Where Were You Yesterday: Privacy Risk of Published Anonymous Trajectories

Shan Chang, Xiaoqiang Liu
and Ting Lu
School of Computer Science
and Technology
Donghua University
Shanghai, China
Email: changshan@dhu.edu.cn

Hongzi Zhu
Department of Computer Science
and Engineering
Shanghai Jiao Tong University
Shanghai, China
Email: hongzi@cs.sjtu.edu.cn

Mianxiong Dong and Kaoru Ota
Department of Information
and Electronic Engineering
Muroran Institute of Technology
Hokkaido, Japan
Email: {mx.dong, ota}@csse.muroran-it.ac.jp

Abstract—**With more and more trajectory traces available, conducting analysis and mining on those trajectories can obtain valuable information. Although the published traces are often made anonymous by substituting the true identities of mobile nodes with random identifiers, the privacy concern remains. In this paper, we propose a new de-anonymization attack based on the movement pattern of moving objects. Since moving objects are open to observe in public spaces, an attacker can easily learn information about a victim's movement either through direct observations or from third parties. After collecting a few trajectory segments of a mobile object, the movement pattern of the victim can be extracted, using an improved TF-IDF method. By comparing the movement pattern of the victim with those extracted from historical anonymous traces, it is possible to identify the victim from the anonymous traces. We conduct extensive trace-driven simulations and the results demonstrate that the attacker is able to de-anonymize anonymous trajectories with high probability.**

I. Introduction

With the blooming development of embedded GPS devices and localization technology, it is convenient to obtain the spatial information of mobile objects. A trajectory of a mobile object consists of a series of consecutive locations reports issued from that object. Such trajectories are very valuable to many applications. For example, the analysis on GPS data of vehicles can facilitate the transportation administration and perceive the traffic condition. Another example is that the government can utilize the mobility information to help improve the urban planning. In recent years, numerous trajectory data have been collected and published for study, such as GeoLife [1], Movebank [2], RAWDAD [3], crowdflow [4] and etc. However, both location and time are sensitive information. An attacker can use these published trajectory dat to infer personal and confidential information like habits, social customs, religious and sexual preferences of individuals, consequently raising serious privacy concerns.

In order to protect the privacy of moving objects, before publishing trajectory data, appropriate privacy protection techniques should be adopted. Currently, the existing methods can be divided into two categories: distortion-based and pseudonym-based. In distortion-based methods, the precision on both temporal and spatial domains of the original trajectories is reduced using cloaking techniques, such as reducing the resolution of the collected trajectories or introducing noise deliberately into trajectories. These methods may induce severe data distortion, which causes low usability of published trajectories. In pseudonym-based methods, the true identity of an object is replaced by a consistent, unique, and random identifier, i.e., pseudonym, which cannot be directly used to get the true identity of that object. Having the advantages of easy-to-implement, low computation cost, and no modification on the original trace, pseudonym-base methods are widely applied, such as GeoLife GPS trajectories by Microsoft [1].

Although the unlinkability between pseudonyms and true identities can be guaranteed, the validity of pseudonym on trajectory anonymization is questioned. Many researchers claim that the de-identification trajectory data are still prone to linkage attacks since spatial and temporal attributes are very powerful quasi-identifiers [5]. C. Y. T. Ma *et al.* indicate that if an attacker can gather some snapshots on a moving object's trajectory, referred to *side information*, then he or she may able to identify the victim's trajectory from a set of anonymous trajectories with high probability[6]. The authors utilize a series of locations as quasi-identifiers. The main assumption of this work is that an attacker can obtain side information of a victim during the time period of the anonymous trajectory of this victim. It is often the case, however, that an attacker can only have historical anonymous trajectory data, which are not temporally overlapped with the available side information.

