Achieving Absolute Privacy Preservation in Continuous Query Road Network Services

Yankson Herbert Gustav[1], Xiao Wu[2], Yan Ren[2,⋆], Yong Wang[1], and Fengli Zhang[1]

[1] School of Computer Science and Technology, University of Electronic Science and Technology of China, Chengdu, China
[2] National Computer Network Emergency Response Technical Team Coordination Center of China, Beijing, China
hgustav.yankson@gmail.com, cla@uestc.edu.cn, ry@cert.org.cn

Abstract. Research have shown that location semantics have lead to privacy leakages especially when two or more users in a cloaked region depict similar semantic locations. This implies that, to achieve absolute privacy(query privacy, location privacy and semantic location privacy) protection for a client on road network, it is important that cloaked users have their locations distinctly diverse with diverse semantics, and making diverse service request thus satisfying the k-anonymity and l-diversity conditions for privacy. Unfortunately, the determination of semantic location of a mobile user online is a challenge which makes the achievement of absolute privacy protection more challenging. In this paper, we developed a privacy preserving algorithm that protects a client's absolute privacy for continuous query road network services. We employed an offline trajectory clustering algorithm and semantic location graph to aid the selection of cloaked users that will effectively protect the absolute privacy of a client. We evaluated the effectiveness of our algorithm on a real world map with two defined metrics, and it exhibited an excellent anonymization success rate in a very good query processing time for the entire period of continuously querying road network services.

Keywords: Location-based Services (LBS), Privacy Preservation Algorithm, Trajectory Clustering Algorithm, Semantic Location Privacy.

1 Introduction

Location based queries have become very common mobile applications because of the convenience that it provide its users. Despite the convenience in its usage, it threatens the privacy of users. The threat of privacy results from the fact that, the disclosed user location for service can be combined with other kinds of data to allow an adversary make unwanted inferences on the activity of the user at some given time in the past [1].

⋆ Corresponding author.

X. Luo, J.X. Yu, and Z. Li (Eds.): ADMA 2014, LNAI 8933, pp. 279–292, 2014.

There are two types of privacy issues namely location privacy and query privacy. Location privacy is related to the disclosure of exact locations that a user has visited, whereas query privacy is related to the disclosure of sensitive information in the query itself and its association to the user. An adversary having knowledge of these two, can reveal places of frequent visit and personal preferences of the user, hence there is the need to protect it [2]. In recent times, location semantics have also shown to leak privacy especially when two or more users in a cloaked region depict similarity in semantic locations [3]. Therefore, to achieve absolute privacy protection(query privacy, location privacy and semantic location privacy) for a client on a road network, it is important that cloaked users must have their locations distinctly diverse with diverse semantics, and making diverse service requests thus satisfying the k-anonymity and l-diversity conditions for privacy. However, determining the semantic location of a mobile user online is a challenge, which makes the achievement of absolute privacy protection more challenging.

The random segment sampling and network expansion cloaking methods are the two known cloaking methods in literature for road networks. The network expansion cloaking method blurs a user's location into a cloaked set of k-users from connected road segments S such that S satisfies client's privacy. Whereas the random segment sampling model blurs a user's location into a cloaked set of k-users from randomly selected road segments [4]. Employing the network expansion cloaking methods to obtain absolute privacy protection may not be achievable, as all cloaked users may assume similar semantics because of the short distances amongst cloaked users. For example, cloaking a client with k-1 other users in a big university campus with such a cloaking technique will have all k cloaked users' location depicting the university as its semantics and hence will leak their privacy. Employing the random segment sampling method seems a better option because of its random segment cloaking, however, the inability to determine the semantics of users online makes it a challenge. We believe a mechanism where the query, location and its semantic of an online mobile user could be determined offline would make the random segment sampling model appropriate. In that way, the characteristics of users that protect the absolute privacy of clients could be determined offline before going ahead to cloak it online.

In this paper, we develop a privacy preserving algorithm that protects the absolute privacy of clients for continuous query road network services. Our contributions in this work are as follows:

1. Develop an algorithm that cluster trajectories of mobile users offline and use the derived movement trends online to aid the selection of cloaked users that will enhance the absolute privacy protection of clients.
2. Develop an algorithm to generate a sematic location graph to aid the determination of users' semantic locations online.
3. Develop a privacy preservation algorithm to protect the absolute privacy of clients continuously querying road network services, using the derived

movement trends from the offline trajectory clustering algorithm and the semantic location graph.
4. We introduce two metrics namely Anonymisation Success Rate and Query Processing Time to test the efficiency of our algorithm on a real world map.

