corresponding to each user's trace $k \in (2, 12)$; in friendship inference, the number of co-occurrences is 5, the temporal range is 24 hours, the number of most influential friends $N_f = 14$, and the length of traces $le = 100$. In addition, Δf and $\Delta f'$ in above five geo-data analysis tasks are: $\{1, 15\}$; $\{0.1, 0.6\}$; $\{0.2, 0.5\}$; $\{0.1, 0.3\}$; $\{0.1, 0.5\}$. Simulations are implemented in C++ and conducted on a desktop PC with an Intel Core i7 3.41G Hz processor and 8G RAM.

B. Location Privacy Preservation

1) Location Privacy Preservation Against Inference Attacks: As shown in Figs. 5(a) and 5(b), in both loc-Gwalla and loc-Brightkite, PULE and W^3-$tess$ outperform PAD, INTER, and PAS, with the inference attack success rate $\overline{P}_{in}$ significantly less than that in PAD, INTER, and PAS. The superiority of PULE and W^3-$tess$ is attributed to that the three simple heuristic algorithms fail to model users' mobility behavior and thus are susceptible to location inference attacks. In addition, compared to PULE, W^3-$tess$ provides comparable privacy preservation against inference attacks, because synthetic traces in both PULE and W^3-$tess$ exhibit similar temporal and spatial features to the real traces. Lastly, $\overline{P}_{in}$ in PAD, INTER, and PAS decreases with the number of synthetic traces k. That is because it is more difficult to distinguish real traces from synthetic traces when more synthetic traces are generated. Furthermore, we have also investigated the impact of the number of most influential friends N_f and the length of traces le on the inference attack success rate $\overline{P}_{in}$, which is shown in Figs. 6(a) and 6(b). It shows that $\overline{P}_{in}$ in the five algorithms and the two datasets slowly decrease with the increasing N_f and le. Larger N_f and le mean more locations of friends used to generate synthetic traces, therefore enhancing the plausibility of synthetic traces. As a result, the $\overline{P}_{in}$ is decreased. In addition, we can observe that the $\overline{P}_{in}$ in algorithms PAD, INTER, and PAS, are less robust to N_f and le. It because PULE and W^3-$tess$ are sophisticated algorithms, and can credibly imitate temporal and spatial features, compared to algorithms PAD, INTER, and PAS.

2) Location Privacy Preservation Against Social Relationship Based De-Anonymization Attacks: Figs. 5(c) and 5(d) show the privacy preservation against social relationship based de-anonymization attacks in both loc-Gwalla and loc-Brightkite. It can be observed that W^3-$tess$ outperforms the four algorithms with $\overline{P}_{so}$ significantly less than that in these algorithms. Because PAD, INTER, PAS, and PULE ignore users' social behavior, and therefore the contract graph in these algorithms are structurally correlated with the social relationship graph. Furthermore, $\overline{P}_{so}$ decreases with k, as more synthetic traces confuse attackers. In addition, $\overline{P}_{so}$ in loc-Gwalla



(a) Impact of N_f on $\overline{P}_{in}$. (b) Impact of le on $\overline{P}_{in}$. (c) Impact of N_f on $\overline{P}_{so}$. (d) Impact of le on $\overline{P}_{so}$.

Fig. 6. Location privacy preservation in both datasets loc-Gwalla and loc-Brightkite varying with the number of most influential friends N_f and the length of traces le. (a) (b) Inference attack success rate $\overline{P}_{in}$ varying with N_f and le. (c) (d) social relationship based de-anonymization attack success rate $\overline{P}_{so}$ impacted by N_f and le.

is a bit larger than that in loc-Brightkite as loc-Gwalla exhibits higher average degree and graph density, and thus more social relationship information in loc-Gwalla is available to correlate users in the contact graph. The superiority of W^3-$tess$ to PULE indicates that the users' social behavior should be protected except for the temporal and spatial behavior. Lastly, as shown in Figs. 6(c) and 6(d), we investigate the impact of N_f and le on $\overline{P}_{so}$. We can observe that $\overline{P}_{so}$ in the five algorithms and the two datasets slowly decrease with the increasing N_f and le. The reasons are as analyzed above, larger N_f and le enhance plausibility of synthetic traces. Moreover, $\overline{P}_{so}$ in the five algorithms and the two datasets are less robust to N_f than $\overline{P}_{in}$. It is attributed to that de-anonymization attacks [6] heavily rely on the social feature of traces.

In summary, W^3-$tess$ outperforms all existing work, protecting users' location privacy against both inference attacks and social relationship based de-anonymization attacks.

C. Trace Data Utility Guarantee

1) Top-k_t POI Extraction: Figs. 7(a) and 7(b) show the distribution of visiting proportion of top-50 POI in both loc-Gwalla and loc-Brightkite. REAL refers to the real traces in the two datasets. It can be observed that W^3-$tess$ and PULE outperform all the heuristic algorithms, with synthetic traces exhibit similar visiting proportion to the real traces. That is because PAD, INTER, and PAS fail to model users' mobility. In addition, W^3-$tess$ performs better than PULE, as PULE ignores users' social behavior and thus cannot credibly imitate users' social locations.

