De-anonymization Attack Method of Mobility Trajectory Data Based on Semantic Trajectory Pattern

Wenshuai Zhang, Weidong Yang(✉), Haojun Zhang, and Zhenqiang Xu

Henan University of Technology, Zhengzhou 450001, China
{yangweidong,zhj,xuzhenqiang}@haut.edu.cn

Abstract. Anonymizing trajectory data based on pseudonyms is a common privacy protection method in data publishing scenarios. The so-called de-anonymization attack associates the anonymized trajectory data with the real identity of mobile object to further obtain the private information. The trajectory of a mobile object contains detailed spatio-temporal and semantic information. For anonymously released trajectory data, we propose a de-anonymization attack method based on semantic trajectory patterns, which uses a semantic trajectory pattern acquisition algorithm to obtain the frequent semantic trajectory pattern set of each mobile object, which is used as trajectory features to construct its mobility profiles, further design the corresponding similarity measure. Experiments on real trajectory datasets show that the method proposed in this paper can obtain a relatively high de-anonymization success rate.

Keywords: Privacy protection · De-anonymization attack · Frequent pattern mining

1 Introduction

Nowadays, with the application and development of mobile sensing, various positioning technologies and location services, the collection of spatio-temporal data has become largely convenient, and location service providers can easily collect the location trajectory data of billions of mobile objects. These data can be used to understand human behavior and to develop various services in various fields, such as vehicle congestion monitoring [1], path planning [2], friend recommendation [3], and travel mode analysis [4].

However, the mobility trajectory data set also contains a large amount of private information of mobile objects, such as residence, work location, physical condition, behavioral habits and other sensitive information. If these data are released or shared without protecting, it will seriously threaten the privacy of mobile objects. The literature [5] points out that continuous trajectory data exposure will be easier for attackers to obtain their behavior habits or interests. In order to protect the privacy of mobile objects' trajectory data, appropriate privacy-preserving mechanisms are usually used to

J. Xiong et al. (Eds.): MobiMedia 2021, LNICST 394, pp. 354–366, 2021.
https://doi.org/10.1007/978-3-030-89814-4_26

anonymize the original trajectory dataset before publishing it. These common privacy-preserving mechanisms can be divided into three main categories: 1) Modify the original trajectory (for example, by generalizing the location to the area to reduce the probability of an attacker identifying a mobile object) to protect privacy. 2) Add noise (for example, in the privacy protection mechanism based on differential privacy [6], by adding appropriate noise to the trajectory, the sensitive data is distorted while ensuring that the processed data can still maintain certain statistical properties). Reduce the trajectory accuracy, so as to achieve the purpose of privacy protection. 3) Use pseudonyms [7] to replace the real identity of the mobile object, and the real identity cannot be associated with the pseudonym in any way.

Among the above privacy protection mechanisms, the first two types of privacy protection mechanisms can have a better privacy protection effect on the trajectory data of mobile objects, but because the accuracy of the trajectory in time and space is reduced, the availability and completeness of the trajectory data set is greatly affected. The pseudonym technology has the advantages of not changing the original trajectory, easy to implement and maximum data availability, and is still one of the commonly used privacy protection methods.

In fact, even if the mobility trajectory is anonymized using pseudonym techniques, the attacker is still able to link the anonymous trajectory to the corresponding real identity with high probability. At present, the existing methods of de-anonymization attacks on mobility trajectory can be divided into two categories. One type is to extract movement features from location data and perform de-anonymization attacks through movement feature similarity. For example, Gambs et al. [8] proposed to construct mobile object behavior profiles for de-anonymization attacks by extracting the frequently occurring location points in the trajectory. This approach only considers single-dimensional location data, which makes the constructed mobility profiles not more accurately reflect the user behavior patterns embedded in the trajectories and affects the success rate of de-anonymization attacks. Another type of de-anonymization attack is to exploit the implied social relationships. This requires the assumption that the attacker can know the more complete real social relations, which is often difficult for the attacker to obtain the more complete real social relations in practical applications.

Researchers point out [9] that each person's mobile trajectory has its own inherent behavior pattern and does not change dramatically in the short term. At the same time, trajectories contain rich semantic information (e.g., semantic knowledge of their geographical location, behavioral preferences at a certain place, etc.) that can better reflect and represent the behavioral characteristics of different mobile objects.