The rest of the paper is organized as follows; Section 2 discusses related work on privacy preservation in LBS. Section 3 discusses the preliminaries including designing goals, system architecture, and the road network model. Section 4 discusses the development of the trajectory clustering algorithm and the semantic location graph. Section 5 discusses the privacy preserving algorithm and its security. Sections 6 presents the experiment, and conclude in section 7.

2 Related Work

In this section, we will categories the related work into privacy preservation in euclidean space and those constrained by the underlying road network.

2.1 Privacy Preservation in Euclidean Space

To protect the semantic location of clients, Damiani et al [5] employed the semantic location cloaking method which allows users to define a personalised privacy profile stating specified sensitive place types and the desired degree of privacy for each type. Unfortunately, semantic location cloaking method have been designed to work only in unconstrained space in which users can move without restrictions, but in real world setting, movement is confined to road network and may therefore lead to privacy leakages. B.Lee et al [3] proposed location privacy protection technique, which protects the location semantics from an adversary. They employed a trusted anonymizing server that uses the location semantic information for cloaking users with semantically heterogeneous locations. Similarly, their work considered euclidean space that will lead to privacy leakages under road network restrictions. C. Chow et al [6] proposed a spatial cloaking technique for snapshot and continuous location-based queries that clearly distinguishes between location privacy and query privacy using k-sharing region and memorization properties. They adopted a minimum area A_{min} within which a client wants to be anonymised. However, their work employed a cloaking technique that did not consider semantic location of its cloaked users hence are likely to include users with same semantics leading to privacy leakages.

2.2 Privacy Preservation on Road Networks

To protect privacy on road networks, Li et al [7] proposed a personality privacy-preserving cloaking framework for the protection of sensitive positions on road network environment. In their client-server architecture scheme, a user expressed his privacy requirements by specifying some types of sensitive semantics and used popularity ratio of those places to measure the degree of semantic diversity.

Yigitoglu et al [8] presented an extension of the semantic location cloaking model for location sharing under road-network constraints, that relies on the trusted anonymizer. Our work is different because we did not consider some places as sensitive but rather all locations of clients as sensitive.

The network expansion method is known to be less attack resilient [4], and therefore not surprising that [4], [9], and [10] employed a modified version of network expansion cloaking method to make it attack resilient. However, their works may lead to privacy leakages when considering semantic locations due to their closeness of segments which will likely make all cloaked users depict same semantics. We intend to use a more attack resilient random segment sampling model employing the semantic location graph and trajectory clustering algorithm to aid the selection of cloaked users that protect the absolute privacy of clients.

In other related research work, Binh Han et al [11] proposed a frame work called NEAT, a road network aware algorithm for fast and effective clustering of spatial trajectories of mobile objects travelling on road networks which takes into account the physical constraints of the road network, network proximity and the traffic flows among consecutive road segments.

3 Preliminaries

3.1 Assumptions and Architectural Systems

We adopt a trusted third party architecture consisting of a mobile client(MC), Anonymous Server(AS) and location based server(LBS) [12]. We assume that the anonymous server has been supplied with the initial trajectory and service request (query content) database of users by the cellular service provider. The location and service request database may be built from clients' regular phone call and query of LBS. If such an initial database does not exist, we assume a location and service request sample collection phase by the anonymous server. The sampling location and service request collection phase should last for some few days. More location data will be obtained from mobile users during their requests for LBS. We also assume that MC is allowed a privacy profile k, being the number of users he would want to be anonymized with.

The core of the system is the AS which consist of the Cloaked Repository (CR) and an offline Trajectory Clustering Engine (TCE). Their functions can be defined as follow:

1. The cloaked repository keeps some previously cloaked results and use them to generate new cloaked regions.
2. The Trajectory clustering Engine performs the clustering of a database of mobile users' trajectories

The architecture is as shown in Fig. 1.