In this paper, we first conduct intensive analysis on two large-scale del-world trajectory traces collected from two metropolises in China, i. e., Shanghai and Shenzhen. We have one key insight that real-world mobile objects such as vehicles have consistent and distinct mobility patterns, which can be utilized as quasi-identifiers of vehicles. Based on this observation, we propose a more practical de-anonymization attack, where an attacker can learn the mobility pattern of a victim by collecting several pieces of trajectories or locations of this victim when needed. Meanwhile, the attacker can also extract mobility patterns of published individual anonymous trajectories. By comparing the similarity between the mobility

978-1-5090-1328-9/16/$31.00 ©2016 IEEE

pattern of the victim and those of historical anonymous trajectories, with high probability, the attacker can identify those anonymous trajectories belonging to the victim. We conduct extensive trace-driven simulations and the results demonstrate the efficacy of the proposed attack. We highlight the main contributions as follows:

1) We conduct intensive analysis on real-world vehicular trace data and find that the preference on road segments selection of an individual vehicle can be used for a quasi-identifier of this vehicle.
2) We present a comprehensive de-anonymization attack strategy, where current mobility of a victim obtained by an attack can be utilized to identify its historical anonymous trajectories, making pseudonyms no longer in force.
3) We conduct extensive experiments to demonstrate that the concern that an adversary could identify the traces of one or more victims in the published data with high probability does exist.

II. Related Work

In recent years, privacy of published data sets has gained much attention in the literature [7] [8] [9] [10]. Different kinds of privacy-persevering techniques were proposed to ensure location privacy of published trajectories. *k*-anonymity was proposed by L. Sweeney [11]. When *k*-anonymity is satisfied, each individual is indistinguishable from *k-1* other individuals. However, achieving optimal k-anonymity with minimal distortion has been proved to be NP-Hard [12]. Cloaking is a general scheme in which the granularity of location information is reduced [13], for example reducing the resolution of the collected trajectories or introducing noise deliberately in trajectories. Pseudonym-based methods replace the true identities by random pseudonyms [14], which have the advantages of easy-to-implement, low computation cost, and no modification on the original trace. However pseudonym is insufficient for privacy protection, since other attributes may able to be used to link identities of individuals. Such attributes are called quasi-identifiers.

There are not many researchers who focus on revealing and assessing the privacy vulnerability of published anonymous location or trajectory data set. H. Zang and J. Bolot examined a large-scale data set of cell phone call records across US and attempted to determine to what extent anonymized location data can reveal private user information [15]. The results show that releasing anonymized location data in its original format, i.e. at the sector level or cell level, poses serious privacy threats as a significant fraction of users can be re-identified from the anonymized data. J. Freudiger *et al.* provide an analysis of the erosion of privacy caused by the use of location-based services (LBS) [16]. They evaluate the success of LBSs in predicting the true identity of pseudonymous users and their points of interest based on small samples of mobility traces. The result explore the relation between the type and quantity of data collected by LBSs and their ability to de-anonymize and profile users. A. Basu *et al.* propose an empirical risk model for privacy based on *k*-anonymous data release [17]. They discuss the relation of risk to the cost of attacks on privacy as well as the utility of the data. Chris Y. T. Ma *et al.* indicate that an attacker who obtains a amount of the snapshot information termed as side information, can infer an extended view of the whereabouts of a victim node appearing in an anonymous trace [6]. Bayesian inference is used to correlate snapshots with anonymous trace. Martin *et al.* quantify the impact of background knowledge possessed by an adversary on privacy breach[18]. The authors represent the background knowledge of adversaries in a language, and provide an algorithm to determine the amount of disclosed private information in the worst case with respect to the amount of the background knowledge.

Our problem differs from the previous work by specific focusing on user location leakage caused by considering the long-term mobility patterns of anonymous traces of users. An adversary, possessing of several pieces of trajectories of a victim, can employs various well grounded strategies to infer the private information from historical trajectory data sets.

III. Preliminaries And Problem Formulation

In this section, we clarify some terms used in this paper and briefly describe the problem.

