Global Privacy and Transportation Mode Homogeneity Anonymization in Location Based Mobile Systems with Continuous Queries

Leon Stenneth
Dept. of Computer Science
University of Illinois at Chicago
Chicago, USA
lstenn2@uic.edu

Phillip S. Yu
Dept. Of Computer Science
University of Illinois at Chicago
Chicago, USA
psyu@cs.uic.edu

Abstract— **A major concern for deployment of location-based mobile systems is the ill-usage of mobile client's location data, which may imply sensitive and private personal information. Also, even if the location is exposed willingly by the mobile client the query should not be linked to the mobile client. Still, many location based systems (store finders, transit itinerary systems, and social networks) are created with a different focus and have little concern for end user privacy. We focused on location based mobile systems where the location of the mobile user may be available; however, an adversary should not be able to link a query to a specific mobile user. Two key contributions of this work are the introduction and experimental evaluation of a novel concept called *transportation mode homogeneity anonymization* that adds another dimension to privacy in mobile location based systems. Also, a novel dynamic layered approach on achieving *K-anonymity* by separating the local privacy requirement on each snapshot and global privacy requirement across snapshots with different privacy goals is proposed to exploit the local privacy anonymization group as candidates to obtain global anonymization group candidates.**

Keywords: GPS, K-anonymity, location based systems (LBS), mobility, privacy, transportation mode homogeneity.

1. INTRODUCTION

Location based systems are becoming more prevalent due to the advancement of mobile devices to include GPS hardware and mapping software. Examples of location based services are TransitGenie [23], Google Maps [25] and NextBus [24]. Several work previously considered guaranteeing the mobile user with a high level of privacy in conjunction with good of quality of service (QoS). QoS in location based systems refers to spatial or temporal variations with the results that are returned by the service provider [12].

We clearly distinguish between *location privacy* [1] and *query linking privacy* [2]. Location privacy ensures that the mobile client's location is not exposed. Query linking privacy ensures that an adversary cannot determine the sender of the query even if the location of the mobile user is known by adversaries.

Consider the following query from Mary, *"Where is the closest Bank of America from my current location"* to some location based service provider from her mobile device.

This work is supported in part by NSF through grants IIS-0914934, IIS-0905215, OIA-0963278 and DBI-0960443.

Location privacy ensures that the adversary cannot determine the exact location such as in a church or at the White House. Knowledge of location may reveal a person's political, religious or health connections. Knowledge of location may also lead to flooding of unwanted advertisements. Several media reports are available where GPS devices are used for stalking individuals [3,4]. Query linking privacy ensures that adversaries cannot determine the sender of the query regardless of the fact that query sender's location is known by the adversary and can be used as a quasi identifier [5]. In this case, adversaries should not learn that Mary is heading to the bank even if the adversaries are aware of her location.

To protect location privacy a mobile client's location point is cloaked into a region, in this way the adversary may not be aware of where in the region the mobile client is located. To protect from query linking the concept used is *K-anonymity* [19], which means that a mobile client is indistinguishable from *K-1* other mobile clients in the region [1]. Therefore, the query cannot be linked to a specific client in the region.

In this paper, we introduce the novel concept of *transportation mode homogeneity* for mobile users submitting a location based request. This implies that at a mobile client is aggregated with at least *K-1* other mobile clients traveling by the same mode of transportation.

We now distinguish between *snapshot queries* and *continuous queries*. A snapshot query is a "one time" request such as *"Where is my nearest bus stop"*. A continuous query is sequence of snapshot queries at discrete time points. For example, "*Continuously send me information on buses that are within five minutes from my current location?"*. Aggregating different continuous snapshots of a mobile user may lead to the query being linked to the sender [2].