Therefore, in order to characterize the mobility profile of different mobile objects more accurately and improve the success rate of de-anonymization. From the perspective of privacy attackers, this paper proposes a mobility trajectory de-anonymization attack method (TP-attack) based on semantic trajectory patterns, which combines the transfer time of mobile objects and the semantic trajectory characteristics, obtain the semantic trajectory pattern and design the corresponding similarity measurement to achieve anonymous trajectory de-anonymization attacks.

The main contributions of this paper are highlighted as follows:

- We propose a novel method to de-anonymization mobility trajectory. Considering the semantic information in the trajectory, we obtain the set of frequent semantic trajectory patterns in the trajectory as "fingerprint" to distinguish different individuals and build a mobility profile of mobile objects.
- Design a new similarity measure to compare the mobility profiles of mobile objects, so as to identify the real identity of the attacker's anonymous trajectory from the anonymous trajectory dataset based on the attacker's existing real trajectory information, and realize the anonymous trajectory de-anonymization attack.
- We perform experiments based on two real datasets. Results show that the set of frequent semantic trajectory patterns describes user behavior more accurately than others, Meanwhile, the experimental results show that the proposed method can obtain a high success rate of de-anonymization.

The rest of this paper is organized as follows. Section 2 introduces related work, Sect. 3 gives basic definitions and problem descriptions, Sect. 4 describes how to obtain semantic trajectory patterns, Sect. 5 introduces the TP-attack method, and Sect. 6 conducts related experiments and analyzes the experimental results, and Sect. 7 makes a conclusion with this paper.

2 Related Works

Protecting the anonymity of personal mobility is notoriously difficult due to sparsity [10] and hence mobility data are often vulnerable to deanonymization attacks. Many studies on trajectory privacy show that even if a person's data is anonymized, they still have unique patterns that may be exploited by malicious attackers who background knowledge. At present, many researchers have conducted in-depth research on the de-anonymization attack methods for mobility trajectories. This section divides them into two categories: One type is to extract movement features from location data and perform de-anonymization attacks through movement feature similarity that is, by extracting characteristic positions from the mobility trajectory (such as places frequently visited by mobile objects, stopping points, etc.) to express the behavior characteristics of mobile objects. Another type of de-anonymization attack is to exploit the implied social relationships.

Mulder et al. [11] proposed a de-anonymization attack method, by establishing a Markov model for each individual in the training set, and then using the model to maximize the likelihood to perform a de-anonymization attack on the individuals in the test set. However, this method only considers location information and does not consider the influence of time at all, and it cannot reflect the differences of different individuals.

Zhong et al. [12] proposed a novel attack method. First, perform characteristics analysis on anonymous mobility trajectories, use an optimized word frequency-inverse document frequency method to construct characteristic vectors, use partial trajectory fragments to analyze the anonymous data set of mobility trajectories, and combine characteristic similarity to analyze and match. Finally, from the anonymous trajectory

data, the mobility trajectory with the highest similarity to the collected trajectory is analyzed to achieve the purpose of de-anonymization. The above methods based on trajectory characteristic positions have high computing efficiency, but do not consider the relevance in the time dimension, resulting in the accuracy of de-anonymization that needs to be improved.

Chris Y. T. Ma et al. [13] proposed a side information (i.e. location snapshots) based de-anonymization attack, where an attacker, obtaining a number of location snapshots of its victims, can recognize the trace of the victims from a set of anonymous traces. The authors use Bayesian inference to break the unlinkability between location snapshots and anonymized trace. H. Wang et al. used two mobile social network datasets as side information to evaluate the performance of de-anonymization attacks using external information. A Gaussian and Markov based algorithm is adapted to deal with spatiotemporal mismatches in different datasets [14]. H. Li et al. [15] measured the similarity between the disclosed locations in the MSN applications and the real mobility pattern, in terms of coverage rate and relative entropy, and presented an attack to infer MSN users' demographics from the disclosed locations by checking their similar point of interests.