3.2 Road Network Model

A road network is represented by a single directed graph G = (V, E), composed of the junction nodes V = (n_0, n_1. . . n_n) and directed edges E = (s_{id},

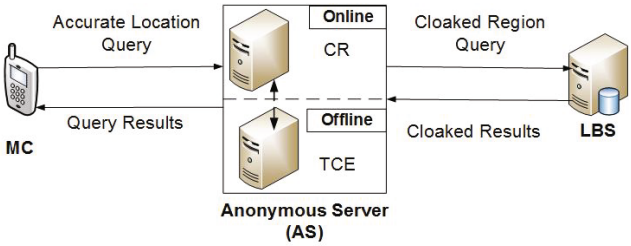

Fig. 1. The System Architecture

$n_in_j)|n_i,n_j \in V$). An edge e = (s_{id}, w_0, w_1, *con*, n_in_j) $\in$ E representing a road segment connecting two junctions n_i and n_j in the real road network with attributes such as the segment identifier s_{id}, the segment classification w_0, the traffic density w_1, and types of service request *con* (s_{r1},s_{r2}...s_{rn} $\in$*con* where each s_r is a service request). The order n_in_j indicates the direction from n_i to n_j of the road segment. For a bi-directional road segment s_{id}, we use edge e = (s_{id}, n_in_j) and e' = (s_{id}, n_jn_i) to denote the bi-directional lanes with road segment identifier s_{id}. The length of a road segment e = (s_{id}, n_in_j) is denoted by length $\lvert n_in_j \rvert$. We classify the segments according to their speed limits. The segments are classified as primary (speed limit <40km/hr), auxiliary (40km/hr$\leq$ speed limit <70km/hr), highway (70km/hr$\leq$ speed limit <100km/hr), express way (speed limit$\geq$100km/hr) denoted by p, a, h, ex respectively. Therefore segment classification w_0 could be represented by p, a, h, or ex. For example w_o = ex represents express road classification. We denote the position of a user on a road segment s_{id} with coordinates (x, y) at a time stamp t by l=(s_{id}, x, y, t).

Definition1 Trajectory: A trajectory denoted by TR = {t_{id}, l_0, l_1....l_n}, is a time-ordered sequence of locations l_0, l_1, ..., l_n of a user on the road network over time and uniquely identified by a trajectory identifier t_{id}. For a mobile user, his/her trips with beginning location and destination location forms a trajectory.

Definition2 Trajectory-fragment: A trajectory-fragment of TR, denoted by tf = {t_{id}, s_{id}, l_kl_{k+m}}, represents a sub-trajectory l_k, l_{k+1}...l_{k+m} consisting of m + 1 consecutive points extracted from TR which lie on the same road segment s_{id}.

Definition3. Base cluster: A base cluster b with respect to a segment is a group of distinct trajectory-fragments with similar characteristics. Each of these trajectory-fragments belongs to a distinct trajectory TR and is associated with a segment s_{id}. A group of base clusters in a segment s_{id} is called a segment cluster $c_{s_{id}}$.

Definition4. Class cluster: For a given set of trajectories T = {TR_1... TR_n} on a road network, a class cluster c_{w_0} is a set of all segment clusters $c_{s_{id}}$ in all segments with similar segment classification w_0. Therefore, class clusters are c_p, c_a, c_h, c_{ex} where each class cluster could be represented generally as c_{w_0}.

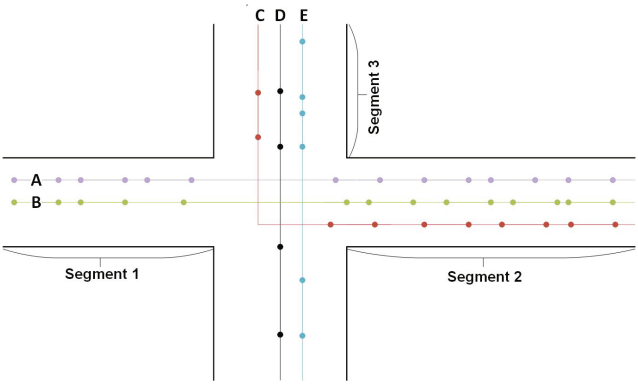

Fig. 2. A section of road network depicting trajectories A,B,C,D,E broken into trajectory fragments along segments

Definition5. Cluster : For a given set of trajectories T = $\{TR_1 ... TR_n\}$ on a road network, cluster C is the set of all class clusters ie C = $\{c_p, c_a, c_h, c_{ex}\}$. C therefore represents all class clusters with segment classifications w_0.

4 Developing Trajectory Clustering Algorithm and Semantic Location Graph

4.1 Trajectory Clustering Algorithm

One technique that is widely known to group objects of a database into a set of meaningful subclasses is the clustering algorithm [11]. We intend to use this tool to cluster trajectories of users into subclasses according to their similar movement characteristics on same road segment offline. We believe users within a segment can be considered close in terms of network proximity and will therefore display a group of subclass characteristics in their movements which will reflect that of any online mobile object on that segment. We therefore can estimate from the offline characteristics if its inclusion will protect privacy of the client before going ahead to cloak a user with such characteristics online. We intend to cluster trajectories of mobile users offline into subclasses along a road segment according to segment id, service request, direction of flow, speed, temporal details, length of the segment and it's speed limits. We use these properties because we believe they influence movement trends of a mobile user on a road segment.

