2016 15th International Symposium on Parallel and Distributed Computing

A Privacy-Preserving Approach Based on Graph Partition for Uncertain Trajectory Publishing

Jianchuan Xiao, Li Xu*, Limei Lin, Xiaoding Wang
Fujian Provincial Key Laboratory of Network Security and Cryptology School of Mathematics and Computer Science
Fujian Normal University, Fuzhou, Fujian, China, 350117. Email:xuli@fjnu.edu.cn.

***Abstract*—Various services such as location-based service (LBS) allow mass collection of spatio-temporal data because the ubiquity of cheap embedded sensors on smart phones. Therefore, the individual privacy-preserving is receiving increasing attention during the data publication. However, the inherent inaccuracy of data acquisition equipments, sampling error and low sampling rate may lead to uncertainty. In this paper, we propose a privacy-preserving approach for trajectory publication with considering the uncertainty in trajectory. The correlation between two trajectories are computed according to the temporal overlap similarity, the trajectory direction similarity and the distance between trajectories with uncertainty. Then a greedy algorithm is proposed to achieve k-anonymity based on graph partition. The analysis and experiment evaluations based on the GeoLife trajectory data set show that significant privacy and QoS benefits can be achieved.**

***Index Terms*—Privacy security, uncertain trajectory publishing, graph partition, clustering.**

I. Introduction

With the course of urbanization is accelerated unceasingly, making a city smart is a new approach to urban development [1]. In the era of smart city and the Internet of Things, various data such as location, time, velocity are produced and collected easily by mobile phones, which are equipped with global positioning system, radio frequency identification an so on. At the same time, lots of smart city applications make use of the information collected by GPS to provide location-based services (LBS). Users get much benefits from the development of LBS, such as local recommendations for shopping or dining, nearest gas station or restaurant recommendation. However, LBS would be a double-edged sword while lack of privacy preserving. More specifically, publishing vast volumes of trajectories for research purposes has raised serious privacy concerns [2]. Trajectories are closely relate to users sensitive information, such as home address, working place, religious faith and so on [2]. Attackers take advantage of background knowledge can match certain trajectory with the corresponding user even though the true trajectory identification is removed [3].

For the purpose of privacy protection for trajectory publishing, it is necessary to model the trajectory appropriately. Trajectories data is not precise because low sampling rate and sampling error [4]. In many practical scenarios, the sampling rate should be reduced considering the communication and update costs [5]. Moreover, the limitation of GPS devices, measurement inaccuracy may lead to sampling error. Privacy-preserving methods based on k-anonymity without considering the uncertainty may result in error clustering, and it is not suitable for practical application.

In this paper, we aim to propose a personalized privacy-preserving approach for uncertain trajectory clustering based on graph partition. The contributions of this paper are summarized below

1) We consider the uncertain direction similarity, the temporal overlap similarity and the uncertain distance to define the correlation of trajectories with appropriate proportion since user's personalized requirements.
2) A greedy algorithm is proposed to partition trajectories into several k-anonymity sets according to the correlation of trajectories.
3) We conduct a comprehensive set of experiments over the GeoLife trajectory data set. The effects of parameters of the temporal overlap similarity, the trajectory direction similarity, the distance between trajectories and the uncertain parameter ξ on the privacy level evaluation and information loss evaluation are analysed. The simulation results show that we can improve the performance of the proposed approach by well designing these parameters.

II. Relate Work

Trajectory k-anonymity has been attracting much attention in research. It is very challenging to ensure a good similarity between trajectories for clustering k-similar trajectories. Some works construct trajectory k-anonymity sets based on graph-partition theory [6-8].

Study in [6] was the first to formalize trajectory k-anonymity problem into a k-node graph partition, then proved the partition is NP-complete. The trajectory data set could be represented by an undirected graph, in which the vertices represent trajectories and the weights of edges represent correlation among trajectories. However, this approach focused on the similarity associate with the overlap sampling points, did not take the movement direction factor into account. The authors in [7] demonstrated that the trajectory directions not only can affect the trajectories privacy protection level, but also help to construct a appropriate trajectory k-anonymity set. Besides, the authors also proposed trajectory angle to evaluate trajectory direction and similarity. A personalized (s, δ)-coverage approach [8] was proposed to trade off between data utility and privacy-preserving considering trajectory angle and distance between trajectories based on a personal trajectory

978-1-5090-4152-7/16 $31.00 © 2016 IEEE
DOI 10.1109/ISPDC.2016.48

graph model. However, they did not take the temporal dimensional similarity into account. All of the above methods regarded the sampling point information as accurate and did not take the uncertainty into account.