In this work we focus on continuous queries and assume that the location of the mobile users may be known, but the query should not be linked to the mobile user. In some systems clients are willing to reveal their location but are not willing for the query to be linked to them, e.g. the query, "*Where is the closest HIV clinic to my current location?*", should not be linked to a mobile user regardless of whether the location of the mobile user that submitted the query is exposed. Some examples of applications whose location is public knowledge but the query should not be linked to the sender are:

(1) *Taxi Control Systems* – In these systems clients of the taxi services normally telephone the dispatching center and request a cab from an origin point to a destination point. In order to guarantee an efficient system the dispatching center is made aware of all the locations of the taxis in order to dispatch the closest taxi to the user. Even though the dispatcher center may be aware of the location of all the cabs, the company may not be allowed to monitor cab driver activities such as the queries that they may issue [2].
(2) *Courier Systems* – This is similar to the Taxi Control Systems. In these systems end users have packages to be collected from their location to some destination. For efficiency purposes the courier company may maintain the location of all the pick-up vehicles in order to allow the closest pick-up vehicle to collect the package from the end users.
(3) *Transit Itinerary Systems* – In these systems users may request the shortest path from a location to some destination. In order to provide accurate results the location of the user must be known by the itinerary system.
In our model we allow the user to specify their own personalized value of *K* in *K-anonymity* [12]. Consider a mobile client called "Mike" that is traveling by a car, and submit the query "*Where is the closest available parking lot for xxx rehab center from my current location?*" from his mobile device to some location based service provider. We assume that the location of the mobile client (Mike) may be known to the adversary, and Mike requires a *K-anonymity* level, e.g. *K=4*. We also assume that the transportation modes of some of the users are known by the adversary. This is a realistic assumption because several papers are available where the mode of transportation can be determined from a set of GPS points [6, 7, 20, 26, 27]. If Mike is anonymized with three other clients that are not traveling by car, then the query can be linked to Mike with high confidence. For example, the other three persons could be walking, running, on bus, on train or stationary. For this reason we introduce the *transportation mode homogeneity anonymity* concept where mobile users traveling by the same mode of transportation are anonymized together. If some transportation modes are indistinguishable from GPS technology, these modes will be aggregated into one mode.
This is related to Chow's and Mokbel's observation that anonymity can be violated based on a sequence of queries [2]. In our case the transportation mode is detected based on a sequence of queries, and then the transportation mode is used to violate privacy.
We use the term *global privacy* to mean that the mobile user is protected from *query linking attacks* in a continuous query system. A global privacy scheme based on a novel dynamic layered approach is proposed by separating the local privacy requirement on each snapshot and global privacy requirement across snapshots with different privacy goals. It exploits the local privacy anonymization group as candidates to obtain global anonymization group candidates.

2. TRANSPORTATION MODE FROM GPS

First, there are several systems whereby the mobile users are willing to reveal their location with the expectation that the query that they submitted cannot be linked back to the sender [2]. Second, if knowledge of location is known by adversaries then the transportation mode can be determined [6, 7, 20, 26, 27]. This work is based on continuous queries such that mobile users submit multiple location based requests at different timestamps and we assume that the adversaries may be aware of the locations of the mobile clients.
Several papers are available that determine a mobile user's transportation mode from GPS data [6, 7, 20, 26, 27]. In these papers the GPS location of users can be extracted over time and aggregated to determine factors such as speed, heading change rate (HCR), stop rate (SR) and other features [6]. Different transportation modes correspond to variations of these features. For example, in [6] it was proven that buses have a higher SR than cars and walking has a higher HCR than buses or cars. Knowing and understanding these features, the authors of [6, 7, 20] used machine learning techniques to inference the transportation modes of mobile clients from a set of GPS location points. Moreover, from a set of GPS points along a trajectory, knowledge of speed may be obtained from these points. Knowledge of speed may reveal the transportation mode, for example, the average speed of a car is different from the average speed of a non-motorized transportation mode [6].
Furthermore, the real time location (latitude and longitude) of public transportation is publicly available in some cities such as Chicago [8]. If the locations of the mobile clients are known then the mobile client's locations at different timestamps can be compared to the real time location of buses in a city in order to determine if the transportation mode of the mobile user is a bus.
In our work, we assume that the mobile user location may be known at each request submission. These systems are common whereby the mobile user is willing to reveal its location but the query submitted by the mobile user should not be connected to the mobile user [2]. We also assume that adversaries are able to determine the transportation mode of a mobile user from the set of GPS location points at different timestamps in a continuous query. Both assumptions are realistic [2, 6, 7, 20, 26, 27].