With this prior information, we intend to develop a trajectory clustering algorithm on users' trajectories. Consider a set of trajectories denoted by T = $\{TR_1$, $TR_2 ... TR_k\}$ in the TCE, we examine a single trajectory $TR_i = \{t_{id}, l_0, l_1 ... l_n\}$ from the first location l_0 to the last location l_n. We take every two consecutive points in the trajectory, say l_i and l_{i+1}, and check to obtain the road junction node that intersects two road segments using the map-matching approach [13].

Next, we insert the obtained junction node(s) as new points in between l_i and l_{i+1} in the trajectory being examined. After examining every point in a given trajectory TR_i, the sequence of junction nodes added to TR_i will serve as the trajectory splitting points used to partition the trajectory into trajectory fragments (tf) along segments. For example, in Fig.2 trajectory C has a two trajectory fragments broken along segment 2 and segment 3. The period of time travelled by the individual trajectory fragments are analyzed and categorized into different time periods of the day(t_b). This procedure is repeated for all set of trajectories in T.

We group the trajectory-fragments by their road segment id, direction of flow, speed (v), segment length, time period of travel(t_b) and speed limits v_l on that segment to form a collection of base clusters. We then compute the resulting traffic density w_1 on each base cluster in a segment as the number of trajectory fragments. The density gives the likely number of users on that segment with a particular characteristics at a given time period of the day. For example in Fig. 2, if all trajectory fragments have similar characteristics then segment2 will have a traffic density of 3, whiles segment1 has a density of 2. A group of base clusters under the same segment id forms a segment cluster $c_{s_{id}}$. The algorithm output a base cluster according to segment id as b= (s_{id}, w_0, s_r, v, w_1, t_b, n_in_j), being the group of users characteristics on the segment at the time t_b. Finally, we abstract all segments clusters under similar road classifications w_0 into class clusters denoted as c_p, c_a, c_h, c_{ex} $\in$ C. We abstract according to segment classification because we believe it will aid cloaking of users with diverse location semantics. A group of class clusters form a cluster C. The detailed description is as shown in Algorithm 1.

Algorithm 1 Trajectory Clustering Algorithm

Input: $< GraphG >$, <T = TR_1, TR_2... TR_n> ,<$TR_i = t_{id}$, l_0 l_1...l_n>
Output: $< clusterC = c_p, c_a, c_h, c_{ex} >$,
1: Let G composed of the junction nodes V(n_i,n_j $\in$ V) and directed edges E (e $\in$ E);
2: **for** TR_i,i=1...n **do**
3: examine every l_i and l_{i+1} to obtain the sequence of road junction nodes n_i
4: insert the obtained junction node(s) as new points in between l_i and l_{i+1}
5: use new junction node to break TR into tf
6: Assign identities s_{id} for each e
7: categorise t_b of all tf
8: group all tf along e according to s_{id}, v, v_l, t_b, n_in_j and ı n_in_j ı into b
9: **end for**
10: Evaluate w_1 in each b
11: group all b in each e as $c_{s_{id}}$;
12: group all $c_{s_{id}}$ with same w_0 as c_{w_0} ;
13: output all c_{w_0} as C;

4.2 Semantic Location Graph

In this subsection, we discuss an algorithm to build a semantic location graph to aid the selection of users for cloaking in order to protect semantic location privacy. We define semantic location privacy as the disclosure of the semantics

associated with a location visited by a user. The essence of protecting semantic location privacy is to avoid a user's activities at that location from being inferred by an adversary. For example, if a location visited by a user is a cancer hospital, it may be inferred from its semantics that a user may be a cancer patient. Therefore, to cloak a location into a region such that it becomes anonymous, it's important that the semantic locations of cloaked users are made l-diverse to avoid privacy leakages. To aid the achievement of l-diversity in semantic locations, we need to build a semantic location graph.