Abul et al. in [9] presented (k,δ)-anonymity privacy preserving model to protect the privacy of uncertain trajectory data. In this privacy model, parameter k has the same meaning as in k-anonymity, while δ represents a lower bound of the uncertainty radius when recording the locations of trajectories. There exist two anonymization methods named Never Walk Alone(NWA, [9]) and Wait for Me(W4M, [10]) regarded an uncertain trajectory as a cylindrical volume of radius δ. However, these researches regarded the radius δ as constant without considering the change of GPS accuracy. A new model was proposed in [11] for uncertain trajectory that the uncertain radius δ is changed with different sampling time. The authors in [12] determined uncertain sampling radius according to the speed of trajectory sampling point.

There are several methods for uncertain trajectories distance measurement. Clustering approaches based on fuzzy logic [13], such as FCM [14], is used in uncertainty processing. In [14], the authors proposed a variant of the Fuzzy C-Means (FCM) clustering algorithm to cluster uncertain trajectories. Authors in UK-means [15], which are based on the classical K-means clustering, proposed the expected distance between uncertain moving objects. An uneven two step sampling approach [12] was proposed to re-sample trajectory based on probability density function and simplify the calculation of distance between uncertain trajectory points. The authors in [16] proposed a definition to measure the maximum distance of uncertain object from an uncertain query issuer.

III. System Model

In this paper, we consider the trajectory direction similarity $S_{dire}[\xi, T_p, T_q]$ and temporal overlap similarity $S_{time}[p, q]$ to evaluate the similarity between trajectories then evaluate the data utility by distance between trajectories with uncertainty $D_{loc}[\xi, T_p, T_q]$.

A moving object dataset include n trajectories which can be denoted as $D = \{T_1, T_2, ..., T_n\}$, where T_k denotes the k-th trajectory of the dataset, $k = 1, 2, ..., n$. We assume that the moving object is moving along a straight line in a constant speed while velocity can be different at two successive sampling points. We regard the sampling points as not accurate because certain inevitable sampling error or law sampling rate. Therefore, trajectory data which is a cylinder, and the radius of the cylinder is the length of the error denote as δ.

A trajectory is considered as a cylinder in three dimensional space, which can be represented as a sequence of spatiotemporal points. The p-th trajectory is denoted as $T_p = \{tid, \xi, (x_p^1, y_p^1, \delta_p^1, t_p^1), (x_p^2, y_p^2, \delta_p^2, t_p^2), ..., (x_p^n, y_p^n, \delta_p^n, t_p^n)\}$, where tid is the identifier of trajectory and ξ is the uncertain parameter, ξ is the uncertain parameter, coordinate (x, y) is the expected location, δ is the uncertain radius and t is the sampling time. For each point (x, y, δ, t) along trajectory, its uncertain area is the horizontal disk with radius δ, and centered at expected coordinate (x, y). Because the sampling points are not accurate for the sake of sampling error or low sampling rate, the actual coordinate of trajectory sampling point at time t may scattered in the cope of the horizontal disk, we represent the disk as a uncertain area of T_p at time t.

The velocity of trajectory T_p at i-th sampling point can be defined as follow

$$v_p^i = \frac{\sqrt{(x_p^{i+1} - x_p^i)^2 + (y_p^{i+1} - y_p^i)^2}}{t_p^{i+1} - t_p^i}. \quad (1)$$

When traffic is very heavy, the velocity of vehicles decrease and GPS accuracy will perform well. Besides, GPS sampling error increase when vehicles driving fast on the road (especially on an expressway). Threrfore, we assume that the velocity and uncertain parameter ξ are the key factors when considering the uncertain radius δ. The uncertain radius of trajectory T_p at i-th sampling point can be denoted as follow

$$\delta_p^i = \xi \cdot v_p^i. \quad (2)$$

Note that the uncertain parameter ξ reflects the level of sampling precision. In the case of low accuracy, the uncertain parameter ξ would be great.