A. Notations

- *Trajectory Data Set*: A trajectory data set can be represented as $D_{mt} = \{\mathcal{V}, \mathcal{T}\}$, where $\mathcal{V} = \{v_1, v_2, ..., v_n\}$ represents the set of vehicles, $\mathcal{T} = \{TR_i, TR_2, ..., TR_n\}$ is the collection of trajectories of all vehicles, where TR_i denotes the trajectory associated with a vehicle $v_i \in \mathcal{V}$.
- *GPS Report*: A GPS report p is composed of a latitude, a longitude coordinate and a timestamp. $p_j^{(v_i)}$ represents the jth report of v_i, which can be simplified as p_j when there is no confusion.
- *Trajectroy*: A trajectory TR_i is represented by a sequence of GPS reports, i.e., $TR_i : p_1 \to p_2 \to \ldots \to p_{len_i} (1 \leq i \leq n)$, where len_i refers to the length of the sequence. A trajectory $p_j \to p_{j+1} \to \ldots \to p_{j+k} (1 \leq j \leq ... \leq j + k \leq len_i)$ is called a *sub-trajectory* of TR_i
- *Road Segment(RS)*: A road segment r is a one-directional edge that connects two intersections $(r.s, r.e)$.
- *Route:* A Route R is a set of consecutive road segment, $R : r_1 \to r_2 \ldots \to r_q$, where $r_k.e = r_{k+1}.s, 1 \leq k < q$.

B. Trajectory Anonymization

Before a trajectory data set is released to public, for each trajectory, the true identity of the corresponding vehicle should be replaced by a random and unique pseudonym. The selection of pseudonyms should follow two principles: 1) the true node identity can not be related in any way to the pseudonym; 2) the same identity is always mapped to the same pseudonym.

TABLE I
Statistical Characteristics of the Trajectory Sets

Data Set	Shanghai	Shenzhen
Number of vehicles	906	1945
From data	Feb. 1, 2007	Oct. 1, 2009
Duration(day)	28	31
Granularity(second)	$15, 60$	60
Number of Trajectories	25,368	60,295

C. Attack Model

We study the *de-anonymization* attacks in which an attacker tries to determine the true identity of an anonymous trajectory. A vehicle whose locations are exposed is called a *victim* object. The objective of the attacker is to determine as many identities of the anonymous trajectories as possible. More precisely, the attack can be divided into the following three steps:

1) The attacker accesses a published set of historical trajectories of its victims and other not-so-interested vehicles.
2) The attacker observes (or obtains) the current movements of victims, to get several pieces of fresh trajectories of the victims.
3) The attacker links each of its victims to all historical anonymous trajectories belonging to this victim.

IV. Trajectory Pre-processing

A. Dataset

- *Trajectories:* we utilize two sets of GPS reports, from more than 10 thousands of taxies, in Shanghai and Shenzhen, two metropolises in China. The specific information contained in a GPS report includes: the ID of a vehicle, the longitude and latitude coordinates of the vehicle, timestamp, moving speed, heading direction and the operational status (i.e., whether a taxi has passengers onboard or a bus is arriving at a bus stop). Reports are sent at a granularity of around one minute. The specific characteristics of the two data sets are shown in Table I.
- *Road Network:* the road networks of Shanghai and Shenzhen have 66,459 and 89,578 road segments, respectively. Each road segment has a unique identity in each road network.

B. Trajectory Segmentation and Map Matching

The whole trajectory TR_i of v_i is divided into several *1-day sub-trajectories* $TR_{i,j}(1 \leq j \leq num_{day})$, which denotes the movement of taxi v_i in the *j*-th day. Due to GPS errors, we adopt the ST-Matching map-matching algorithm [19] to map each GPS report of a vehicle to the most likely road segment where the vehicle was located at that time. As a result, a trajectory TR_i (or TR_{ij}) is converted to a sequence of road segments, denoted as a route R_i (or R_{ij}).

V. Trajectory Analysis

In this section, we conduct analysis on the two trajectory sets and try to answer the following two questions: 1) whether different routes have distinct road segment? 2) whether a vehicle has a consistent preference of road segments over a long period of time? If the answers to both questions are ”yes”, it implies that road segments traversed by a vehicle can be used as mobility features of this vehicle to distinguish this vehicle from others.

A. Road Segment Preferences

It is straightforward that the probabilities for a vehicle to traverse different road segments are not evenly distributed. Instead, there are several factors, such as the district this driver lives or works in, his or her habit and the characteristics of road segments, affecting a driver to select his or her travel routes. As a result, route R_i contains different road segments with each road segment appearing in the route for different times.