To build such a graph, we need to determine what constitutes a semantic location. We define semantic location as a location where a kind of service is provided and therefore visited by many people who stay for a period of time. People visit locations mostly with a reason. We go to restaurants to have food, schools to attend classes, and hospitals to see a doctor. Since we have reasons for a visit, we stay for a while at a location for those reasons. Moreover, we spend different amount of time based on these reasons from which an adversary is able to link a location to a user's activity. Using the duration of stay and the service provided at the location as factors, different point of interest (POI) representing different semantic locations can therefore labelled [3]. Unfortunately labelling is not the focus of this research, so we assume point of interest (POI) collections and their location coordinates which are open source information. [14].

Let's consider a set of POI locations SL=(L_1, L_2, L_3....L_n) where each L represents a semantic location. We categorise the semantic locations according to their similarity of service provided at that location denoted SL_u=(L_1,... L_m), where u is a category of service provided at L_1,.. L_m. For example, we group clinic, health post, hospitals, dental clinics, etc under the category health, where health is category of service provided. We do the categorisation to help us avoid cloaking user locations with similar service to enhance l-diversity in their semantic locations. Using the set of semantic locations SL, we employ the map matching approach to locate their exact positions on various segments on our road network model. We then generate a sematic location graph G' depicting various semantic locations and their services provided. The algorithm is as shown below;

Algorithm 2 Semantic location Graph Algorithm

Input: $< GraphG >$, <SL = (L_1, L_2, L_3.....L_n)>
Output: $< SemanticLocationGrapgG' >$ and $< SL_u >$
1: Let G composed of the junction nodes V(n_i,n_j ∈ V) and directed edges E (e ∈ E);
2: **for** L_i,i=1...n **do**
3: label each L_i according to service provided
4: group all L_i with same u
5: output each group as SL_u
6: **end for**
7: **for** L_i,i=1...n **do**
8: insert in G
9: output G'
10: **end for**

5 Privacy Preserving Algorithm and Security Analysis

5.1 Privacy Preserving Algorithm

In this subsection, we develop the privacy preserving algorithm using trajectory clustering algorithm and the semantic location graph. The mobile client(MC) sends a new query q in the form q=(q_{id}, l, t_i, t_f, k, s_r), where l is the location coordinates, t_i is the query initiation time and t_f is the expiration time. The service request is s_r, privacy profile k, and q_{id} is the client id. On receiving a new query q, AS determines the time of q, and use the location to find the segment from which it was issued, classification of the segment, semantic location L associated with q from G' and the category of service provided SL_u to which it belongs.

We define a traffic density threshold σ, below which it will not be appropriate for online cloaking to be executed. $w_1 \geq \sigma$ implies we are sure to find enough mobile users on that segment at the time period t_b. Traffic density threshold σ is determined by AS based on history.

To find users online to anonymise MC, TCE searches through all class clusters c_{w_0} other than that containing MC, to find the k-1 other segments at random and select a base cluster each with time t_b that satisfies the time of q from the selected segments. The selected base clusters must satisfy the cloaking conditions designed to aid the achievement of absolute privacy protection employing the random segment sampling method.

Cloaking Conditions:- All selected base clusters must satisfy;

1. k-1 in number,
2. The time period $t_{b_{k-1}}$ of the $(k-1)^{th}$ selected base cluster b_{k-1} must satisfy the time t of the query q and the time t_{b_1} of first selected base cluster b_1, ie $t \in t_{b_1}$; $t \in t_{b_{k-1}}$.
3. The traffic density of any selected base cluster $w_1(b) \geq \sigma$ ie $w_1(b_1) \geq \sigma$, $w_1(b_{k-1}) \geq \sigma$.
4. The service request $s_r(b_{k-1})$ of the $(k-1)^{th}$ selected base cluster b_{k-1} must not be the same as that of q and that of the first selected base cluster b_1, ie $s_r(b_1) \neq s_r(b_{k-1}) \neq s_r(q)$.
5. The category of service provided $SL_u(b_{k-1})$ of the semantic location L on the selected segment of the $(k-1)^{th}$ selected base cluster should not be in the same category as q and that of the first selected base cluster b_1 , ie $SL_u(b_1) \neq SL_u(b_{k-1}) \neq SL_u(q)$,
6. The segment classification $c_{w_0}(b_{k-1})$ of $(k-1)^{th}$ selected base cluster should not be the same as that of q and the first selected base cluster b_1, ie $c_{w_0}(b_{k-1}) \neq c_{w_0}(b_1) \neq c_{w_0}(q)$.