3 Definitions and Attack Model

3.1 Definitions

The source of the mobility trajectory dataset is mainly the dataset generated by emerging Internet applications such as the Internet of Vehicles and mobile social networks. It has the following description:

Definition 1 (Stay Region). A stay region refers to a geographic location area where a mobile object stays within an area for a given time range. The semantic label of each stay area is the semantic information corresponding to the location point with the highest importance score of the stay area, symbolized as SR_i. In addition, the transfer time between two consecutive stay regions is $\alpha_i = t_{i+1\_arv} - t_{i\_lev}$. Where $t_{i_{arv}}$ represents the time when the first point p_1 reaches the stay region, and $t_{i_{lev}}$ is the time when the last point p_n leaves the stay region. In this way, a set of semantic stay points with transfer time for each mobile object can be obtained.

Definition 2 (Semantic Trajectory). A semantic trajectory is a sequence of n stay regions with time information and semantic annotations, denoted as $SR = (< SR_1, \alpha_1 >, \cdots, < SR_i, \alpha_i >)$, the length of the semantic trajectory is the number of semantic stay points.

Definition 3 (Semantic Trajectory Pattern). The trajectory pattern represents the regular movement of a mobile object. In this paper, it is usually denoted as sequences of semantic stay point with transfer time annotated, denoted as $SR_1 \xrightarrow{\alpha_1} \cdots \xrightarrow{\alpha_i} SR_i$.

Definition 4 (Frequent Semantic Trajectory Pattern Set). Given a set of semantic stay point sequences G, the minimum support θ, the frequent semantic trajectory pattern set of G is formalized as $FT_\theta^G = \left\{F_i | support^G(D_i) \geq \theta\right\}$ where the minimum support represents the percentage of trajectories containing D_i in the set G.

3.2 Attack Model

In order to protect the privacy of the trajectory of mobile objects, the mobility trajectory dataset needs to be anonymized before it is released. For any trajectory T_i, a pseudonym $w_i \in \mathrm{W}$ is used to replace the real identity $v_i \in \mathrm{V}$ of the creator of each trajectory, and each moving object corresponds to a unique pseudonym, so that the attacker cannot associate his real identity with his real position, thus protecting the mobile Object trajectory privacy.

The attacker has access to a set of anonymous trajectory dataset H, which includes the anonymous mobile trajectory of one or more attack targets, and the attacker can obtain several mobility trajectory fragments of the attack target at any time period in the future through observation or other methods. We call these trajectory fragments as background knowledge dataset K, in which user identities are known. The attacker's goal is to find out the true identity of the target's trajectory from the anonymous trajectory dataset by analyzing the trajectory characteristics of different mobile objects.

4 Semantic Trajectory Mode Acquisition

Frequent semantic trajectory patterns appearing in the trajectory of mobile objects reflect personal habits and behavior patterns, and can be used as quasi-identifiers to distinguish different individuals [16]. By extracting the stay region in the mobility trajectory and making it semantic, the semantic trajectory pattern that characterizes the life and behavior of the mobile object is further obtained, which is used as the trajectory feature to construct its mobility profile.

4.1 Extraction of Stay Region

When a mobile object generates activity in a certain geographic location, the speed will be lower than the average speed of the entire trajectory. The speed-based stay region recognition algorithm in [17] is used to cluster the stay points belonging to the same stay region to obtain the stay region sets.

4.2 Semantic Stay Region

Considering that mobile objects will have different behavioral information in the stay region for the same longitude and latitude, the semantics of the stay region is important to reveal the behavioral patterns. The literature [18] uses a predefined category of locations to represent the semantics of a stay region. However, each mobile object has different interest preferences in the stay region, and it is not very accurate to represent the semantics of mobile object activities in the stay region by this predefined approach. In fact, the semantics of an individual's important location can be derived from a person's long-term trajectory data. In this section, the semantic of stay region can be expressed by selecting the position with the highest importance score in the stay region.

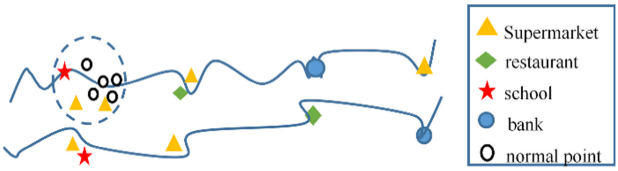

Fig. 1. Examples of stay region

Suppose that the position of the dash coil in Fig. 1 is a stay region of mobile object, which contains spatio-temporal points and POI. After identifying the stay region, calculate the importance score of each location in the area, select the highest score value as the stay point, and then call the Google API interface to semantically the stay point, so that the location data Correlate with semantic information to obtain a set of semantic stay region sequences. The importance score is defined as:

$$\text{score}(p_i) = \frac{sigp_i}{Dp_i} \tag{1}$$

where $sig(p_i) = \frac{|p_i|}{|sr|} \cdot \frac{|T_{pi}|}{|T|}$ as the importance of the location. $D(p_i)$ represents the distance between the position p_i and the nearest POI. $|p_i|$ indicates the number of times the mobile object visited the location, $|sr|$ indicates the number of positions in the stay region. $\left|T_{p_i}\right|$ indicates the number of traces containing the position. $|T|$ represents the total number of traces of mobile object u.

It is known that the POI (supermarket) location points in the stay region have a higher importance score, so the supermarket is used as the stay point in the stay region. In this way, a set of semantic stay points with transfer time for each mobile object can be obtained.

4.3 Semantic Trajectory Pattern Acquisition Algorithm

After obtaining the set of semantic stay region sequences, PrefixSpan [19] algorithm is improved, combined with the transfer time between stay regions, to obtain the semantic trajectory pattern with a support value less than θ, thereby obtaining the frequent trajectory pattern set of each mobile object. The specific process is shown in Algorithm 1:

Algorithm 1: Semantic trajectory pattern acquisition algorithm

Input: Semantic stay region sequence set R, minimum support θ, time thresholds τ
Output: frequent semantic trajectory pattern set Q

1: k ← 0
2: $P_0 \leftarrow \{T \times \{<>\}\}$
3: while $P_k \neq \emptyset$ do
4: for all P∈ P_k do
5: lastSR ← getlastStayRegion(P. prefix)
6: for all i ∈ P do
7: if support(P, i, τ) ≥ θ then
8: interval← ExtractIntervals(lastSR, i, P)
9: P_{k+1}. prefix ← link(prefix, i)
10: D_i ←GTP(P_{k+1}. prefix, $intervals$)
11: Output(D_i)
12: $P_{k+1} \leftarrow generateProjection(P, i)$
13: end if
14: end for
15: end for
16: k++
17: end while

Algorithm 1 gives the procedure of the semantic trajectory pattern acquisition algorithm. In order to obtain frequent sequences of stay regions with transfer times, the PrefixSpan algorithm is improved on by first taking the given set of semantic stay region sequences as the initial values of the projection database, and then finding the set of all frequent stay regions in the projection database as the initial set of prefixes. In each recursive process, the last set of stay regions of the current projection database prefix is first obtained, and then the set of stay regions with support satisfying the minimum threshold is found. The function in lines 8–11 functions to obtain the frequent time interval between two consecutive stay regions and combine the set of frequent stay regions with the current prefixes to generate a frequent semantic trajectory pattern with time interval. A projection database of this set of frequent stay regions is also obtained.

5 De-anonymization Attack Methods

5.1 An Approach Overview

From the de-anonymization attack model, it is clear that the attacker aims to identify the true identity corresponding to the attack target's trajectory from the anonymous trajectory dataset based on the trajectories in the background knowledge dataset. This section gives a strategy for an attacker to perform a de-anonymization attack. This is shown in Fig. 2.

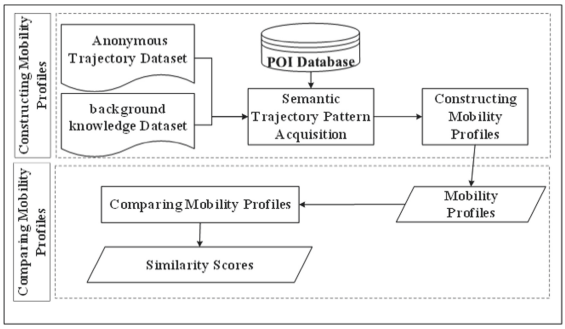

Fig. 2. The framework of our method.

The de-anonymization attack of mobile trajectory data can be divided into two phases:

- Constructing mobility profiles. The attacker constructs a mobility profile of the mobile object for several trajectories of the mobile object in the background knowledge dataset and the anonymous trajectory dataset, respectively.
- Comparing mobility profiles. A suitable similarity metric is designed to compare the similarity of the user mobility profiles constructed in the first phase so that the trajectories of the attack targets can be identified from the anonymous trajectory set.