3. MOTIVATING EXAMPLE

Consider the following scenario with 10 mobile clients (*A,B,C,D,E,F,G,H,I,J*) with corresponding transportation modes. Location *K-anonymity* ensures that the mobile user is cloaked in a region with at least *K-1* other mobile clients. In this way the adversary cannot claim a user in that region with more than *1/K* confidence. This may be enough for snapshot query based systems [11, 12, 13, 14, 15, 16], as we will highlight this approach maybe ill-suited for continuous query based systems [2, 9, 10].Consider TABLE 1 which is for explanation purposes only and is not a part of the system.
In Figure 1 ten mobile clients (*A* to *J*) are in the system. The continuous query is submitted by mobile user B and consists of three different timestamp readings ($\mathbf{t_0}$, $\mathbf{t_1}$, $\mathbf{t_2}$).

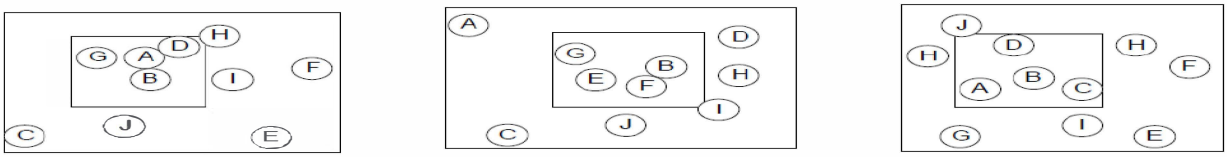


Fig 1, k=4 (a) time = t_0 **(b) time =t_1** **(c) time = t_2**

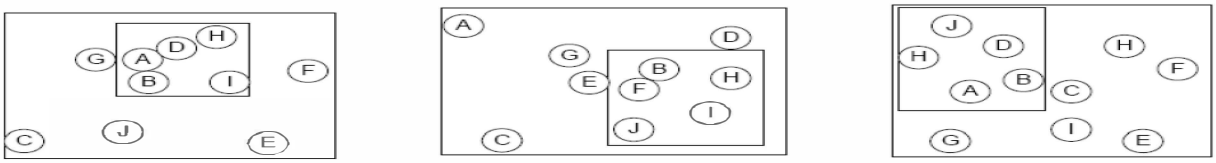


Fig 2, K_{local}=4,K_{global}=2,t_{mode}= *car* (a)time = t_0 **(b) time =t_1** **(c) time = t_2**

The large rectangles in Figures 1(a, b, c) represent the map of the region where the mobile clients reside. The small rectangles represent a region request containing the mobile users that are anonymized together. The desired local anonymity level is K_{local}=*4*. K_{local} is the local *K-anonymity* constraint for the snapshot. For this example, consider the query "*Where is the closest fast food restaurant to my current location on my route to the 'abc' strip bar?*". We refer to this query as *Q1*.

TABLE 1

Mobile Client	**Transportation Mode**
A	Car
B	Car
C	Walk
D	Run
E	Bus
F	Stationary
G	Train
H	Car
I	Car
J	Car

To make our discussions clear to the reader we separated the observations from the example above into three sections:

(1) *Transportation Mode*

First, at time $\mathbf{t_0}$ mobile user *B* submits *Q1* with K_{local}=*4* anonymity requirement. This implies that the user would only have a ¼ chance of being linked to the query. The mobile user *B* is cloaked encompassing three other mobile clients *A, G* and *D*. Observe clearly that if the modes of the mobile clients are known by the adversary then, at $\mathbf{t_0}$ the adversary may filter the mobile clients *G* and *D* since *G* is on a train and mobile client *D* is using non-motorized transportation. Knowing the transportation mode at $\mathbf{t_0}$ will relate query *Q1* to either mobile user *A* or mobile user *B* since they both are traveling by car. Hence, the query linking possibility is now ½. Moreover, at time $\mathbf{t_1}$, the query may be positively linked to mobile client *B.* If at least K_{local} *-1* other mobile users in the region were traveling by the same mode of transportation query linking via transportation mode would have been eliminated. We refer to the policy of anonymizing with at least K_{local} *-1* of the same transportation mode as *transportation mode homogeneity anonymization.* This is a key concept that we introduced and evaluated in this work.