The cloaking conditions ensures that we are likely to find k-1 mobile users online with k-1 different locations associated with k-1 different semantics from k-1 different segments requesting k-1 different service request at the time of the query. When the cloaking conditions are met, AS cloaks k-1 other online users q' with the characteristic of the k-1 selected base clusters from their respective

segments into a cloaked region with q. All queries not meeting these conditions are suppressed. Client's id is removed and replaced with quasi-id and put into a cloaked region R_i where i represent the i^{th} snapshot with cloaked region identity R_{id}. The cloaked region R_i containing k users is then forwarded to the LBS. The LBS provides the result for all k users and forwards it to AS. AS knowing the exact location of MC and his request submits the appropriate result to MC.

For continuous query LBS, a query will continuously be issued periodically by AS within the period (t_f - t_i). AS cloaks users requesting continuous service for the first snapshot if client is requesting continuous service so as to maintain same cloaked users at t_i throughout the query period, whiles ensuring cloaking conditions at all times. Further, we keep a repository of already cloaked users request to use at a later time when its related to the same segment. The number of cloaked sets that meets cloaking conditions is denoted by n. The privacy preserving algorithm is as described in algorithm 3.

Algorithm 3 Privacy Preserving Algorithm

Input: <queryq = q_{id}, l, k, t_i, t_f, s_r>, < $classcluster c_{w_0} = (c_p, c_a, c_h, c_{ex})$ >, < G' >, < SL_u >,
Output: < $cloakedregion R$ >
1: for q issued at t=t_i =i;
2: determine t_i of q
3: identify e, c_{w_o}, L and SL_u of q
4: randomly output a base cluster (b_1) in e of $c_{w_0}(b_1) \neq c_{w_0}$(q) and $t_i \in t_{b_1}$
5: ensure $w_1(b_1) \geq \sigma$, $SL_u(b_1) \neq SL_u$(q) and $s_r(b_1) \neq s_r$(q)
6: goto line 11
7: **for** $k > 2$ **do**
8: ensure $w_1(b_{k-1}) \geq \sigma$; $c_{w_0}(b_1) \neq c_{w_0}(q)\neq c_{w_0}(b_{k-1})$; $s_r(b_1) \neq s_r(b_{k-1}) \neq s_r(q)$; $SL_u(b_1) \neq SL_u(b_{k-1}) \neq SL_u$(q)
9: **end for**
10: **if** line 8 is satisfied **then**
11: select q' online with characteristics of b_1 to b_{k-1} at their respective e
12: cloak q with k-1 other q' into cloaked region R_i
13: replace q_{id} with quasi id
14: assign region identity R_{id}
15: **else**
16: suppress query
17: **end if**
18: forward R_i to LBS
19: **for** t>t_i=i **do**
20: Repeat 2-18
21: **end for**

5.2 Security Analysis

An adversary with the exact position of users, sample of cloak set of some snapshot on road segments, knowledge of some sample query contents and cloaking algorithm may be able to launch query tracking attack [6], query homogeneity attacks [15], query semantic homogeneity attacks and location similarity attack [3]. We employed the randomly selected segment cloaking technique which is attack resilient to enhance the security of our algorithm.

Let us consider a scenario, in which all users from a cloaked region are requesting for the same type of service such as the location of a special club.

In this case, even if an adversary cannot link an individual query to a specific user, it is still known to the adversary that all the users in the cloaked region including the client are interested in that special club. This kind of attack is referred to as query homogeneity attack. To avoid query homogeneity attack, we avoided cloaking together any two users requesting the same service such that the probability of linking a request to a client is $\frac{1}{k}$.

Assuming cloaked users in a region show a primary school, high school, and a university as their semantic locations, an adversary may be able to infer from the semantics that the users are likely students and so is the query client. This attack is called query semantic homogeneity attack. If the cloaked users locations depict a big university campus, an adversary may be able to infer from their locations that they may be students of that university from their location similarity. This attack is termed location similarity attacks. To overcome query semantic homogeneity attacks, we avoided cloaking users with similar semantic locations and with similar services provided at that location whiles for location similarity attacks we cloaked users at diverse location from different road segment with different classifications such that linking a client to its semantics or location is $\frac{1}{k}$.

Query tracking attacks are attacks in which an adversary aggregates different snapshots of continuous query and takes an intersection of all snapshots such that the user that appears across all regions is identified as the query client. Alternatively, an adversary could model the mobility of any user from a cloaked region at t_i to a cloaked region at t_f as a Markov chain [16] where each state of the Markov chain represents a cloaked region. The m-step transition probability of a user can be defined as the probability that a user at an initial cloaked region t_i will be at a final region t_f across m cloaked regions, which can be expressed as;

$$P^m(t_i, t_f) = P(X_m = t_f \mid X_0 = t_i) \tag{1}$$

The adversary evaluates the transition probability for all users in the cloaked regions, and the user that appear across all regions will have $P^m(t_i, t_f)$ =1 indicating its the query client otherwise its not.