In this paper, we consider the uncertainty in the measure of direction similarity. The expected angle consin between i-th trajectory segments of T_p and T_q is represented by $\cos[\theta]_{pq}^i$, which is shown as follows

$$\cos[\theta]_{pq}^i = \frac{\overrightarrow{T_p^i} \cdot \overrightarrow{T_q^i}}{|\overrightarrow{T_p^i}||\overrightarrow{T_q^i}|}. \quad (3)$$

If $[\cos\theta]_{pq}^i < 0$, it means that the two trajectory segments move toward different direction. Then we set $[\cos\theta]_{pq}^i$=0, and ignore this trajectory segment in the measure of direction similarity.

Sampling point is uncertain in each uncertain area, therefore, the trajectory segment direction is uncertain. However, we can determine the angle variation range for the reason that the uncertain radius is limited by ξ and velocity. As shown in the Fig. 1, θ_2 and θ_3 are the angle variation ranges of the i-th trajectory segments of T_p and T_q, respectively. The vector $\overrightarrow{CB}$ is denoted as $(x_p^{i+1} - x_p^i + \delta_{i+1} + \delta_i, y_p^{i+1} - y_p^i)$ and vector $\overrightarrow{DA}$ is denoted as $(x_p^{i+1} - x_p^i - \delta_{i+1} - \delta_i, y_p^{i+1} - y_p^i)$. It is found that the angle between vector $\overrightarrow{CB}$ and horizontal plane reaches the minimum and the angle between vector $\overrightarrow{DA}$ and horizontal plane reaches the maximum. The angle variation range in the i-th trajectory segment of T_p can be calculated by

$$[\cos\theta_2]_p^i = \frac{\overrightarrow{CB} \cdot \overrightarrow{DA}}{|\overrightarrow{CB}||\overrightarrow{DA}|}. \quad (4)$$

In order to measure the uncertain direction similarity between the ith trajectory segments of T_p and T_q, we define the angle variation ranges difference as $\cos([\theta_2]_p^i - [\theta_3]_q^i)$. A smaller angle indicates a higher value of cosx, because cosx is monotonically decreasing function in [0,π]. Therefore, $\cos([\theta_2]_p^i - [\theta_3]_q^i)$ is near to 1 when the angle variation ranges of two trajectories segments are similar. If $\cos[\theta_2]_p^i \cdot \cos[\theta_3]_q^i < 0$, it means there exists large difference of the angle variation ranges. In this case, we set $\cos([\theta_2]_p^i - [\theta_3]_q^i) = 0$.

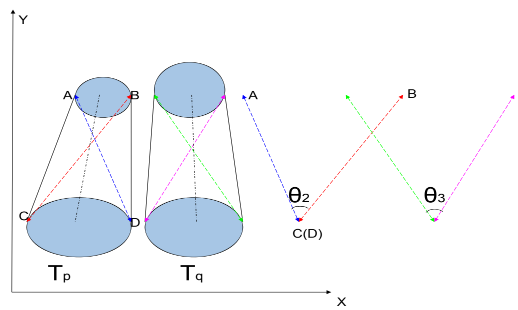

Fig. 1: Trajectory direction similarity.

In this paper, the expected angle consin $\cos[\theta]_{pq}^i$ and angle variation ranges difference $\cos([\theta_2]_p^i - [\theta_3]_q^i)$ are proposed to describe the direction similarity between two trajectories reasonably. It can be measured by III.5.

$$S_{dire}[\xi, p, q] = \frac{\sum_{i=1}^{n-1}\left(\frac{[\cos\theta]_{pq}^i + \cos([\theta_2]_p^i - [\theta_3]_q^i)}{2}\right)}{n-1}. \tag{5}$$

Let t_p^{beg} and t_q^{beg} denote the time of beginning points of T_p and T_q, respectively, t_p^{end} and t_q^{end} represent the time of ending points of T_p and T_q. Temporal overlap similarity can be measured by III.6. If the two trajectory do not share the same temporal segment, we set $S_{time}[p, q] = 0$, otherwise,

$$S_{time}[p, q] = \frac{\min(t_p^{end}, t_q^{end}) - \max(t_p^{beg}, t_q^{beg})}{\max(t_p^{end}, t_q^{end}) - \min(t_p^{beg}, t_q^{beg})}. \tag{6}$$

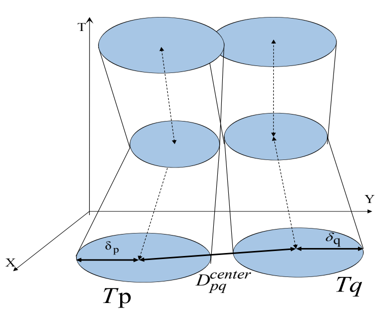

Fig. 2: The distance of uncertain area.