(2) *Global Privacy*

Local privacy ensures that each individual snapshot is transportation mode anonymous with respect to some *local K-anonymity* value. Global privacy ensures that the aggregation of all the submitted snapshots is also transportation mode anonymous with respect to some *global K-anonymity* value.

At time $\mathbf{t_0}$ the local *K-anonymity* of K_{local}=*4* submitted by the mobile client *B* is satisfied because mobile client *B* is anonymized with mobile clients *A*, *G*, and *D* as shown by the small rectangle in Figure 1 (a). Also, at $\mathbf{t_1}$ the local *K-anonymity* of the mobile client is satisfied (K_{local} =4). Since the adversary is aware of the location of the mobile client, the adversary may take the intersection of the two snapshots **($\mathbf{t_0}$, $\mathbf{t_1}$)** and conclude that only mobile clients *B* and *G* are present in both snapshots, hence the query linking is reduced to ½. Furthermore, at time = $\mathbf{t_2}$, if the intersection of all three snapshots **($\mathbf{t_0}$,$\mathbf{t_1}$,$\mathbf{t_2}$)** is taken, mobile client *B* will be positively linked to the query. We refer to this as a reduction of global privacy of the mobile client. In our work we allow the mobile clients to specify the global privacy they desire as K_{global} and we ensure that at least K_{global} *-1* mobile clients that are transportation mode homogeneous are common in the region each time a request is submitted.

Different adversaries can have distinct levels of knowledge. Since there are fewer adversaries with more sophisticated knowledge, it may be good to provide a stronger level of privacy protection against the more common adversaries with weaker knowledge, i.e., maintaining a higher local *K-anonymity*. An adversary with weaker knowledge may be capable of deciphering only individual snapshots hence higher local *K-anonymity* is a deterrent. An adversary with sophisticated knowledge may be able to intersect or aggregate multiple snapshots. This will be addressed by the global *K-anonymity,* which is smaller than the local *K-anonymity* value.

(3) *Aggregation of Global Privacy and Transportation mode*

It should be clear to the reader that if we consider transportation mode and global privacy simultaneously the mobile client may be linked to the query much easier.

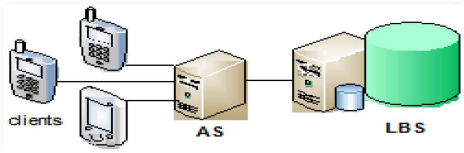
The example above highlights two important issues that are

the main motivations behind our work. The first issue is the fact that if adversaries have knowledge of the transportation mode then queries can be linked to the mobile client even if regular (local) *K-anonymity* is satisfied. The second issue is the reduction of global privacy if multiple snapshots at different time stamps in a continuous query are aggregated. The reduction of global privacy influences query linking.
We therefore introduce an approach that guarantees global privacy and transportation mode granularity anonymization. We aim for transportation mode homogeneity instead of diversity.

4. ARCHITECTURE and ASSUMPTIONS

Our work is focused on the trusted third party architecture. The mobile clients first send the requests to the anonymization server (AS). The request also contains the user and request identification parameters, transportation mode, personalized value of *K* and location. The AS then cloak the mobile client location point into a region containing *K-1* other users, as depicted by the rectangular region in Figures 1(a, b and c). The AS then forwards the region request to the LBS. We sometimes refer to a region request as a snapshot. In previous work [2, 12, 16], the other *K-1* users that are in the cloaked region can be of any transportation mode, which influences query linking. In our approach at least *K-1* other mobile clients in the region must be of the same transportation mode.
After the AS forwards the region request to the LBS, the region request is processed by the LBS and a response is sent back to the or to the mobile client. This response returned by the LBS is generic and should be filtered to decipher accurate results. Filtering can be done on the AS or by the mobile client. If a large region consisting of the *K* users is sent to the LBS then we may have numerous unwanted results been returned. Unwanted results imply more filtering and processing by the AS or the mobile client. A diagram depicting our architecture is shown below in Figure 3.