2) Map Inference: The KL-divergence of real traces to the synthetic traces is shown in Table I. KL in PAD, INTER, and PAS is larger than that in PULE and W^3-$tess$. This indicates the distribution of visiting proportion among locations are preserved in PULE and W^3-$tess$. Furthermore, the KL in W^3-$tess$ is a bit less than that in PULE, as PULE fails to model users' social locations.

3) Modeling Users' Mobility: The mobility similarity is shown in Table I. The synthetic traces in PULE and W^3-$tess$ exhibit more similar mobility behavior, because the two algorithms credibly model users' temporal and spatial behavior. In addition, W^3-$tess$ outperforms PULE, as PULE ignores the social behavior.

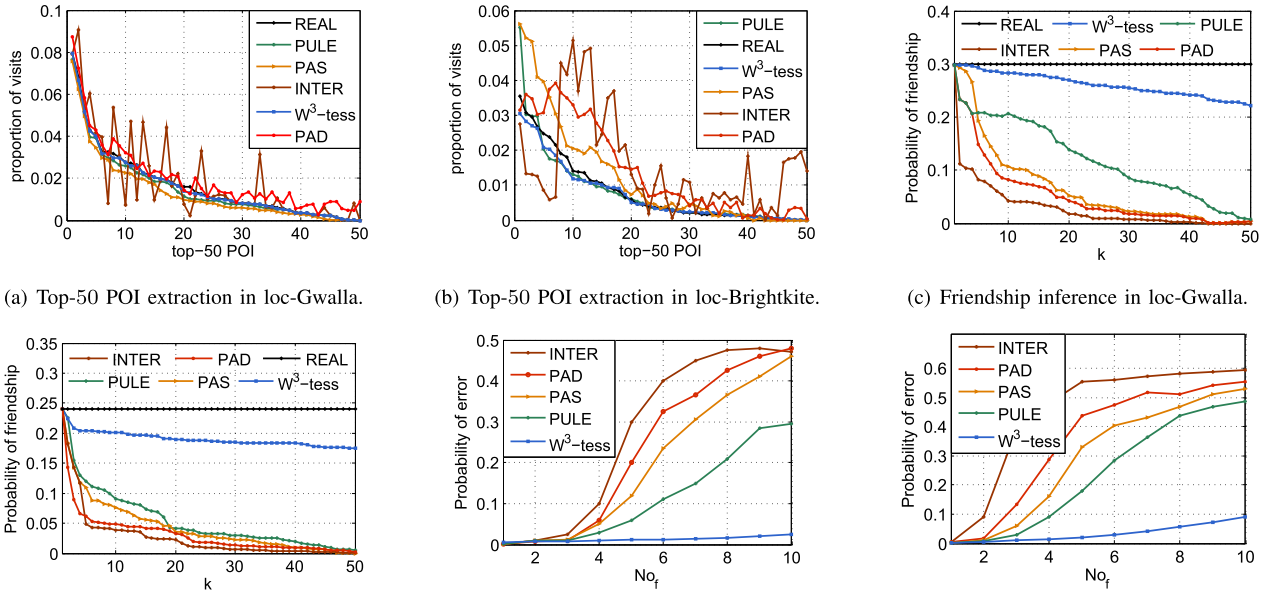
4) Friendship Inference: It shows in Figs. 7(c) and 7(d) that the probability of friendship decreases with the number of synthetic traces in the four algorithms except for REAL. Because REAL does not cloak real traces with synthetic ones while the other four algorithms do, and more synthetic traces result in more fake users. Furthermore, W^3-$tess$ significantly outperforms the other four algorithms, as the fake users in W^3-$tess$ exhibit similar social behavior to users. In addition, the probability of friendship in loc-Gwalla is larger than that in loc-Brightkite, since more friendships exist in loc-Gwalla.

5) Influential Friends Inference: As shown in Figs. 7(e) and 7(f), the probability of error increases with the number of influential friends. That is not surprising, as inferring more influential friends definitely leads to more inference errors. In addition, the probability of error in W^3-$tess$ is much less than that in other four algorithms, as W^3-$tess$ syntheses traces considering both temporal, spatial, and social behavior. Lastly, the probability of error in loc-Gwalla is less than that in loc-Brightkite, since more friendships exist in loc-Gwalla.

In summary, W^3-$tess$ outperforms all existing work, as it can guarantee trace data utility in all the above geo-data analysis tasks.

D. Computation Cost

The computation cost in PAD, INTER, PAS, PULE, and W^3-$tess$ in the two datasets loc-Gwalla and loc-Brightkite is shown in Fig. 8. It can be observed that the computation cost in the two datasets loc-Gwalla and loc-Brightkite increases with the number of synthetic traces corresponding to each user's trace k, since algorithms in PAD, INTER, PAS, PULE, and W^3-$tess$ have to generate more traces when parameter k is enlarged, thus incurring more computation cost. Moreover, in both Figs. 8(a) and 8(b), computation cost in W^3-$tess$ is less than that in PULE and PAS, and a bit larger than that in PAD and INTER. It is because that PAD and INTER only considered the spatial and temporal characteristics of locations when generate fake traces, and that in contrast W^3-$tess$ consider temporal, spatial, and social behavior of each user's mobility. Moreover, PAS iteratively generated trajectory with pauses, and PULE processed all users' locations both in geographic and semantic spaces. In contrast, W^3-$tess$ relies on light-weight algorithm to process the friends' locations of each user rather than the whole users' locations. As a result, W^3-$tess$ incurs less computation cost. Lastly, computation



(a) Top-50 POI extraction in loc-Gwalla. (b) Top-50 POI extraction in loc-Brightkite. (c) Friendship inference in loc-Gwalla.