5.2 Constructing Mobility Profiles

Mobility profiles describe the daily behavior patterns of mobile objects, which reflect the patterns of commonly visited locations. However, the longer the semantic mobility is, the greater the number of subsequences it generates. In order to avoid the error problem caused by repeated calculation, the maximum pattern set [20] is used as the mobility profiles of the mobile object. In this case, the maximal pattern set is composed of trajectory patterns that are not subsequences of any other pattern. The maximum trajectory pattern set is defined as:

Definition 5: (Maximum pattern set). Given mobile object u's frequent trajectory pattern set FT_θ^G, where Q and P are the trajectory patterns of mobile object u, then the maximum trajectory pattern set of user u is defined as:

$$\mathrm{M}\left(FT_\theta^G\right) = \left\{P \in FT_\theta^G \middle| \nexists Q \in FT_\theta^G (P \sqsubseteq Q)\right\} \tag{2}$$

Then for the anonymous dataset G of the anonymous mobile object u and the auxiliary data set G' of the real mobile object u' to construct the mobility profiles:

$$\mathrm{M}\left(FT_\theta^G\right) = (D_1, D_2, \cdots, D_i) \tag{3}$$

$$M\left(FT_\theta^{G'}\right) = \left(V_1, V_2, \cdots, V_j\right) \tag{4}$$

Where D_i and V_j are the frequent patterns of u and the frequent patterns of u' obtained by frequent trajectory pattern acquisition algorithm, respectively.

5.3 Comparing Mobility Profiles

The more similar two mobile objects u and $u^{'}$ are, the more similar the corresponding mobility profile is. The essence of measuring the mobility profile of two mobile objects is to measure the similarity between the set of trajectory patterns. This section measures the distance between two maximal trajectory patterns on the basis of DTW [21]. Traditional distance calculation methods such as Euclidean distance are very sensitive to even small mismatches and require that the two time series are of equal length. In contrast, DTW is more suitable for our scenario and overcomes these drawbacks well.

Definition 6: (DTW): Give u_i and $u_i^{'}$, along with their corresponding maximal trajectory pattern $D_i = SR_1 \xrightarrow{\alpha_1} \cdots \xrightarrow{\alpha_n} SR_n$ and $V_j = SR_1 \xrightarrow{\alpha_1} \cdots \xrightarrow{\alpha_m} SR_m$, DTW aims to calculate the minimum distance between each semantic stay region SR_i and SR_j, which requires constructing an $n \times m$ Distance Matrix and finding a path $\varphi = \{\varphi(1), \cdots, \varphi(T)\}$, making the path through the elements minimizes the distance of SR_i and SR_j. We represent each element $\varphi(k) \in \varphi$ as the distance between stay region SR_i and SR_j. Where $\varphi(k) = (\varphi_{D_i}(k), \varphi_{V_j}(k))$, k = 1,2,⋯,T.

Obtaining the optimal Warping Curve Φ, the distance between SR_i and SR_j is minimized by:

$$DTW\left(D_i, V_j\right) = \min_{\varphi} \sum_{k=1}^{T} \varphi(k) \tag{5}$$

Different mobile objects have different transfer times in the continuous stay region, reflecting different intentions. For any two maximum trajectory patterns D_i and V_j, we use the average of the overlapping time ratios between all consecutive staying regions in all the longest common subsequences as the time weight DOF $\left(D_i, V_j\right)$ to distinguish different mobile objects.

$$\mathrm{DOF}\left(D_i, V_j\right) = \frac{\sum_{S \in lcs\left(D_i, V_j\right)} \sum_{i=1}^{len(S)-1} ot_S^{u,u^{'}}(i)}{\left|lcs\left(D_i, V_j\right)\right| \cdot \left(LCS\left(D_i, V_j\right) - 1\right)} \tag{6}$$

where $ot_S^{u,u^{'}}(i) = \frac{\sum_{1 \le i \le m} \alpha_{i\_max}^{'} - \alpha_{i\_min}^{'}}{\sum_{1 \le i \le k} \alpha_{i\_max} - \alpha_{i\_min}}$ represent the overlap time ratio between any two consecutive SR_{i-1} and SR_i in the longest common subsequence between u and $u^{'}$, $\sum_{1 \le i \le m} \alpha_{i\_max}^{'} - \alpha_{i\_min}^{'}$ represents the overlapping transition time between any two consecutive SR_{i-1} and SR_i, $\sum_{1 \le i \le k} \alpha_{i\_max} - \alpha_{i\_min}$ represents the all the occurring transition time between any two consecutive SR_{i-1} and SR_i. Therefore, the similarity between the maximum trajectory patterns D_i and V_j is calculated as:

$$\mathrm{SIM}\left(D_i, V_j\right) = DTW\left(D_i, V_j\right) \cdot DOF\left(D_i, V_j\right) \tag{7}$$