At time t_i, we denote D_{pq}^{center}=$dis((x_1, y_1), (x_2, y_2))$ as the distance between two points. As shown in Fig. 2, the uncertain distance of trajectories at time t_i is the the sum of radius of two uncertain areas and the center distance, which can be calculated by

$$D_{pq}^{pa}[i] = \delta_p^i + \delta_q^i + D_{pq}^{center}[i]. \tag{7}$$

In this paper, distance between two uncertain trajectory is calculated based on the Frechet Distance. We define the uncertain distance between T_p and T_q as the minimum uncertain distance of uncertain areas. The uncertain distance $D_{loc}[p, q]$ between T_p and T_q is defined as follows

$$D_{loc}[\xi, p, q] = \min\max D_{pq}^{pa}[i], (i = 1, 2, ..., n-1). \tag{8}$$

IV. Uncertain Trajectories Publishing Privacy Protection

A. Pre-processing

In order to achieve high similarity between trajectories and reduce the information loss in the trajectory anonymization, how to choose the time span, we define the notion of trajectory synchronization and construct trajectory equivalence class have been the important tasks in the trajectory pre-processing phase. Trajectories collected by users are different in time dimension because their different behaviors and life styles, therefore, it is difficult to construct trajectory equivalence class whose trajectories are in the same intervals.

Trajectories include the same period $[t_p, t_q]$ in the time dimensional form the equivalence class. The size of time interval $[t_p, t_q]$ can be adjusted flexibly according to the trajectory sparseness. In our approach, we make sure the number of trajectory sampling points are the same in the trajectory equivalence class.

B. Trajectory graph model construction

In this subsection, we consider the trajectories clustering based on graph model. The trajectory graph is defined as follows

Definition 4.1 (Trajectory graph). A trajectory graph $G(V, E, W)$ is a undirected graph where vertices represent trajectories. There exists an edge between two vertices V_p and V_q if they are in the same trajectory equivalence class.

The trajectory weight matrix, denoted as $M = (W_{pq})_{n*n}$, is a upper triangular matrix, where each element in the diagonal of the matrix equals to 0. We should consider how to minimize the information loss and maximize the privacy-preserving when we construct the trajectory k-anonymity set. Therefore, the primary preparation work in this phase is to measure the weight between two vertices appropriately.

We propose a optimal weight function in a trajectory equivalence class $C = \{T_1, T_2, T_3, ..., T_n\}$ as follows

$$\begin{aligned} W_{pq} &= \alpha \cdot (1 - S_{time}[p, q]) + \\ &\beta \cdot (1 - S_{dire}[\xi, T_p, T_q]) + \gamma \cdot D_{loc}[\xi, T_p, T_q]. \end{aligned} \tag{9}$$

The parameters satisfy $\alpha, \beta, \gamma \in [0, 1]$, and $\alpha+\beta+\gamma = 1$. The trajectory graph and weight construction algorithm, which is shown in Algorithm 1.

Algorithm 1 Trajectory Graph and Weight Construction

Input:

A trajectory equivalence class $C = \{T_1, T_2, T_3, ..., T_n\}$;
The parameters α,β,γ;
The parameters ξ to determine the radius.

Output:

Trajectory Graph G=$\{V, E, W\}$;

$len \leftarrow n; V \leftarrow C; E \leftarrow \varnothing$
for $p = 1$; $p < len$; $p++$ **do**
  for $q = p+1$; $q < len$; $q++$ **do**
    $W_{ij} \leftarrow \alpha \cdot (1 - S_{time}[T_p, T_q]) + \beta \cdot (1 - S_{dire}[\xi, T_p, T_q]) + \gamma \cdot D_{loc}[\xi, T_p, T_q]$;
    $E \leftarrow (T_p, T_q, W_{pq})$;
  end for
end for

The Algorithm 1 shows the procedure of trajectory graph and weight construction. At the beginning, the edges set E is empty, and each vertex in the vertices set V denotes one trajectory in C. The weight of edge is computed and stored by the sum of $S_{time}[T_p, T_q]$, $S_{dire}[\xi, T_p, T_q]$ and $D_{loc}[\xi, T_p, T_q]$ with appropriate proportion. Trajectory Direction Similarity and Uncertain Distance, which are shown in Algorithm 2.