For each route $R_i, 1 \leqslant i \leqslant n$, we analyze the number of times that each road segment r_j appears in it, denoted as $t_{i,j}$. For a constant number k, the number of road segments which appear in a route more than k times can be seen as a random variable, denoted as X. Figure 1(a) and Figure 2(b) plot the CDFs of X using Shanghai and Shenzhen traces, respectively, under different values of k. For example, in Figure 1(a), point A in the curve of $k = 100$ means that 50% trajectories possess no more than 30 road segments satisfying $t_{i,j} \geqslant 100$. The top of the purple line indicates 100% trajectories possess no more than 200 road segments satisfying $t_{i,j} \geqslant 100$. It can be seen from Figure 1(a) and (b) that as k increases, the number of road segments satisfying $t_{i,j} \geqslant k$ decreases dramatically.

In summary, we have the following two observations from the above studies. First, there exist some road segments which are frequently travelled by certain vehicles. Second, although there are tens of thousands of road segments in both Shanghai and Shenzhen road networks, a certain vehicle only travels over a quite small part of a road network.

B. Difference between Trajectories

First, we define the following notations:

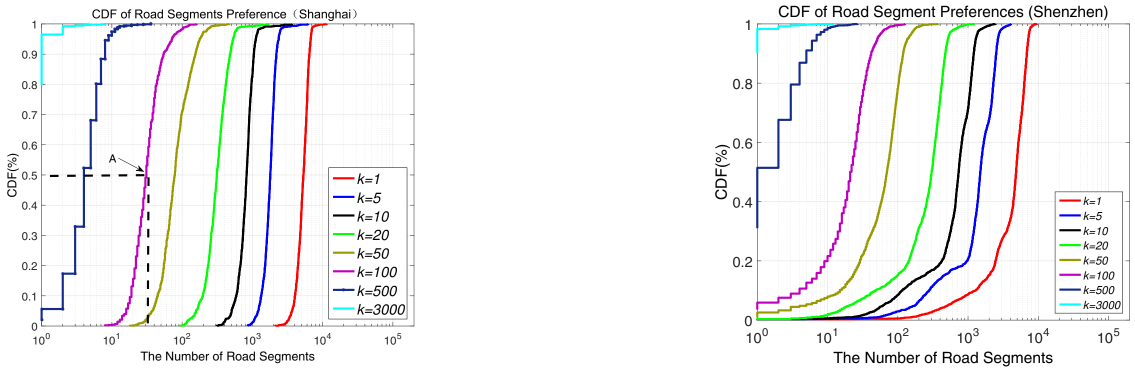
1) $S_i = \{r_j|t_{i,j} > 0\}$, i.e., S_i is a collection of road segments in route R_i.
2) $S_i^{(k)} = \{r_j|t_{i,j} \geqslant k\}$, i.e., $S_i^{(n)}$ is a collection of road segments which occur in route R_i at least k times.
3) $S_i^{(k)} \cap S_j^{(k)} = \{r_l|r_l \in S_i^{(k)} and\ r_l \in S_j^{(k)}\}$
4) $S_i^{(k)} \cup S_j^{(k)} = \{r_l|r_l \in S_i^{(k)} or\ r_l \in S_j^{(k)}\}$

Then, the similarity between trajectories TR_i and TR_j is defined as:

$$Sim_{i,j}^{(k)} = \frac{|S_i^{(k)} \cap S_j^{(k)}|}{|S_i^{(k)} \cup S_j^{(k)}|} \quad (1)$$

where $|S|$ represents the cardinality of a set S.

Given a threshold k, the similarity between a pair of trajectories can be calculated according to (1). Figure 2(a) and (b) plot the CDFs of similarities between trajectory pairs in Shanghai and Shenzhen data sets, respectively. It can be seen that the similarity of trajectory pairs in the two trace sets are both smaller than 0.45, which implies that individual



(a) Shanghai

(b) Shenzhen

Fig. 1. CDF of Road Segment Preferences

vehicles have unique routing preference. In addition, as k increases, the similarity is further decreased. It implies that, after eliminating the interference from low-frequency road segments, the road segment preference of individual vehicle will be more significant. For example, as $k = 100$, the similarities of 90% trajectory pairs are smaller than 0.1. In other words, on the perspective of road segment preferences, there are huge differences between trajectories.