(d) Friendship inference in loc-Brightkite. (e) Influential friends inference in loc-Gwalla. (f) Influential friends inference in loc-Brightkite.

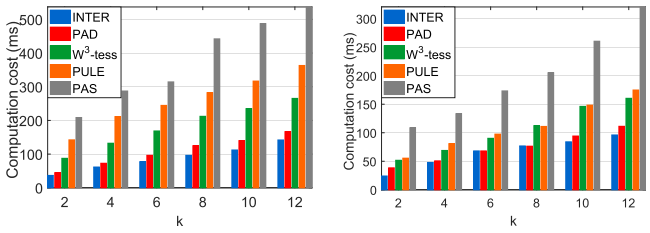
Fig. 7. Trace data utility in both loc-Gwalla and loc-Brightkite. (a) (b) Top-50 POI extraction; (c) (d) Friendship inference varies with the number of synthetic traces k; (e) (f) Influential friends inference.

TABLE I

KL AND sim BETWEEN SYNTHETIC AND REAL TRACES

Datasets	Metrics	PAD	INTER	PAS	PULE	W^3-$tess$
loc-G	KL	1.22	3.50	1.03	0.436	0.42
loc-B	KL	1.34	3.70	1.40	0.42	0.401
loc-G	sim	0.02	0.005	0.11	0.745	0.879
loc-B	sim	0.04	0.003	0.147	0.747	0.84

loc-G and loc-B refer to datasets loc-Gwalla, loc-Brightkite.



(a) Computation cost in loc-Gwalla. (b) Computation cost in loc-Brightkite.

Fig. 8. Computation cost in both datasets loc-Gwalla and loc-Brightkite varies with the number of synthetic traces corresponding to each users trace k.

cost in dataset loc-Gwalla is larger than that in dataset loc-Brightkite, as the more users' locations are included in dataset loc-Gwalla. In summary, W^3-$tess$ is preferable in terms of computation cost, since it provides more location privacy preservation and trace data utility guarantee.

VI. RELATED WORK

A. Literature Concerning Synthesizing Traces

To protect users' location privacy against the untrusted data aggregator, a prevalent method to protect users' location privacy is to cloak real traces with synthetic traces. Naive heuristic algorithms were proposed in [7], [8] that generated dummy users' data by applying interpolation strategies, or according to circles, grids, etc. These work cannot model users' mobility behavior and thus was vulnerable to mobility model based inference attacks. To this end, the follow-up work [11] generated dummy users stopping at several locations to visit attractions. The same problem is tackled by [10] that first characterizes users' driving behavior from their traces and then on this basis generate dummy traces using probabilistic models. Further, authors in [9] made efforts towards generating dummy traces with consistent movement patterns, taking three kind of privacy disclosure. *However, all these work [7]–[11] are susceptible to the inference attacks [12] where adversaries infer users' location privacy according to the semantic features of locations.* To this end, research [5] considered the semantic of users' data by characterizing users' temporal and spatial behavior when cloaking users' data. Furthermore, the latest research [38] used interest points and the road network topology to make the generated dummy more realistic and effective in time and space dimension. Moreover, study [39] proposed an estimation-based dummy trajectory generation without any assumptions about users' movements. Another work [40] utilized dummy based technique to generate dummy locations in spatial crowdsourcing. *Unfortunately, it is vulnerable to the social relationship based de-anonymization attacks [6], [15].*

To tackle the above problem, in this paper, we propose W^3-$tess$, the first work that synthesizes privacy preserving traces via enhancing the plausibility of synthetic traces with social networks.

不足+我是怎么做的，你的论文中只要在B中另外写上你是怎么的，A只写不足就可。

B. Other Somehow Related Work

Other studies concerning privacy preservation are somehow related to our work, which mainly includes the studies based on differential privacy and work based on encryption.

1) Studies Based on Differential Privacy: Several papers [41]–[45] concerning applying differential privacy to location privacy are somehow related to our work. Specifically, existing work [41] introduced a generalized version of differential privacy, geo-indistinguishability. The follow-up study [42] proposed spatial or temporal distance function, and on this basis, combined differential privacy and δ-neighbourhood. Similarly, the work [43] proposed δ-location set based differential privacy and taked account for the temporal correlations in location data. Thereafter, the study [44] inferred users' social relationships from their locations in a differentially private manner. Moreover, research [45] considered the temporal correlations among continuous data and quantified the corresponding privacy disclosure of differential privacy. The latest work [46] proposed a privacy-preserving method that meets differential privacy constraint to protect location data privacy and maximize data utility. Moreover, study [47] generated synthetic integrated dataset so that the publication of high-dimensional data meets differential privacy. Another work [48] utilized the geo-indistinguishability based on differential privacy to preserve the sensitive location information. In contrast, literature [49] designed a differentially private GAN that meets differential privacy under GANs.

However, all these work only considered the temporal and spatial behavior implied in users' locations, and thus suffered from the social relationship based de-anonymization attacks. Different to these studies above, W^3-$tess$ takes temporal, spatial, and social behavior into consideration, and it can defend against such social relationship based de-anonymization attacks.