Figure 3 – LBS with Anonymization Server (AS)



Once a mobile user submits a continuous request to the AS, the AS continues to issue the request at different time intervals on behalf of the user based on his or her current location until the query is expired.
We assume an environment containing mobile users with positioning capabilities, location based services (LBS), wireless networks and our algorithms running on a middle-ware of the anonymization server (AS). We assume that adversaries loiter between the AS and the LBS or adversaries directly aim for the LBS. We also assume that adversaries may know the exact location of some of the mobile users along with the time when they have submitted the request. This assumption is realistic as many location based systems reveal their mobile client's location [2]. However, the query that they submitted must not be linked to the mobile client. Since the locations of the mobile clients are available it is possible for an adversary to determine the mode of transportation [6, 7, 20, 26, 27]. We also assume that the transportation modes of the mobile clients may be known by the adversaries.

5. MODELS AND NOTATIONS

Each mobile client submit a query q in the form of $\{usr_{id}, q_{id}, K_{local}, K_{global}, t_{mode}, exp_{time}, \{x,y\}, q_{time}, q_{con}\}$. Where usr_{id} and q_{id} is the unique identification of the mobile client that submitted the query and the identification of the query respectively. Before submitting the request to the service provider (SP), usr_{id}, and q_{id} are hashed and replaced by usr'_{id} and q'_{id}. The purpose of the user identification parameter is to re-identify the mobile clients when the query results are returned from the service providers[16]. K_{local} is the local *K-anonymity* for each snapshot in a continuous query. K_{global} is the global anonymity level that controls the query linking privacy requirement by restricting mobile user selection in the cloaking region in order to guarantee global privacy, when multiple snapshots in a continuous query is aggregated. The parameter t_{mode} is an optional parameter, sometime the users want to specify their transportation mode in order to get the right service. Even if t_{mode} is not specified, the AS may infer the transportation mode as discussed in Section 2.
A continuous query will be issued periodically by the AS on behalf of the user to the LBS before the exp_{time} has elapsed. For example, the AS could be configured such that for a continuous query the AS executes a snapshot query every 30 seconds. Each query that is issued by the AS will take the mobile client's current location into consideration. The parameter *{x,y}* represents the latitude and longitude of the location of the mobile client, where the value is determined by the GPS or other positioning component on the mobile device. The time that the query was submitted is represented by the parameter q_{time}. The content of the query is represented as q_{con}.
A query can have one of three states: (1) *fresh,* (2) *active*, and (3) *expired*. A *fresh* query is a newly created query, at time = q_{time}. An *active* query is one that was issued before and has not yet expired, as a continuous query contains multiple queries submitted for the same mobile user. A query is active for a period of $exp_{time} - q_{time}$. Expired query as the name implies is a query that exp_{time} has elapsed. The exp_{time} is used by the AS to determine the number of snapshots in a continuous query.
Multiple users are cloaked together to form a *region request* (R_i) for each snapshot query in the continuous query *CS*. A continuous query consists of multiple snapshots for the same user. We represent a *region request* R_i as the i^{th} region request (small rectangular region in Figures 1 (a,b,c) and Figures 2 (a,b,c)), and the total number of region requests in a continuous query is n. Since the user specifies exp_{time}, the AS determines the n based on the duration of the *active* period i.e., $exp_{time} - q_{time}$. A longer active period corresponds to a larger n. For example, the AS could be configured such

that for a continuous query the AS executes a snapshot query every 60 seconds. Thus for a query with an active period of 5 minutes, the AS would generate $n=5$. A region request R_i = {R_{id} , q_{set} , BR}, such that R_{id} is a region request identification attribute , q_{set} is the set of queries contained in R and BR is the bounding rectangle for the cloaked region. For example, in Figure 1(a) the rectangle containing mobile users (A, B, G, D) is a region request. Note that K_{local} is the local K-anonymity requirement for all R_i.