To overcome query tracking attacks, we ensure that all users in a cloaked region for the first snapshot are querying continuously. This is realistic since our trajectory clustering algorithm gives us a fair idea of the segment in which to find such a user. However, different users query for different time periods, so where the time for the query client is greater than the other cloaked users, we reuse repository query for the same segment where necessary to protect the privacy of the query client at all times. Where the time for the query client is less than the other cloaked users, the query client will then be fully protected across its query period. In this way, aggregating all snapshots and taking intersection will lead to the same number of users appearance in all cloaked regions thus making the probability of identifying the query client $\frac{1}{k}$. An adversary using equation (1) will also have $P^m(t_i, t_f)$ =1 for all cloaked users since they will all appear at all cloaked regions hence making it difficult to identify the query client. An adversary having the algorithm is not anticipated because AS is trusted.

6 Experiment and Evaluation

6.1 Evaluation Criteria and Metrics

Success Rate. Anonymisation Success Rate (S) measures the ability of our algorithm to avoid suppression of a query. We evaluate this metric as the ratio of the number of successfully cloaked snapshots n to the total number of successful and unsuccessful cloaked regions C_{total} within an active query period.

$$SuccessRate(S) = \frac{n}{C_{total}} \times 100 \tag{2}$$

Query Processing Time. Query Processing Time is the time required by the algorithm to find the k-1 other users. The average cloaking time T_{avg} for continuous query that has just elapse its active period consisting of n snapshots can be evaluated as;

$$T_{avg} = \frac{\Sigma_{i=1}^{n} T_{R_i}}{n} \tag{3}$$

where T_{R_i} is the cloaking time of the query with region R_i.

6.2 Experimental Setup

Using the Thomas Brinkhoff Network-based Generator of Moving Object [17], we generated 10000 mobile users moving along the map of shanghai with varied speeds and assigned varied service requests. All road segments were assigned speed limits according to our classifications and with segment id. We employed 20148 POI GPS data set consisting of 50 different categories of service provided in Shanghai city obtained from GPS Data Team [18] as our semantic locations. We recorded 100 snapshots at an interval of 5seconds. With the simulated data, we implemented our algorithms using a laptop with 6GB memory and a Core (TM) i3-2330M 2.20 GHz Intel processor.

6.3 Discussions

We studied the effects of our defined metrics with some snapshot queries for a cloaked region of k-users and l-diverse segment ($k - l$) values. From Fig.3, there was a high processing time at the first snapshot which is due to lots of time required for processing to meet the initial cloaking conditions. The querying process time decreased sharply from the initial snapshot until the first ten snapshots. The decrease may be the results of the cloak users requesting continuous service at the initial query hence there was less time required because it involved same users. There was a slight increase in the processing time from the 10^{th} to 20^{th} snapshot which might be due to users changing segments and therefore some processing was required. Thereafter, there was a steady decrease in processing time. This trend might be due to the introduction of repository

queries hence a reduction in the processing time. Generally, the query processing time decreased with increase in snapshots even when k-l values was increased.

From Fig.4, the anonymization success rate remained almost constant for the first ten snapshots which is because users cloaked at t_i were querying continuously hence most of snapshots met the cloaking principles. There was a sharp decrease in success rate thereafter until the 20^{th} snapshot which may be due to mobile objects changing segments with different classification and different semantic locations hence most of the snapshots could not meet the cloaking conditions. There was a steady increase in success rate thereafter which is due to the gradual introduction of repository queries hence most queries met cloaking conditions. When k-l was increased, similar trends was exhibited. Generally, our algorithm exhibited an average success rate of about 87.8 percent per snapshot within the 100 snapshots evaluated.