2) Studies Based on Encryption: Work [50] proposed a hybrid homomorphic encryption via combining public-key encryption and homomorphic encryption. Study [51] designed a distance-based encryption to apply biometrics in identity-based encryption. Thereafter, work [52] used a key-aggregate approach and a proxy re-encryption scheme to design a key-aggregate proxy re-encryption scheme. Another work [53] focused on BSS-based speech encryption to encrypt speech signal via linearly combining it with secret key signal. Literature [54] designed a novel Dynamic Searchable Symmetric Encryption to provide higher level of privacy. Moreover, work [55] protected query range and individual IoT device's data using BGN homomorphic encryption technique. Studies [56], [57] utilized a cryptographic technique and differential privacy to allow the collector to privately compute the statistics and support the dynamic dropouts. *Nevertheless, these encryption-based approaches incurred a large amount of computation and communication overheads.* In contrast, we propose a lightweight yet effective scheme that synthesizes privacy preserving traces via enhancing the plausibility of synthetic traces with social networks.

VII. Conclusion

In this paper, we present W^3-$tess$, the first work that synthesizes privacy preserving traces through enhancing the plausibility of synthetic traces with social networks to protect users' location privacy against social relationship based de-anonymization attacks. In addition, both location privacy preservation and trace data utility guarantee are theoretically provable. Moreover, it is applicable to most geo-data analysis tasks. Extensive experiments on two real world datasets have demonstrated the effectiveness of W^3-$tess$.

Appendix A
Proof of Theorem 3

Proof: According to the composition property (cf. Theorem 1), we only need to prove W^3-$tess$ satisfy (ε, δ)-differential privacy in each dimension (i.e., temporal, spatial and social behavior). We first prove W^3-$tess$ meets (ε, δ)-differential privacy in terms of temporal behavior.

Denote p the procedure of selecting trace fragments using $(\mathcal{F}, k, \rho)$ sampling in W^3-$tess$; $\mathcal{F}$ the trace data analysis task; $p(t)$ the three-dimension attributes; D the set of users' traces and the synthetic traces; D_{-t} is the dataset where the tuple (i.e., trace) t is removed from D. Assume n tuples t' in D meet $p(t').temporal = p(t).temporal$. Denote i the number of $p(t).temporal$ in S. Denote $F(i, n, \rho) = \sum_{j=0}^{i} f(i, n, \rho)$, where $f(i, n, \rho)$ is the probability of getting i heads in n trials where each trial succeeds with probability ρ. According to Section III-B, $(\mathcal{F}, k, \rho)$ sampling in W^3-$tess$ is like the game, tossing a coin. According to Theorem 2, we get the following results.

First, we have

$$\frac{Pr[\mathcal{F}(D)=S]}{Pr[\mathcal{F}(D_{-t})=S]} = \frac{F(i,n,\rho)}{F(i,n-1,\rho)} = \begin{cases} \frac{n(1-\rho)}{n-i}, & n \geq i \\ 1, & n < i. \end{cases} \tag{14}$$

Obviously, $e^{-\varepsilon} \leq \frac{Pr[\mathcal{F}(D)=S]}{Pr[\mathcal{F}(D_{-t})=S]} \leq e^{\varepsilon}$ holds for $i > n$. When $n \geq i$, we get $\frac{n(1-\rho)}{n-i} > 1-\rho$. Since $\varepsilon \geq -ln(1-\rho)$, $1-\rho \geq e^{-\varepsilon}$ holds. Next, we only need to consider the following case where $\frac{n(1-\rho)}{n-i} > e^{\varepsilon}$ so that Eq. (1) does not hold.

$$\frac{n(1-\rho)}{n-i} > e^{\varepsilon}, \quad s.t \begin{cases} n \geq i > \dfrac{(e^{\varepsilon}-1+\rho)}{e^{\varepsilon}} = n\varrho, \\ i \geq k. \end{cases} \tag{15}$$

Then the possibility that $\mathcal{F}(D)$ and $\mathcal{F}(D_{-t})$ give bad outcomes (i.e., $\frac{Pr[\mathcal{F}(D)=S]}{Pr[\mathcal{F}(D_{-t})=S]} \leq e^{\varepsilon}$ does not hold) is

$$\max_{n} \sum_{i:i \geq k \wedge i > \varrho n}^{n} F(i,n,\rho), \tag{16}$$

$$\max_{n} \sum_{i:i \geq k \wedge i > \varrho n}^{n-1} F(i,n-1,\rho). \tag{17}$$

Since both equations above increase with n, Eq. (16) is thus larger than Eq. (18). Furthermore, δ is bound by

$$\delta = \max_{n} \sum_{i:i \geq k \wedge i > \varrho n}^{n} F(i,n,\rho)$$

$$\frac{Pr[\mathcal{F}(D)=S]}{Pr[\mathcal{F}(D_{-\mathcal{T}})=S]}=\frac{C_n^i\rho^i(1-\rho)^{(n-i)}}{C_{n-(k-1)}^i\rho^i(1-\rho)^{(n-(k-1)-i)}}=\begin{cases}\dfrac{n(n-1)\cdots(n-i+1)(1-\rho)^{k-1}}{[n-(k-1)][n-(k-1)-1]\cdots[n-(k-1)-(i-1)]}, & n\geq i;\\ 1, & n<i;\end{cases} \tag{21}$$

$$=\begin{cases}\max\limits_{n:n<\lceil\frac{k}{\varrho}-1\rceil}\sum\limits_{i:i\geq k}^{n}F(i,n,\rho),\\ \max\limits_{n:n\geq\lceil\frac{k}{\varrho}-1\rceil}\sum\limits_{i:i>\varrho n}^{n}F(i,n,\rho).\end{cases} \tag{18}$$