Given two mobile objects u and $u^{'}$, we only consider the trajectory pattern with the highest similarity value of the maximum trajectory pattern, denoted as $SIM_{MAX}\left(D_i, V_j\right)$

to measure between the maximum pattern sets, and obtain the mobility profiles similarity score, which is calculated as follows:

$$sim\left(u|u^{'}\right) = \frac{\sum_{D_i \in \mathrm{M}(FT_{\theta}^{G})} \sum_{V_j \in \mathrm{M}(FT_{\theta}^{G'})} SIM_{MAX}(D_i, V_j) \cdot \mu(D_i, V_j)}{\sum_{D_i \in \mathrm{M}(FT_{\theta}^{G})} \sum_{V_j \in \mathrm{M}(FT_{\theta}^{G'})} \mu(D_i, V_j)} \tag{8}$$

where $\mu\left(D_i, V_j\right) = \frac{support_u(D_i) + support_{u'}(D_i)}{2}$ denoted as a weight function constructed using the support value of the trajectory pattern [20].

6 Evaluation

6.1 Experimental Setup

The trajectory datasets used in the experiment are the Geolife dataset [22] and the CabSpotting dataset [23]. The specific information of the data set is shown in Table 1.

Table 1. Description of datasets.

Datasets	Mobile objects	Localization
Geolife	42	Beijing
CabSpotting	536	San Francisco

The experiment selected the most frequent 30-day trajectory data of mobile objects from these two datasets, and regarded the mobility of the entire time period as a mobility trajectory. In the experiment described in the article, we use the trajectory data of the mobile objects in the first 15 days of these two datasets as the anonymous trajectory data set H, that is, randomly renumber the trajectory of each mobile object in the previous 15 days to anonymize the trajectory data. The trajectory data of mobile objects in the next 15 days is the background knowledge data set K (trajectory collection of known user identities) for the de-anonymization attack stage.

Specifically, by de-anonymizing the anonymous trajectory dataset, Measuring the similarity of mobility profiles between the trajectory in the anonymous trajectory dataset H and the trajectory in the dataset K. The anonymous trajectory with the highest similarity to the trajectory in K is regarded as the same mobile object trajectory, and then compare the trajectory Whether the number of the mobile object of is consistent with the number of the known trajectory, if it is the same, the result of de-anonymization of the trajectory is accurate. Finally, the effect of the de-anonymization attack is measured by the following formula (9):

$$\mathrm{H}_{accuracy} = \frac{n}{N} \tag{9}$$

where n represents the number of trajectories that match correctly, and N represents the total number of trajectories in the anonymous dataset.

6.2 Competitors

In order to verify the effectiveness of the TP-attack method, we compare it with the following two mobility trajectory de-anonymization methods for de-anonymization success rates: 1) POI-attack [24]: The attacker extracts the trajectory from the anonymous trajectory feature POI, using POI collection to build mobility profile. Compare with the background knowledge you have to identify the true identity corresponding to the anonymous trajectory. 2) MMC-attack [8]: The attacker extracts the trajectory from the anonymous trajectory feature POI, constructing Markov chain model and compare the similarity of two Markov chains.

6.3 Experiment Analysis

In this section, we first studied the influence of the amount of background knowledge the attacker has on the success rate of de-anonymization. The experiment uses the two datasets mentioned in Table 1, and randomly selects $d(1 \leq d \leq 15)$ days of trajectory data from K to form the attacker's known dataset. Then use the TP-attack, POI-attack and MMC-attack methods to de-anonymize them. The X-axis and Y-axis represent the background knowledge of the attacker and the accuracy of trajectory matching. Figures 3 and 4 respectively show the influence of the amount of background knowledge the attacker has on the success rate of de-anonymization on the two datasets of Geolife and CabSpotting.