VI. Strategies of The Adversary

Based on the above investigation, it is possible to calculate the frequency of occurrence of each road segment in a trajectory, constructing a feature vector, which can be used as quasi-identifier of the trajectory. In this section, we present one of the possible strategies used by an adversary to construct quasi-identifiers of trajectories and to achieve de-anonymization.

We assume that an attacker can access anonymous trajectory dat set $\mathcal{T} = \{TR_1, TR_2, ..., TR_n\}$, which contains anonymous trajectories of n vehicles. Meanwhile, the attacker obtains a segment of trajectory $\widetilde{TR}_\sigma$ of any time of a victim, named a *compromised trajectory*. The purpose of the attacker is to identify the TR_σ (called the *target trajectory*) from $\mathcal{T}$ which has the same identity of this victim according to $\widetilde{TR}_\sigma$. The attacker carries out three steps to launch the de-anonymization attack: 1) it transfers $\mathcal{T}$ and $\widetilde{TR}_\sigma$ to $\mathcal{R} = \{R_1, R_2, ..., R_n\}$ and $\widetilde{R}_\sigma$, respectively; 2) then it constructs the road segment feature vector of each route R_i using the improved *term-frequency-inverse document frequency* (TF-IDF) values of road segments, which reflect the importance of road segments in each route R_i; 3) finally, it conducts the feature vector matching algorithm using cosine similarity, and chooses the trajectory R^*, called the *suspect trajectory*, whose RS feature vector is most similar to that of $\widetilde{TR}_\sigma$. The attacker succeeds in the attack if the suspect trajectory R^* has the same identity as that of $\widetilde{TR}_\sigma$.

A. RS Feature Vector Extraction

As mentioned above, the importance of a road segment to a trajectory can be measured by using an improved TF-IDF value.

Specifically, for a given TR_i, we formulate an RS feature vector, $f_i = (w_{i,1}, w_{i,2}, ..., w_{i,\phi})^T$ where $w_{i,j}$ is the improved TF-IDF value of the $r_j (1 \leq j \leq \phi)$ and ϕ is the total number of road segments in the corresponding road network. The improved TF-IDF value $w_{i,j}$ is given by:

$$w_{i,j} = \frac{t_{i,j}}{\mathcal{N}_i} \times \log(c_{i,j} \times \frac{n}{c_j})$$

where $t_{i,j}$ is the number of times that road r_j appears in R_i and $\mathcal{N}_i$ is the length of R_i, i.e., the total number of road segments in R_i; therefore, $\frac{t_{i,j}}{\mathcal{N}_i}$ represents the frequency that r_j occurs in TR_i; c_j is the number of trajectories containing road r_j in D_{mt} and $c_{i,j}$ indicates how many 1-day trajectories of TR_i, i.e., $TR_{i,j}$, contains r_j.

B. RS Feature Vector Matching

Let $\tilde{f}_\sigma = (\mu_{\sigma,1}, \mu_{\sigma,2}, ..., \mu_{\sigma,\phi})^T$ represent the road segment feature vector of $\widetilde{TR}_\sigma$, where the value of $\mu_{\sigma,j}$ is calculated as:

$$\mu_{\sigma,j} = \frac{\tilde{t}_{\sigma,j}}{\sum_{j=1}^{\phi} \tilde{t}_{\sigma,j}}$$