Lastly, we get

$$\delta=\max_{n:n\geq\lceil\frac{k}{\varrho}-1\rceil}\sum_{i:i>\varrho n}^{n}F(i,n,\rho). \tag{19}$$

In summary, W^3-$tess$ meets (ε,δ)-differential privacy in terms of temporal behavior.

Similarly, we can prove W^3-$tess$ meets (ε,δ)-differential privacy in terms of spatial and social behavior.

Overall, Theorem 3 holds. ∎

APPENDIX B
PROOF OF THEOREM 4

Proof: We only need to prove the following holds with the possibility $(1-\delta)$,

$$e^{-\varepsilon}\leq\frac{Pr[\mathcal{F}(D)=S]}{Pr[\mathcal{F}(D_{-\mathcal{T}})=S]}\leq e^{\varepsilon}, \tag{20}$$

where $\mathcal{T}$ is the set of any k tuples in D.

As any tuple $p(t)$ appears no less than k times, then $i\geq k$, and we get Eq. (21), shown at the top of this page.

Eq. (20) does not hold for some values of i. So next we try to bound the possibility δ that Eq. (20) is violated. It can be deduced from Eq. (21) that Eq. (20) hold when $i>n$.

We now consider the case when $n\geq j\geq k$. Note that $\phi(n,k,i)=\frac{n(n-1)\cdots(n-i+1)}{[n-(k-1)][n-(k-1)-1]\cdots[n-(k-1)-(i-1)]}>1$, so $\phi(n,k,i)\ (1-\rho)^{k-1}>(1-\rho)^{k-1}\geq e^{-\varepsilon}$. Hence next we only need to consider the values of i that make $\phi(n,k,i)(1-\rho)^{k-1}>e^{\varepsilon}$. The formula in Eq. (21) can be simplified into

$$f(i)=\frac{n(n-1)\cdots[n-(k-1)+1](1-\rho)^{k-1}}{(n-i)(n-i-1)\cdots[n-(i-1)-(k-1)]}. \tag{22}$$

Denote $g(i)=(n-i)(n-i-1)\cdots[n-i+1-(k-1)]$. Obviously, $g(i)$ decreases with increasing i. Thus Eq. (22) increases with i. So we can get $i=\varrho n$ through dichotomy so that $\varrho=\frac{1}{n}\arg\min_i[f(i)\geq e^{\varepsilon}]$. Thus, when $i>\varrho n$, Eq. (20) is violated.

So far, we have proved

$$\frac{Pr[\mathcal{F}(D)=S]}{Pr[\mathcal{F}(D_{-\mathcal{T}})=S]}>e^{\varepsilon},\quad s.t\begin{cases}n\geq i\geq k,\\ i>\varrho n.\end{cases} \tag{23}$$

The possibilities that $\mathcal{F}(D)$ and $\mathcal{F}(D_{-\mathcal{T}})$ output bad outcomes are

$$\max_n\sum_{i:i\geq k\wedge i>\varrho n}^{n}F(i,n,\rho), \tag{24}$$

$$\max_n\sum_{i:i\geq k\wedge i>\varrho n}^{n-1}F(i,n-1,\rho). \tag{25}$$

Since $\sum_i^n F(i,n-1,\rho)$ increases with n, the possibility in Eq. (24) is larger than that in Eq. (25). Thus, we only need to bound the possibility in Eq. (24). So the error possibility δ is bound by

$$\delta=\max_n\sum_{i:i\geq k\wedge i>\varrho n}^{n}F(i,n,\rho)=\begin{cases}\max\limits_{n:n<n_o}\sum\limits_{i:i\geq k}^{n}F(i,n,\rho),\\ \max\limits_{n:n\geq n_o}\sum\limits_{i:i>\varrho n}^{n}F(i,n,\rho),\end{cases} \tag{26}$$

where $n_o=\lceil\frac{k}{\varrho}-1\rceil$ which satisfies $n_o\varrho<k$ and $(n_o+1)\varrho\geq k$. Lastly, we get the error possibility δ

$$\delta=\max_{n:n\geq n_o}\sum_{i:i>\varrho n}^{n}F(i,n,\rho). \tag{27}$$

In summary, Theorem 4 holds. ∎

REFERENCES

[1] X. Wu *et al.*, "GLP: A novel framework for group-level location promotion in Geo-social networks," *IEEE/ACM Trans. Netw.*, vol. 26, no. 6, pp. 2870–2883, Dec. 2018.

[2] P. Zhao *et al.*, "P^3-LOC: A privacy-preserving paradigm-driven framework for indoor localization," *IEEE/ACM Trans. Netw.*, vol. 26, no. 6, pp. 2856–2869, Dec. 2018.