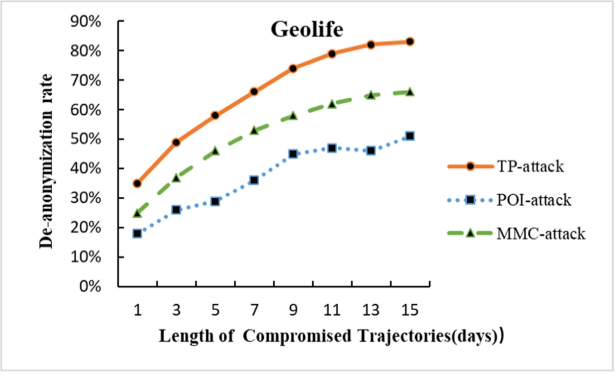

Fig. 3. Influence of the amount of background knowledge (Geolife)

By gradually increasing the attacker's background knowledge, the success rate of de-anonymity attacks is significantly increased; but with the increase of $d(d \geq 7)$, the rise gradually tends to be flat. In contrast, the success rate of TP-attack is better than the POI-attack and MMC-attack methods. The POI-attack method does not consider the time sequence; the MMC-attack method also does not consider the time factor, and the Markov chain is calculated multiple times, and the method of measuring the similarity of the Markov chain will have a certain impact on the success rate of de-anonymization. From Figs. 3 and 4, it can be seen that the more background knowledge an attacker has, the easier it is to carry out a de-anonymization attack.

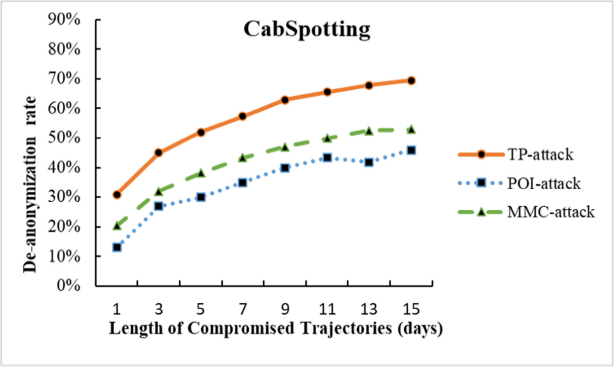

Fig. 4. Influence of the amount of background knowledge (CabSpotting)

7 Conclusion

In this paper, we propose a de-anonymization attack method based on the semantic trajectory pattern. By obtaining the semantic trajectory pattern of the mobile object, and using the transfer time of the mobile object between the stay region as the weight of the similarity measure, it can be more accurately characterize the trajectory patterns of different mobile objects, a unique mobility profiles of the mobile objects can be constructed for similarity measurement. The experimental results show that as the attacker has more background knowledge, the easier it is for the attacker to de-anonymization attack. At the same time, by comparing other mobility trajectory de-anonymization attack methods, the method based on the semantic trajectory pattern can obtain a higher success rate. Trajectory privacy protection technology is constantly improving, and in the next step, we will conduct research on de-anonymization attacks against other privacy protection technologies and explore the problems in trajectory privacy protection technology.

Acknowledgement. This work was supported by National Natural Science Foundation of China (61772173); Program for the Innovative Talents of the Higher Education Institutions of Henan Province (19HASTIT027); Open fund of Key Laboratory of Grain Information Processing and Control (under Grant No. KFJJ-2018105).

References

1. Kamble, S.J., Kounte, M.R.: Machine learning approach on traffic congestion monitoring system in internet of vehicles. Procedia Comput. Sci. **171**, 2235–2241 (2020)
2. Aggarwal, S., Kumar, N.: Path planning techniques for unmanned aerial vehicles: a review, solutions, and challenges. Comput. Commun. **149**, 270–299 (2020)
3. Uplavikar, N.M., Vaidya, J., Lin, D., Jiang, W.: Privacy-preserving friend recommendation in an integrated social environment. In: Kanhere, S., Patil, V.T., Sural, S., Gaur, M.S. (eds.) ICISS 2020. LNCS, vol. 12553, pp. 117–136. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-65610-2_8