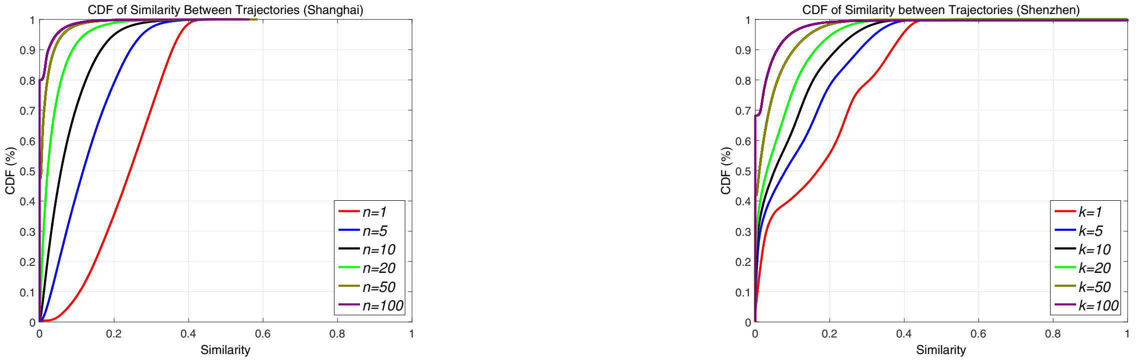
where $\tilde{t}_{\sigma,j}$ represents the number of times r_j occurs in $\tilde{R}_\sigma$.

We use cosine similarity to calculate the matching score $\mathcal{S}$ between $f_i (1 \leqslant i \leqslant n)$ and $\tilde{f}_\sigma$ as follows:

$$\begin{aligned}\mathcal{S}^{(i,\sigma)} &= \frac{f_i \cdot \tilde{f}_\sigma}{\|f_i\| \|\tilde{f}_\sigma\|} \\ &= \frac{\sum_{j=1}^{\phi} w_{i,j} \times \mu_{\sigma,j}}{\sqrt{\sum_{j=1}^{\phi} (w_{i,j})^2} \times \sqrt{\sum_{j=1}^{\phi} (\mu_{\sigma,j})^2}}\end{aligned}$$

C. Dimension Reduction of RS Feature Vector

In Subsection VI, the dimension of road segment feature vectors f_i extracted should be very high, since the value of ϕ depends on the scale of the road network. Fortunately, because f_i is a sparse vector, thus Principle Components Analysis (PCA) can be used to reduce the dimension of f_i. A detail description of dimension reduction is as follows.



(a) Shanghai

(b) Shenzhen

Fig. 2. CDF of Similarities Between Trajectory Pairs

1) Use the trajectories in D_{mt} to construct a $\phi \times n$ matrix $\mathbb{F} = [f_1, f_2, ..., f_n]$.
2) Carry out zero-mean normalization on $\mathbb{F}$, i.e., the average value of a row is subtracted from each element in the corresponding row.
3) Compute $Cov(\mathbb{F}) = \frac{1}{n}\mathbb{F} \cdot \mathbb{F}^T$.
4) Eigenvalue decomposition: $Cov(\mathbb{F}) = U \Lambda U^T$, where each row in U is a feature vector of $Cov(\mathbb{F})$ and $\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_\phi)$ is an ϕ order diagonal matrix, where diagonal elements are the eigenvalues of the corresponding feature vector, sorted in descending order.
5) Pick up top-d eigenvalues to satisfy the inequality:
$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{\phi} \lambda_i} \geqslant \alpha, 0 \leqslant \alpha \leqslant 1$$
which implies that the information contained in the top-d eigenvalues is at least α.
6) Choose the first d rows of U to build the projection matrix U_d, then a truncated $d \times n$ matrix $\mathbb{F}_d = U_d \mathbb{F}$ can be obtained. The i-th row of F_d refers to the dimension-reduced RS feature vector of TR_i.
7) After the zero-mean normalization of $\tilde{f_\sigma}$ (i.e., the average value of the a row in $\mathbb{F}$ is subtracted from the corresponding row of $\tilde{f_\sigma}$), the dimension-reduced road segment feature vector is: $\tilde{f_d} = U_d \tilde{f_\sigma}$.

VII. Evaluation

A. Methodology

We conduct trace-driven simulations to evaluate the proposed de-anonymization attack using the traces described in Section IV. Specifically, we pick trajectories collected in the first 20 (22) days from the Shanghai (Shenzhen) trace to constitute the anonymous trajectory dat set $\mathcal{T}$. The rest trajectories of eight (nine) days in the Shanghai (Shenzhen) trace, denoted as $\mathcal{T}'$, are compromised trajectories with the ID information of vehicles available to an attacker, who tries to use such trajectory to de-anonymize the anonymous trajectory dat set $\mathcal{T}$.