[3] M. Bradbury and A. Jhumka, "Understanding source location privacy protocols in sensor networks via perturbation of time series," in *Proc. IEEE Conf. Comput. Commun.*, May 2017, pp. 1–9.

[4] C. Luo *et al.*, "Predictable privacy-preserving mobile crowd sensing: A tale of two roles," *IEEE/ACM Trans. Netw.*, vol. 27, no. 1, pp. 361–374, Feb. 2019.

[5] V. Bindschaedler and R. Shokri, "Synthesizing plausible privacy-preserving location traces," in *Proc. IEEE Symp. Secur. Privacy*, May 2016, pp. 546–563.

[6] M. Srivatsa and M. Hicks, "Deanonymizing mobility traces: Using social network as a side-channel," in *Proc. ACM Conf. Comput. Commun. Secur.*, Oct. 2012, pp. 628–637.

[7] H. Lu, C. S. Jensen, and M. L. Yiu, "Pad: Privacy-area aware, dummy-based location privacy in mobile services," in *Proc. 7th ACM Int. Workshop Data Eng. Wireless Mobile Access*, Jun. 2008, pp. 16–23.

[8] A. Gkoulalas-Divanis and V. S. Verykios, "A privacy-aware trajectory tracking query engine," *ACM SIGKDD Explor. Newslett.*, vol. 10, no. 1, pp. 40–49, 2008.

[9] T.-H. You, W.-C. Peng, and W.-C. Lee, "Protecting moving trajectories with dummies," in *Proc. Int. Conf. Mobile Data Manage.*, May 2007, pp. 278–282.

[10] J. Krumm, "Realistic driving trips for location privacy," in *Proc. Int. Conf. Pervasive Comput.*, 2009, pp. 25–41.

[11] R. Kato *et al.*, "A dummy-based anonymization method based on user trajectory with pauses," in *Proc. 20th Int. Conf. Adv. Geographic Inf. Syst.*, Nov. 2012, pp. 249–258.

[12] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux, "Quantifying location privacy," in *Proc. IEEE Symp. Secur. Privacy*, May 2011, pp. 247–262.

[13] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: User movement in location-based social networks," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2011, pp. 1082–1090.

[14] D. Mok, B. Wellman, and J. Carrasco, "Does distance matter in the age of the internet?" *Urban Stud.*, vol. 47, no. 13, pp. 2747–2783, 2010.

[15] J. Qian, X.-Y. Li, C. Zhang, and L. Chen, "De-anonymizing social networks and inferring private attributes using knowledge graphs," in *Proc. 35th Annual IEEE Int. Conf. Comput. Commun.*, Apr. 2016, pp. 1–9.

[16] J. Zeng *et al.*, "Mobile r-gather: Distributed and geographic clustering for location anonymity," in *Proc. 18th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2017, p. 7.

[17] H. Jiang, P. Zhao, and C. Wang, "RobLoP: Towards robust privacy preserving against location dependent attacks in continuous LBS queries," *IEEE/ACM Trans. Netw.*, vol. 26, no. 2, pp. 1018–1032, Apr. 2018.

[18] P. Zhao *et al.*, "ILLIA: Enablingκ-anonymity-based privacy preserving against location injection attacks in continuous LBS queries," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1033–1042, Apr. 2018.

[19] M. Backes, M. Humbert, J. Pang, and Y. Zhang, "Walk2friends: Inferring social links from mobility profiles," 2017, *arXiv:1708.08221*. [Online]. Available: https://arxiv.org/abs/1708.08221

[20] C. Y. T. Ma, D. K. Y. Yau, N. K. Yip, and N. S. V. Rao, "Privacy vulnerability of published anonymous mobility traces," *IEEE/ACM Trans. Netw.*, vol. 21, no. 3, pp. 720–733, Jun. 2013.

[21] C. Dwork, "Differential privacy," in *Proc. 33rd Int. Colloq. Automata, Lang. Program.*, 2006, pp. 1–12.

[22] J. Xu, Z. Zhang, X. Xiao, Y. Yang, and G. Yu, "Differentially private histogram publication," in *Proc. IEEE 28th Int. Conf. Data Eng.*, 2012, pp. 32–43.

[23] M. Hay, V. Rastogi, G. Miklau, and D. Suciu, "Boosting the accuracy of differentially private histograms through consistency," *VLDB Endowment*, vol. 3, nos. 1–2, pp. 1021–1032, 2010.

[24] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, nos. 3–4, pp. 211–407, 2014.

[25] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Proc. EUROCRYPT*, 2006, pp. 486–503.

[26] C. Dwork, "Differential privacy: A survey of results," in *Proc. Int. Conf. Theory Appl. Models Comput.*, 2008, pp. 1–19.

[27] D. Falcone, C. Mascolo, C. Comito, D. Talia, and J. Crowcroft, "What is this place? Inferring place categories through user patterns identification in geo-tagged tweets," in *Proc. 6th Int. Conf. Mobile Comput., Appl. Services*, Nov. 2014, pp. 10–19.

[28] N. Lathia, K. K. Rachuri, C. Mascolo, and P. J. Rentfrow, "Contextual dissonance: Design bias in sensor-based experience sampling methods," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Sep. 2013, pp. 183–192.