Algorithm 2 Trajectory Direction Similarity and Uncertain Distance

Input:

Two trajectories T_p and T_q in equivalence class C;
Trajectory sampling points n;
Parameter ξ to determine the radius of uncertainty area.

Output:

$S_{dire}[\xi, T_p, T_q], D_{loc}[\xi, T_p, T_q]$;

$S_{dire} \leftarrow 0, D_{loc} \leftarrow 0$
$\cos[\theta_1]_{pq} \leftarrow 0, \cos[\theta_2]_p \leftarrow 0, \cos[\theta_3]_q \leftarrow 0$
for $i = 1$; $i < n$; $i++$ **do**
  compute v_p^i, v_q^i according to III.1.
  compute D_{pq}^{center} (Euclidean distance).
  if $\cos[\theta_1]_{pq}^i < 0$ **then**
    $\cos[\theta_1]_{pq}^i \leftarrow 0$
  end if
  if $\cos[\theta_2]_p^i \cdot \cos[\theta_3]_q^i < 0$ **then**
    $\cos([\theta_2]_p^i - [\theta_3]_q^i) \leftarrow 0$
  end if
  $S_{dire} \leftarrow S_{dire} + (\frac{[\cos\theta]_{pq}^i + \cos([\theta_2]_p^i - [\theta_3]_q^i)}{2})$
  $\max D_{pq}^{pa}[i] \leftarrow \delta_p^i + \delta_q^i + D_{pq}^{center}[i]$;
end for
$D_{loc}[\xi, T_p, T_q] \leftarrow \min \max D_{pq}^{pa}[i]$;
$S_{dire}[\xi, T_p, T_q] \leftarrow \frac{S_{dire}}{n-1}$;

C. Trajectory k-anonymization set construction

After the trajectory graph construction, we try to find trajectory k-anonymity sets R_i, $i = 1, 2, ..., n$. from the trajectory graph. In order to select the k-anonymity set, we adopt a greedy strategy partition the trajectory sets $R = \{R_1, R_2, ..., R_n, D_{drop}\}$. In this method, partition with k vertices are kept in R_i, while partitions with less than k vertices are dropped in D_{drop}.

At the beginning, we sort the edge set E, and search the edge with minimum weight and specify the two vertexes affiliated with the edge as the start points. The start points are put into k-anonymity set respectively. Then we search the edges that affiliated with the points in the k-anonymity set, find out the smallest weight, and put the affiliated vertex into the k-anonymity set. We redo the operation until the vertex number of anonymity set $|R_i|$ more than k. After the partition, the vertices left are not enough to construct a k-anonymity set. We put those vertices into D_{drop} to measure the information loss. We denote $E(v, R_i)$ as the set of all edges between node v and the node set from R_i.

Algorithm 3 Greedy k-node Partition

Input:

Trajectory graph G=(V,E,W);
Parameter k to construct the k-anonymity set;
Parameter MAX .

Output:

partition result $R = \{R_1, R_2, ..., R_n, D_{drop}\}$

$num \leftarrow 0$;
$R_i, D_{drop} \leftarrow \varnothing$
for $i = 1$; $i < N$; $i++$ **do**
  $num \leftarrow 0$;
  Sort Edge set E in ascending order;
  Find the edge with the smallest weight e=$(v_s, v_e.)$;
  $R_i \leftarrow v_s, v_e, num = num + 2$;
  while $num < k$ **do**
    for each vertex $v \in V \setminus \{R_1 \bigcup \cdots \bigcup R_i\}$ **do**
      Compute the sum of weights of
      edges from $E(v, R_i)$,denoted as $W(v, R_i)$.
    end for
    Find out the node v satisfy minimum $W(v, R_i)$.
    $R_i \leftarrow v, num++$;
    if $num > k$ **then**
      break;
    end if
  end while
end for
$D_{drop} \leftarrow V \setminus \{R_1 \bigcup R_2 \bigcup ... \bigcup R_n\}$

D. Privacy-preserving and Information Loss

Trajectories data publisher would publish the k-anonymity sets $\{R_1, R_2, ..., R_n\}$ while not publish the drop set D_{drop}. We suppose that the adversary has not aware of the privacy-preserving model, in this scenario, adversary is difficult to tell the victim's trajectory with probability under $1/k$.