We define $|R_i|$ to represent the number of mobile clients in R_i and $|R_i|^{mode}$ to define the number of users in R_i with a specific transportation mode. For example, $|R_i|^{bus}$ means the number of mobile clients in $|R_i|$ traveling on bus.

Our algorithm is bounded by the following definitions:

Definition 1.1 (LBS Quasi-Identifier) *A set of attributes {Q1, . ., Qn} of a mobile query is called a LBS quasi-identifier if these request parameter can be linked with external data to uniquely identify at least one mobile client in the system.*

One example of quasi identifier (QI) in location based systems is the location {latitude,longitude} of the mobile user submitting the request.

Definition 1.2 (Transportation mode homogeneity anonymity) *A continuous query (CS) satisfies transportation mode homogeneity if all the region requests (snapshots) are transportation mode anonymous. A region request (R_i) is transportation mode anonymous iff the region contains at least K_{local}-1 other users with the same t_{mode} as the request. Therefore,* $|R_j|^{request.mode} >= K_{local}$ *for j = 1 to n.*

Definition 1.3 (Local K-anonymity) *A region request (R) satisfies local K-anonymity (K_{local}) if for every mobile client m ϵ R there exist at least K_{local} -1 other mobile clients $m_1, m_2, m_3, m_4.....m_{k-1}$ ϵ R such that any identifiers such as transportation mode that influences query linking is the same for $m_1, m_2, m_3, m_4.....m_{k-1}$.*

Definition 1.4 (Global K-anonymity) *A continuous query (CS) satisfies global K-anonymity if the intersection of all the region requests or snapshots in the continuous query is at least K_{global}. Therefore,* $|R_1 \cap R_2 \cap R_3 \cap R_4 \cap ..... Rn| >= K_{global}$ and $K_{global} <= K_{local}$

In Figure 2, consider the mobile clients in TABLE 1 and query $Q1$ submitted by mobile client B where K_{local} = 4, K_{global} = 2, t_{mode} = "*car*". The remaining parameters needed to complete the query parameter set is omitted for this explanation purpose only. We refer to the rectangular region containing the mobile clients in Figures 2 (a,b,c) as R_1, R_2 and R_3 respectively.

First, observe that in Figure 2(a) the local *K-anonymity* required by mobile client B (K_{local} =4) is satisfied. However, the cloaking region (R_1) contains at least three other mobile clients with the transportation mode equal "car". Query linking is not possible at this stage even if the transportation modes are known.

Second, in Fig 2(b) K_{local} =4 is also fulfilled to the transportation mode anonymization granularity (*Definition 1.3*). Also, K_{global} =2 is satisfied as the intersection of R_1 and R_2 gives mobile clients {*B, H, I*}. This conforms to *Definition 1.4*. At this point, even if the adversary aggregates both R_1 and R_2 and also is aware of the locations and transportation modes, the adversary cannot link the query directly to a particular mobile client. Alternatively, the approach could have considered including mobile client A in the region request in Figure 2(b). Including mobile client A would imply a large compromise on spatial tolerance. We also take a constraint on spatial tolerance when we attempt to cloak our mobile clients.

Third, we analyze Fig 2(c) and we observe that the required local anonymity of K_{local} =4 is satisfied at the transportation mode level as proposed in this paper. Furthermore, the global anonymity (K_{global}) requirement of mobile client B is also fulfilled by aggregating R_1, R_2 and R_3 to get {B, H} of size 2.

Observe that we may pay a price in spatial expansion in order to guarantee *global privacy* and *transportation mode homogeneity*.