[29] A. Thomason, N. Griffiths, and V. Sanchez, "Context trees: Augmenting geospatial trajectories with context," *ACM Trans. Inf. Syst.*, vol. 35, no. 2, p. 14, 2016.

[30] A. Sadilek, H. Kautz, and J. P. Bigham, "Finding your friends and following them to where you are," in *Proc. 5th ACM Int. Conf. Web Search Data Mining*, 2012, pp. 723–732.

[31] Y. Jia, Y. Wang, X. Jin, and X. Cheng, "Location prediction: A temporal-spatial Bayesian model," *ACM Trans. Intell. Syst. Technol.*, vol. 7, no. 3, p. 31, 2016.

[32] C. Cheng, H. Yang, I. King, and M. R. Lyu, "Fused matrix factorization with geographical and social influence in location-based social networks," in *Proc. 26th AAAI Conf. Artif. Intell.*, Jul. 2012, pp. 17–23.

[33] C. C. Aggarwal, "On κ-anonymity and the curse of dimensionality," in *Proc. 31st Int. Conf. Very Large Data Bases*, Aug. 2005, pp. 901–909.

[34] *Gwalla and Brightkite Data*. Accessed: 2011. [Online]. Available: http://snap.stanford.edu/data/index.html

[35] P. Galdames and Y. Cai, "Efficient processing of location-cloaked queries," in *Proc. IEEE INFOCOM*, May 2012, pp. 2480–2488.

[36] D. J. Crandall *et al.*, "Inferring social ties from geographic coincidences," *Proc. Nat. Acad. Sci. USA*, vol. 107, no. 52, pp. 22436–22441, 2010.

[37] D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A.-L. Barabási, "Human mobility, social ties, and link prediction," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2011, pp. 1100–1108.

[38] Y. Li and S. Li, "A real-time location privacy protection method based on space transformation," in *Proc. 14th Int. Conf. Comput. Intell. Secur.*, Nov. 2018, pp. 291–295.

[39] S. Hayashida, D. Amagata, T. Hara, and X. Xie, "Dummy generation based on user-movement estimation for location privacy protection," *IEEE Access*, vol. 6, pp. 22958–22969, 2018.

[40] R. Alharthi, E. Aloufi, A. Alqazzaz, I. Alrashdi, and M. Zohdy, "DCentroid: Location privacy-preserving scheme in spatial crowdsourcing," in *Proc. IEEE 9th Annu. Comput. Commun. Workshop Conf.*, Jan. 2019, pp. 0715–0720.

[41] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: Differential privacy for location-based systems," in *Proc. 20th ACM Conf. Comput. Commun. Secur.*, 2013, pp. 901–914.

[42] C. Fang and E.-C. Chang, "Differential privacy with δ-neighbourhood for spatial and dynamic datasets," in *Proc. 9th ACM Symp. Inf., Comput. Commun. Secur.*, 2014, pp. 159–170.

[43] Y. Xiao and L. Xiong, "Protecting locations with differential privacy under temporal correlations," in *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur.*, 2015, pp. 1298–1309.

[44] C. Shahabi, L. Fan, L. Nocera, L. Xiong, and M. Li, "Privacy-preserving inference of social relationships from location data: A vision paper," in *Proc. 23rd SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2015, pp. 901–904.

[45] Y. Cao, M. Yoshikawa, Y. Xiao, and L. Xiong, "Quantifying differential privacy under temporal correlations," in *Proc. IEEE 33rd Int. Conf. Data Eng.*, Apr. 2017, pp. 821–832.

[46] C. Yin, J. Xi, R. Sun, and J. Wang, "Location privacy protection based on differential privacy strategy for big data in industrial Internet of things," *IEEE Trans. Ind. Informat.*, vol. 14, no. 8, pp. 3628–3636, Aug. 2018.

[47] X. Cheng *et al.*, "Multi-party high-dimensional data publishing under differential privacy," *IEEE Trans. Knowl. Data Eng.*, to be published. doi: 10.1109/TKDE.2019.2906610.

[48] D. Shi *et al.*, "Deep Q -network-based route scheduling for TNC vehicles with passengers' location differential privacy," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7681–7692, Oct. 2019. doi: 10.1109/JIOT.2019.2902815.

[49] C. Xu *et al.*, "GANobfuscator: Mitigating information leakage under GAN via differential privacy," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 9, pp. 2358–2371, Sep. 2019. doi: 10.1109/TIFS.2019.2897874.

[50] J. H. Cheon and J. Kim, "A hybrid scheme of public-key encryption and somewhat homomorphic encryption," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 5, pp. 1052–1063, May 2015.

[51] F. Guo, W. Susilo, and Y. Mu, "Distance-based encryption: How to embed fuzziness in biometric-based encryption," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 2, pp. 247–257, Feb. 2016.

[52] W.-H. Chen, C.-I. Fan, and Y.-F. Tseng, "Efficient key-aggregate proxy re-encryption for secure data sharing in clouds," in *Proc. IEEE Conf. Dependable Secure Comput.*, Dec. 2018, pp. 1–4.

[53] A. Farhati, A. B. Aicha, and R. Bouallegue, "On the strengthening of the speech encryption schemes for communication systems based on blind source separation approach," in *Proc. 14th Int. Wireless Commun. Mobile Comput. Conf.*, Jun. 2018, pp. 108–111.