In this paper, we evaluate the privacy level by analyzing the similarity among trajectories. Let the selected k trajectories in the k-anonymity set R_i=$\{T_1, T_2, ..., T_k\}, (i = 1, 2, ..., n)$.

The weight matrix $M = (W_{pq})_{n*n}$ denotes the correlation between trajectories T_m and T_n in trajectory equivalence class C. According to IV.1, the correlation between T_m and T_n can be divided into two parts. The sum of trajectory direction average similarity $S_{dire}[\xi, T_m, T_n]$ and temporal overlap similarity $S_{time}[T_m, T_n]$ can be denoted as S_{mn} to evaluate the similarity between trajectories, and the data utility is evaluated by uncertain distance $D_{loc}[\xi, T_m, T_n]$, which is denoted as D_{mn}.

The greater the S_{mn}, the similar between T_m and T_n. Therefore, the greater the S^i_{avg}, the more average similarities in the trajectory k-anonymity set R_i, $i = 1, 2, ..., n$; S^i_{avg} can be denoted as follows

$$S^i_{avg} = \frac{2 \cdot \sum_{m=2}^{k} \sum_{n=1}^{m-1} S_{mn}}{k \cdot (k-1)}. \quad (10)$$

The privacy level, denoted as P, is defined as follows.

$$P = \frac{\sum_{i=1}^{n} S^i_{avg}}{n} - \frac{k}{|C|}. \quad (11)$$

In our definition of privacy level, the size of k-anonymity set k and the average similarity are taken into consideration. There is a negative correlation between privacy level and the size of k-anonymity set k. In other words, the greater the average value of S^i_{avg}, the similar the trajectories in each trajectory k-anonymity sets. The privacy level will be higher when trajectories are similar with each other in each trajectory k-anonymity sets. At the same time, the size of k-anonymity may has an impact on the privacy level.

In k-anonymity sets $R_i, (i = 1, 2, ..., n)$; the maximum uncertain distance in each R_i and the proportion of dropped trajectories $\frac{|D_{drop}|}{|C|}$ are taken into consideration to measure information loss. The greater the $\frac{|D_{drop}|}{|C|}$, the higher the information loss is. In our definition, the maximum uncertain distance in each R_i is denoted as D^i_{max}.

The information loss in $R = \{R_1, R_2, ..., R_n, D_{drop}\}$, denoted as I, can be computed by

$$I = \frac{\sum_{i=1}^{n} D^i_{max}}{n} + \frac{|D_{drop}|}{|C|}. \quad (12)$$

In this paper, we consider the trajectory direction average similarity and distance between trajectories with uncertainty. The uncertain parameter ξ has an impact on the information loss and privacy level. In the simulation section, we evaluate the relationship between information loss and uncertain parameter ξ.

V. Simulation

We conduct a set of simulations to evaluate the performance of the strategies in this section. In our experiment, we use the GeoLife Trajectory Dataset, which is a GPS trajectory dataset from Microsoft Research GeoLife project [19], collected by 182 users from April 2007 to August 2012. We select a trajectory equivalence class with 60 trajectories randomly after the pre-processing phrase. Each trajectory in the equivalence class is represented as a sequence of 30 spatiotemporal sampling points.

In order to evaluate the privacy level, we set a group of different settings on three parameters, which are the proportion of the temporal overlap similarity α, the proportion of the trajectory direction similarity β, the proportion of the distance between trajectories γ. Privacy level evaluation is shown in Fig. 3 by Group 1: $(\alpha = 0.5, \beta = 0.5, \gamma = 0)$, $(\alpha = 0.45, \beta = 0.45, \gamma = 0.1)$, $(\alpha = 0.4, \beta = 0.4, \gamma = 0.2)$. It is noted from

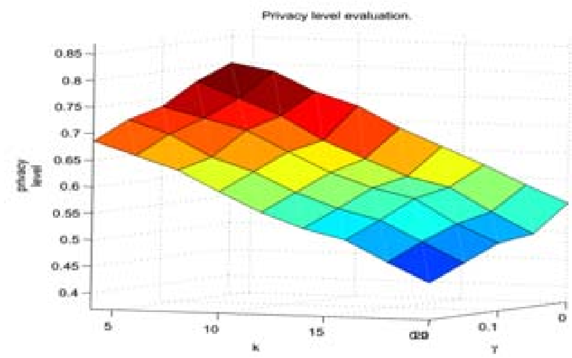

Fig. 3: Privacy level evaluation.