We evaluate the performance of the attack using the metric of *trajectory matching accuracy* (TMA), defined as,
$$TMA = \frac{n_{crt}}{n},$$
where n_{crt} denotes the number of trajectories in $\mathcal{T}$ which have been successfully identified and n is the total number of anonymous trajectories in $\mathcal{T}$.

B. The Impact of the Length of $\widetilde{TR}_\sigma$

We first investigate the impact of the length of a compromised trajectory $\widetilde{TR}_\sigma$ obtained by an attacker to the de-anonymization attack effect. We use all available trajectories in $\mathcal{T}'$ and vary the length of $\widetilde{TR}_\sigma$ from one day to eight (nine) days using the Shanghai (Shenzhen) trace with an increment of one day to conduct the attack.

Figure 3 plots the trajectory matching accuracy as a function of the length of $\widetilde{TR}_\sigma$. It can be seen that, even a short segment of a compromised trajectory can lead to a relatively high matching accuracy. For instance, the accuracy reaches around 71% and 52% when the length of $\widetilde{TR}_\sigma$ is only one day long for Shanghai and Shenzhen, respectively. In addition, as the attacker obtain more open trajectory, he/she can achieve a higher trajectory matching accuracy. For example, the matching accuracy can reach up to 95% and 70% when the length of $\widetilde{TR}_\sigma$ becomes eight and nine days for Shanghai and Shenzhen, respectively.

C. The Impact of the Size of Suspect Trajectory Set

In the original de-anonymization attack, only the anonymous trajectory with the highest matching score is treated as the suspect trajectory of a compromised trajectory. In practice, given a compromised trajectory, it also makes a lot of sense if the anonymous target trajectory which has the same ID with the compromised trajectory can be restricted within a small set of suspect trajectories. We generalise the definition of TMA as follows,
$$TMA = \frac{n_{crt}^m}{n},$$
where n_{crt}^m denotes the number of trajectories in $\mathcal{T}$ which have been successfully identified within a set of m suspect

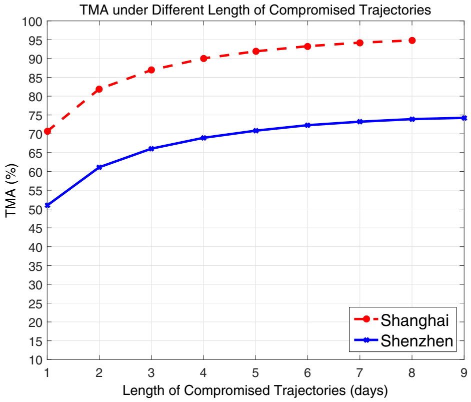


Fig. 3. TMA under Different Length of Compromised Trajectories

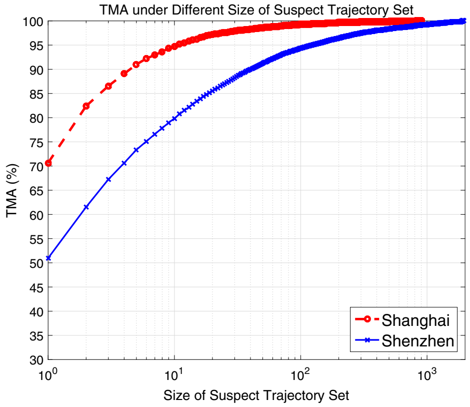


Fig. 4. TMA under Different Size of Suspect Trajectory Set

trajectories and n is the total number of anonymous trajectories in $\mathcal{T}$.

We study the impact of the size of a suspect trajectory set m (e.g., $m = 1$ in the above experiment) to the trajectory matching accuracy in this experiment. For a compromised trajectory, to constitute a set of m suspect trajectories, we rank all anonymous according to their matching scores and take the top m trajectories. Figure 4 plots the TMA as a function of m. It can seen that the TMA can reach up to 93% and 75% when m takes a small value of five using the Shanghai trace and the Shenzhen trace, respectively. Furthermore, TMA keeps increasing with greater m values and tends to stabilise.