4. Zhao, X., Yan, X., Yu, A., et al.: Prediction and behavioral analysis of travel mode choice: a comparison of machine learning and logit models. Travel Behav. Soc. **20**, 22–35 (2020)
5. Gao, Q., Zhang, F.L., Wang, R.J., Zhou, F.: Trajectory big data: a review of key technologies in data processing. J. Softw. **28**(4), 959–992 (2017)
6. Andrés, M.E., Bordenabe, N., Chatzikokolakis, K., Palamidessi, C.: Geo-indistinguishability: differential privacy for location-based systems. In: Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security, pp. 901–914 (2013)
7. Pfitzmann, A., Köhntopp, M.: Anonymity, unobservability, and pseudonymity—a proposal for terminology. In: Designing Privacy Enhancing Technologies, pp. 1–9 (2001)
8. Gambs, S., Killijian, M.O., del Prado Cortez, M.N.: De-anonymization attack on geolocated data. J. Comput. Syst. Sci. **80**(8), 1597–1614 (2014)
9. Chang, S., Li, C., Zhu, H., Lu, T., Li, Q.: Revealing privacy vulnerabilities of anonymous trajectories. IEEE Trans. Veh. Technol. **67**(12), 12061–12071 (2018)
10. Aggarwal, C.C., Philip, S.Y.: A general survey of privacy-preserving data mining models and algorithms. In: Aggarwal, C.C., Yu, P.S. (eds.) Privacy-Preserving Data Mining, vol. 34, pp. 11–52. Springer, Boston (2008). https://doi.org/10.1007/978-0-387-70992-5_2
11. De Mulder, Y., Danezis, G., Batina, L., Preneel, B.: Identification via location-profiling in GSM networks. In: Proceedings of the 7th ACM Workshop on Privacy in the Electronic Society, pp. 23–32 (2008)
12. Zhong, J., Chang, S., Liu, X., Song, H.: De-anonymization attack method for mobile trace data. Computer Engineering (2016)
13. Ma, C.Y., Yau, D.K., Yip, N.K., Rao, N.S.: Privacy vulnerability of published anonymous mobility traces. In: Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking, pp. 185–196 (2010)
14. Wang, H., Gao, C., Li, Y., Wang, G., Jin, D., Sun, J.: De-anonymization of mobility trajectories: dissecting the gaps between theory and practice. In: The 25th Annual Network & Distributed System Security Symposium (NDSS 2018) (2018)
15. Li, H., Zhu, H., Du, S., Liang, X., Shen, X.: Privacy leakage of location sharing in mobile social networks: Attacks and defense. IEEE Trans. Dependable Secure Comput. **15**(4), 646–660 (2016)
16. Sun, L., Yu, K.: Research on big data analysis model of library user behavior based on Internet of Things. Comput. Eng. Softw **40**(6), 113–118 (2019)
17. Lin, Z., Zeng, Q., Duan, H., Liu, C., Lu, F.: A semantic user distance metric using GPS trajectory data. IEEE Access **7**, 30185–30196 (2019)
18. Mazumdar, P., Patra, B.K., Lock, R., Korra, S.B.: An approach to compute user similarity for GPS applications. Knowl.-Based Syst. **113**, 125–142 (2016)
19. Cai, G., Lee, K., Lee, I.: Mining semantic trajectory patterns from geo-tagged data. J. Comput. Sci. Technol. **33**(4), 849–862 (2018)
20. Chen, X., Pang, J., Xue, R.: Constructing and comparing user mobility profiles for location-based services. In: Proceedings of the 28th Annual ACM Symposium on Applied Computing, pp. 261–266 (2013)
21. Kate, R.J.: Using dynamic time warping distances as features for improved time series classification. Data Min. Knowl. Disc. **30**(2), 283–312 (2015). https://doi.org/10.1007/s10618-015-0418-x
22. Zheng, Y., Xie, X., Ma, W.Y.: Geolife: a collaborative social networking service among user, location and trajectory. IEEE Data Eng. Bull. **33**(2), 32–39 (2010)
23. Piorkowski, M., Sarafijanovic-Djukic, N., Grossglauser, M.: CRAWDAD data set epfl/mobility. 24 February 2009 Downloaded from (2009)
24. Primault, V., Mokhtar, S. B., Lauradoux, C., Brunie, L.: Differentially private location privacy in practice. In: Third Workshop on Mobile Security Technologies (MoST) (2014)