[54] T. Hoang, A. A. Yavuz, and J. G. Merchan, "A secure searchable encryption framework for privacy-critical cloud storage services," *IEEE Trans. Services Comput.*, to be published. doi: 10.1109/TSC.2019.2897096.

[55] R. Lu, "A new communication-efficient privacy-preserving range query scheme in fog-enhanced IoT," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2497–2505, Apr. 2019. doi: 10.1109/JIOT.2018.2871204.

[56] Y. Chen, J.-F. Martínez-Ortega, P. Castillejo, and L. López, "A homomorphic-based multiple data aggregation scheme for smart grid," *IEEE Sensors J.*, vol. 19, no. 10, pp. 3921–3929, May 2019. doi: 10.1109/JSEN.2019.2895769.

[57] A. Abdallah and X. S. Shen, "A lightweight lattice-based homomorphic privacy-preserving data aggregation scheme for smart grid," *IEEE Trans. Smart Grid*, vol. 9, no. 1, pp. 396–405, Apr. 2016.

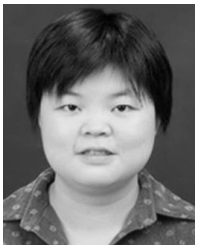
Ping Zhao received the Ph.D. degree from the Huazhong University of Science and Technology, China. She is currently a Visiting Assistant Professor with Hunan University and an Assistant Professor with Donghua University. Her research interests include privacy preservation, mobile computation, and the Internet of Things.

Hongbo Jiang (SM'15) received the Ph.D. degree from Case Western Reserve University in 2008. He was a Professor with the Huazhong University of Science and Technology. He is currently a Full Professor with the College of Computer Science and Electronic Engineering, Hunan University. His research concerns computer networking, especially algorithms and protocols for wireless and mobile networks. He is also serving as an Editor for the IEEE/ACM TRANSACTIONS ON NETWORKING, an Associate Editor for the IEEE TRANSACTIONS ON MOBILE COMPUTING, and an Associate Technical Editor for the *IEEE Communications Magazine*.

Jie Li received the B.S. degree from the College of Computer Science and Electronic Engineering, Hunan University, China, in 2016, where she is currently pursuing the Ph.D degree. Her research interests include mobile and wireless networks, especially privacy preserving in mobile systems.

Fanzi Zeng received the Ph.D. degree in signal and information processing from Beijing Jiaotong University, Beijing, China, in 2005. Since 2005, he has been with the School of information science and engineering, Hunan University, Changsha, China, where he is currently a Professor. His general interests are in the areas of signal processing for wireless communications, estimation, and detection theory. His current research focuses on cognitive radio technology.

Zhu Xiao (M'15–SM'19) received the M.S. and Ph.D. degrees in communication and information system from Xidian University, China, in 2007 and 2009, respectively. From 2010 to 2012, he was a Research Fellow with the Department of Computer Science and Technology, University of Bedfordshire, U.K. He is currently an Associate Professor with the College of Computer Science and Electronic Engineering, Hunan University, China. His research interests include mobile communications, wireless localization, the Internet of Vehicles, and mobile computing.



Kun Xie received the Ph.D. degree in computer application from Hunan University, Changsha, China, in 2007. She was a Post-Doctoral Fellow with the Department of Computing, The Hong Kong Polytechnic University, from 2007 to 2010. She was a Visiting Researcher with the Department of Electrical and Computer Engineering, The State University of New York at Stony Brook, from 2012 to 2013. She is currently a Professor with Hunan University. She has authored over 60 articles in major journals and conference proceedings, including the journals IEEE/ACM TRANSACTIONS ON NETWORKING, the IEEE TRANSACTIONS ON MOBILE COMPUTING, the IEEE TRANSACTIONS ON COMPUTERS, and the IEEE TRANSACTIONS ON WIRELESS COMPUTING and conferences INFOCOM, ICDCS, SECON, and IWQoS. Her research interests include wireless networks and mobile computing, network management and control, cloud computing and mobile cloud, and big data.

Guanglin Zhang received the Ph.D. degree in information and communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2012. From 2013 to 2014, he was a Post-Doctoral Research Associate with the Institute of Network Coding, The Chinese University of Hong Kong. From 2013 to 2017, he was an Associate Professor with the Department of Communication Engineering, Donghua University, Shanghai, where he is currently a Professor and the Department Chair with the Department of Communication Engineering. His research interests include capacity scaling of wireless networks, vehicular networks, smart microgrids, and mobile edge computing. He served as a Technical Program Committee Member for the IEEE Global Communications Conference in 2016 and 2017, the IEEE International Conference on Communications in 2014, 2015, and 2017, the IEEE Vehicular Technology Conference in Fall 2017, the IEEE/CIC International Conference on Communications, China, in 2014, the International Conference on Wireless Communications and Signal Processing in 2014, the Asia–Pacific Conference on Communications in 2013, and the International Conference on Wireless Algorithms, Systems, and Applications in 2012. He serves as the Local Arrangement Chair of ACM TURC 2017 and the Vice TPC Co-Chair of ACM TURC 2018. He serves as an Editor on the Editorial Board of *China Communications* and the *Journal of Communications and Information Networks*. He is an Associate Editor of IEEE ACCESS.