Fig. 3 that the privacy level decreases when the number of trajectories k increase. The reason is that with the increase of the size of trajectory k-anonymity set, it is more difficult to find more similar trajectories.

Users have different emphases on the aspect of privacy level and data utility, and Group 1 simulate different kinds of demands. γ increases with $\alpha + \beta$ decreases for the reason that $\alpha + \beta + \gamma = 1$. When $\alpha + \beta$ decreases, user pay less attention to the privacy level, then the privacy level will decrease at the same time.

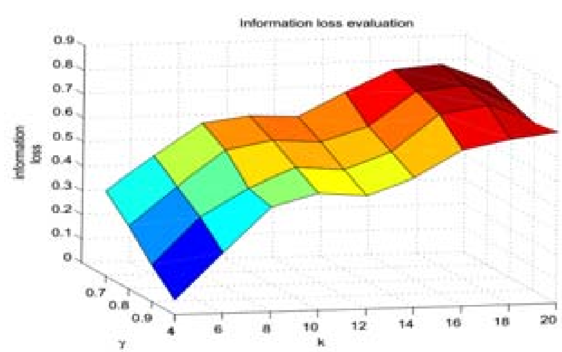

Fig. 4: Information loss evaluation.

The information loss evaluation shown in Fig. 4 by Group 2: $(\alpha = 0.15, \beta = 0.15, \gamma = 0.7)$, $(\alpha = 0.1, \beta = 0.1, \gamma = 0.8)$, $(\alpha = 0, \beta = 0, \gamma = 1)$. The information loss roughly increase when the k increase, since a larger k-anonymity set may cause bigger difference between trajectories in the k-anonymity set and the maximum uncertain distance in each R_i would increase. In some cases, the equivalence class cannot be divided into several parts neatly, the dropped trajectories have an impact on the information loss. In particular, in the case of $k = 20$, it is noted that the information loss is decrease for the reason that the equivalence class C can be divided into 3 k-anonymity sets without any dropped trajectories. The C can be divided into 3 k-anonymity sets with 12 dropped trajectories when $k = 16$, therefore the information loss is bigger than $k = 20$.

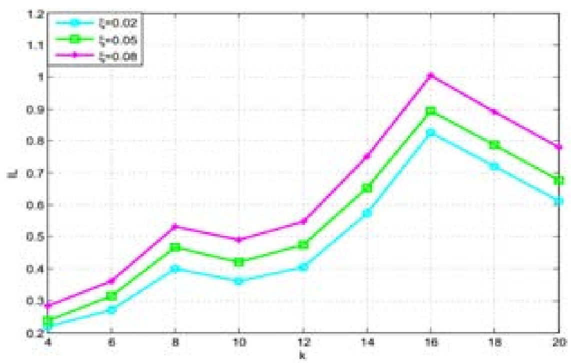

Fig. 5: Uncertain parameter ξ evaluation.

Uncertain parameter ξ evaluation is shown in Fig. 5 by Group 3: $(\xi = 0.02, \xi = 0.05, \xi = 0.08)$ is under the same condition of $(\alpha = 0.1, \beta = 0.1, \gamma = 0.8)$. There exits uncertainty since the inherent accuracy of data acquisition e-quipments, sampling error and so on. However, if the uncertain parameter ξ is too large, the partition can lead to incorrect partition results. Therefore, we evaluate the relationship between information loss and uncertain parameter ξ with appropriate ranges. As shown by Fig. 5, when ξ increase, the information loss will be more evident.

VI. Conclusion

In this paper, we measure the correlation between two trajectories according to the temporal overlap similarity, the trajectory direction similarity and the distance between trajectories with uncertainty. Then the trajectory graph model is constructed according to the correlation. Based on the trajectory graph model, a greedy algorithm is proposed to achieve k-node partition, where k-node partition denotes a trajectory k-anonymity set. We investigate the effects of the model parameters, α, β, γ, and ξ, on the information loss and the privacy level, respectively with different proportion of the temporal overlap similarity, the trajectory direction similarity and the distance between trajectories. Further studies on the performance improvement of privacy level and information loss can then be done based on the results in this paper.