VIII. Conclusion

In this paper, we have considered the privacy risk of publishing trajectories of moving objects whose true identities are made anonymous. We presented a comprehensive strategy for an attacker to well utilize the movement pattern of victim, extracted from several trajectory segments obtained either through direct observations or indirect information sources, to facilitate de-anonymization attacks. We experiment with two real data sets of mobility traces. Our results show that the movement feature of the victims could identify corresponding trajectories with high probability. These results stem from the fact that the spatial feature of trajectories tends to be unique to individuals and persistent. Our results show that the release of trajectory data in its original form except removing real identities is vulnerable to de-anonymization attacks. Furthermore, it's necessary to consider that the movement pattern might act as quasi-identifier and to design more elaborate privacy-preserving trajectory publishing method cloaking movement patterns and minimizing the reduction of utility.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (Grant No. 61300199, 61472255, 61420106010, and 61402101); by the Fundamental Research Funds for the Central Universities (Grant No. 2232014D3-21, and 2232015D3-29); by Shanghai Municipal Natural Science Foundation (Grant No. 14ZR1400900); by JSPS KAKENHI (Grant No. 16K00117, and 15K15976).

References

[1] GeoLife GPS Trajectories: http://research.microsoft.com/en-us/downloads/b16d359d-d164-469e-9fd4-daa38f2b2e13/
[2] Movebank: http://www.movebank.org/
[3] RAWDAD: a community resource for archiving wireless data at Dartmouth: http://crawdad.cs.dartmouth.edu/
[4] crowdflow.net: http://crowdflow.net/
[5] M. E. Nergiz, M. Atzori and Y. Saygin. Towards trajectory anonymization:a generalization-based approach. *In proceedings of the ACM SPRINGL'08*, 2008, pp. 52-61.
[6] 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 Transactions on Networking*, Vol. 21, Issue: 3, 2013, pp. 720-733.
[7] A. R. Beresford and F. Stajano. Location privacy in pervasive computing. *IEEE Pervasive computing*, 2003(1), pp. 46-55.
[8] M. Gruteser and X. Liu. Protecting privacy in continuous location-tracking applications. *IEEE Security & Privacy*, 2004(2), pp. 28-34.
[9] J. Krumm. A survey of computational location privacy. *Personal and Ubiquitous Computing*, 2009, 1(6), pp. 391-399.
[10] L. Kulik. Privacy for real-time location-based services. *SIGSPATIAL Special*, 2009, 1(2), pp. 9-14.
[11] L. Sweeney. k-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5), 2002.
[12] A. Meyerson and R. Williams. On the complexity of optimal k-anonymity. *In the proceedings of the PODS'04*, 2004, pp. 223228.
[13] M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. *in Proceedings of the ACM international conference on Mobile systems, applications and services*. ACM, 2003: 31-42.
[14] L. Buttyán, T. Holczer and I. Vajda. On the Effectiveness of Changing Pseudonyms to Provide Location Privacy in VANETs. *Chapter Security and Privacy in Ad-hoc and Sensor Networks of the series Lecture Notes in Computer Science*, Vol. 4572, pp. 129-141.
[15] H. Zang and J. Bolot. Anonymization of location data does not work: A large-scale measurement study. *In proceedings of the ACM Mobicom'11*, 2011, pp. 145-156.
[16] J. Freudiger, R. Shokri and J-P. Hubaux. Evaluating the privacy risk of location-based services. *Lecture Notes in Computer Science Chapter Financial Cryptography and Data Security*, Vol. 7035, 2012, pp. 31-46.
[17] A. Basu, A. Monreale, J. C. Corena, F. Giannotti, D. Pedreschi, S. Kiyomoto, Y. Miyake and T. Yanagihara. A privacy risk model for trajectory data. emphIn proceedings of the IFIP'14, 2014, pp. 125-140.
[18] D. J. Martin, D. Kifer, A. Machanavajjhala, J. Gehrke and J. Y. Halpern. Worst-case background knowledge for privacy-preserving data publishing. *In proceedings of the IEEE ICDE'07*, February 2007, pp. 126-135.
[19] Y. Lou, C. Zhang, Y.Zheng, X. Xie, W. Wang and Y.Huang. Map-matching for low-sampling-rate GPS trajectories. *In proceedings of the ACM SIGSPATIAL GIS'09*, 2009, pp. 352-361.