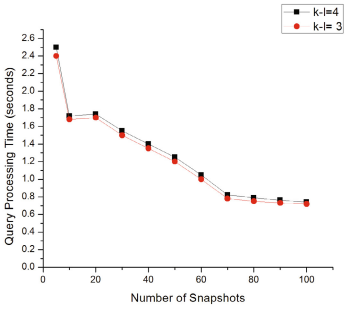

Fig. 3. Graph showing a measure of Query Processing Time

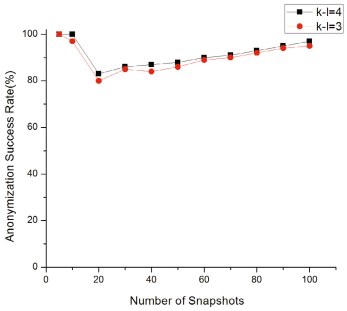

Fig. 4. Graph showing a measure of Anonymization Success Rate

7 Conclusion

In this paper, we developed a privacy preserving algorithm that protects a client's absolute privacy for continuous query road network services. We employed an offline trajectory clustering algorithm and semantic location graph to aid the selection of cloaked users that will effectively protect the absolute privacy of a client. We evaluated the effectiveness of our algorithm on a real world map with two defined metrics, and it exhibited an excellent anonymization success rate in a very good query processing time for the entire period of continuously querying road network services.

References

1. Silvestri, C., Yigitoglu, E., Damiani, M.L., Abul, O.: Sawlnet: Sensitivity aware location cloaking on road-networks. In: 2012 IEEE 13th International Conference on Mobile Data Management (MDM), pp. 336–339. IEEE (2012)

2. Dewri, R., Ray, I., Whitley, D.: Query m-invariance: Preventing query disclosures in continuous location-based services. In: 2010 Eleventh International Conference on Mobile Data Management (MDM), pp. 95–104. IEEE (2010)
3. Lee, B., Oh, J., Yu, H., Kim, J.: Protecting location privacy using location semantics. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1289–1297. ACM (2011)
4. Wang, T., Liu, L.: Privacy-aware mobile services over road networks. Proceedings of the VLDB Endowment 2, 1042–1053 (2009)
5. Damiani, M.L., Silvestri, C., Bertino, E.: Fine-grained cloaking of sensitive positions in location-sharing applications. IEEE Pervasive Computing 10, 64–72 (2011)
6. Chow, C.-Y., Mokbel, M.F.: Enabling private continuous queries for revealed user locations. In: Papadias, D., Zhang, D., Kollios, G. (eds.) SSTD 2007. LNCS, vol. 4605, pp. 258–275. Springer, Heidelberg (2007)
7. Li, M., Qin, Z., Wang, C.: Sensitive semantics-aware personality cloaking on road-network environment. International Journal of Security & Its Applications 8 (2014)
8. Yigitoglu, E., Damiani, M.L., Abul, O., Silvestri, C.: Privacy-preserving sharing of sensitive semantic locations under road-network constraints. In: 2012 IEEE 13th International Conference on Mobile Data Management (MDM), pp. 186–195. IEEE (2012)
9. Hossain, A., Hossain, A.A., Chang, J.W.: Spatial cloaking method based on reciprocity property for users' privacy in road networks. In: 2011 IEEE 11th International Conference on Computer and Information Technology (CIT), pp. 487–490. IEEE (2011)
10. Chow, C.Y., Mokbel, M.F., Bao, J., Liu, X.: Query-aware location anonymization for road networks. GeoInformatica 15, 571–607 (2011)
11. Han, B., Liu, L., Omiecinski, E.: Neat: Road network aware trajectory clustering. In: 2012 IEEE 32nd International Conference on Distributed Computing Systems (ICDCS), pp. 142–151. IEEE (2012)
12. Wang, Y., He, L.P., Peng, J., Zhang, T.T., Li, H.Z.: Privacy preserving for continuous query in location based services. In: Proceedings of the 2012 IEEE 18th International Conference on Parallel and Distributed Systems, pp. 213–220. IEEE Computer Society (2012)
13. Weber, M., Liu, L., Jones, K., Covington, M.J., Nachman, L., Pesti, P.: On map matching of wireless positioning data: a selective look-ahead approach. In: Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 290–299. ACM (2010)
14. Haklay, M., Weber, P.: Openstreetmap: User-generated street maps. IEEE Pervasive Computing 7, 12–18 (2008)
15. Liu, F., Hua, K.A., Cai, Y.: Query l-diversity in location-based services. In: Tenth International Conference on Mobile Data Management: Systems, Services and Middleware MDM 2009, pp. 436–442. IEEE (2009)
16. Papoulis, A., Pillai, S.U.: Probability, random variables, and stochastic processes. Tata McGraw-Hill Education (2002)
17. Brinkhoff, T.: A framework for generating network-based moving objects (2008), `http://iapg.jade-hs.de/personen/brinkhoff/generator/`
18. Gps data team website, `http://www.gps-data-team.com/`