Acknowledgment

The authors wish to thank National Natural Science Foundation of China (NOs. 61072080, U1405255). Fujian Normal University Innovative Research Team (NO. IRTL1207). The Natural Science Foundation of Fujian Province (NO. 2013J01222). Major science and technology project in fujian province (NO. 2014H61010105).

References

[1] Nam T, Pardo T A. Smart city as urban innovation: focusing on management, policy, and context, in Proceedings of International Conference on Theory and Practice of Electronic Governance. 2011:185-194.

[2] Z. Huo, X. Meng and H. Hu. You can walk alone : trajectory privacy-preserving through significant stays protection. Lecture Notes in Computer Science, 2012: 351-366.

[3] F. Bonchi, L. V. S. Lakshmanan and H. Wang. Trajectory anonymity in publishing personal mobility data. Acm Sigkdd Explorations Newsletter, 2011, 13: 30–42.

[4] J. Chen, P. Chen, Q. Huo and X. Xu. Clustering network-constrained uncertain trajectories, in Proceedings of Fuzzy Systems and Knowledge Discovery (FSKD), 2011, 3: 1657-1662.

[5] Y. Lou, C. Zhang, Y. Zheng, X. Xie, X, W. Wang and Y. Huang. Map-matching for low-sampling-rate GPS trajectories, in Proceedings of Acm Sigspatial International Conference on Advances in Geographic Information Systems(GIS. ACM), 2009: 352-361.

[6] Z. Huo, H. Hu, X. Meng. History Trajectory Privacy-preserving through Graph Partion, in Proceedings of the 1st Int Conf on Mobile Location-based service(MLBS'II). Beijing: ACM, China, 2011. 88-97.

[7] S. Gao, J. Ma and Cong. Sun. Balancing trajectory privacy and data utility using a personalized anonymization model. Journal of Network and Computer Applications, 2014, 38: 125-134.

[8] J. YANG, B. ZHANG, J. ZHANG and J. XIE. Personalized trajectory privacy preserving method based on graph partition. Journal on Communications, 2015, 36(3): 1-11.

[9] O. Abul, F. Bonchi and M. Nanni. Never Walk Alone: Uncertainty for Anonymity in Moving Objects Databases, in Proceedings of the 2008 IEEE 24th International Conference on Data EngineeringIEEE Computer Society, 2008: 376–385.

[10] O. Abul, F. Bonchi and M. Nanni. Anonymization of moving objects databases by clustering and perturbation. Information Systems 2010, 35(8): 884-910.

[11] F. Zhou, He. X, S. Wang, J. Xu, M. Wang and L.W. A Clustering-Based Privacy-Preserving Method for Uncertain Trajectory Data, in Proceedings of Trust, Security and Privacy in Computing and Communications (Trust-Com), 2014:1-8.

[12] L. Zhang, J. Li and Z. Wang. Uncertain Trajectory Clustering Based on Uneven Two-step Sampling. Journal of Convergence Information Technology, 2011, 6(5).

[13] G. Klir and B. Yuan. Fuzzy Sets and Fuzzy Logic: Theory and Applications. International Encyclopedia of Human Geography, 1995: 283-287.

[14] Bezdek, C. James, R. Ehrlich, and W. Full. FCM: The fuzzy c-means clustering algorithm. Computers and Geosciences10.84 (1984): 191-203.

[15] H. P. Kriegel, and M. Pfeifle. Density-based clustering of uncertain data, in Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data miningACM, 2005:672-677.

[16] Y. Chen , X. Qin and L. Liu. Uncertain Distance-Based Range Queries over Uncertain Moving Objects. Journal of Computer Sciences and technology, 2010, 25(5): 982-998.

[17] C. Wang, J. Yang and J. Zhang. Privacy preserving algorithm based on trajectory location and shape similarity. Journal of Communications, 2015, 36(2).

[18] H. Zhu, J. Luo, and H. Yin. Mining Trajectory Corridors Using Frechet Distance and Meshing Grids, in Proceedings of Advances in Knowledge Discovery and Data Mining, the 14th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2010) 2010: 228-237.

[19] Y. Zheng, X. Xie and W. Ma, GeoLife: A Collaborative Social Networking Service among User, location and trajectory. Invited paper, in IEEE Data Engineering Bulletin. 2010, 33(2): 32-40.