6. ALGORITHM

In this section we present and discuss the Dynamic Transportation Mode Cloaking (*D-TC*) algorithm which is different from the previous approach in [2]. In [2] the algorithm continues to search for the same set of mobile clients until the query expires. However, the same mobile clients anonymized before may be much further apart in future snapshots. This implies enlargement of the spatial region to find the same mobile clients as before. We note that enlargement of spatial regions has several implications. First, large regions overwhelm the LBS with load, and also reduce the QoS of the results sent back by the LBS to the clients or the AS. Finally, large regions add more processing and filtering cost to the AS or the mobile client.

We introduce a novel dynamic layered approach to guarantee global privacy across snapshots. The dynamic layered approach separates the local privacy on each snapshot and global privacy across snapshots with different privacy goals and exploits the local privacy anonymization group as candidates to obtain global anonymization group candidates. It uses a dynamic snapshot suppressing strategy to guarantee global privacy. It will strike a balance between the QoS and the number of snapshots issued for the continuous queries, referred to as the *completeness* of the continuous queries.

The *D-TC* cloaking methodology is a bottom up cloaking strategy [16] where it continues to expand the region around the mobile user that submitted the request until it finds the K_{local}-*1* closest mobile clients traveling via the same transportation as the request submitter. The region chosen to submit to the LBS will be the bounding rectangle surrounding at least K_{local} mobile clients with similar transportation mode as shown in Figures 2 (a,b,c).

More specifically, the user submits a request in the form of {usr_{id}, q_{id}, K_{local}, K_{global}, t_{mode}, exp_{time}, {x,y}, q_{time}, q_{con}}, the parameters are explained in section 4. The mobile user only submits this request once and the AS continues to issue the request at different time intervals on behalf of the user. Each request that is issued by the AS on behalf of the mobile client must consider the mobile client's current location The parameter K_{local} is not mandated to be entered by the user, as

K_{local} can be learned by the AS from experience or from user profiling. For example, the AS may learn from experience on how to select the K_{local} value for a specific mobile client to satisfy his or her requirement on completeness and QoS.

The *D-TC* algorithm ensures that for each snapshot the mobile client is cloaked in a region encompassing at least K_{local} -1 other mobile clients using the same mode of transportation.

In the *D-TC* algorithm, the total number (*n)* of region requests (*snapshot queries*) (R_1.....R_n) is determined by the AS based on the query expiration time (exp_{time}). The *active* period of a query can be determined from the query's exp_{time} - q_{time}. A long active period corresponds to a larger value for *n*. For example, the AS could be configured such that for a continuous query the AS executes a snapshot query every 30 seconds, for a query with an active period of 7 minutes the AS would generate *n=14* and attempt to execute 14 (R_1.....R_{14}) region requests (snapshots).

At each snapshot (R_1.....R_{14}), the *D-TC* algorithm finds at least K_{local} closest mobile clients with the same transportation mode. If K_{global} common mobile clients can be found from the intersection with all previous snapshots, *D-TC* proceeds to issue the snapshot to the LBS. Otherwise, if K_{global} common mobile clients cannot be found in this snapshot, *D-TC* considers cloaking with the mobile clients from the previous snapshot that had just satisfied the global constraint. A snapshot that cannot satisfy the global constraint it will be suppressed, instead of reducing the global privacy, e.g. considering the case where some of clients used to provide the global privacy anonymization in the previous snapshots are out of the map or cannot be found for this snapshot. This way *D-TC guarantees global privacy*. If a snapshot is suppressed, *D-TC* continues to check subsequent snapshots. For example, if K_{global} is satisfied at R_1, R_2 , R_3 and *D-TC* cannot satisfy K_{global} at R_4, it suppresses R_4 and continues to check subsequent snapshots R_5.....R_{14}. Thus *D-TC* guarantees global privacy, however it may suppress some snapshots.

7. EXPERIMENTAL EVALUATION

Our work is the first to consider transportation mode anonymization and the dynamic layered approach for achieving global privacy with high quality of service. We evaluated three algorithms: (1) Robust Spatial Cloaking (*RSC*) – This algorithm uses the memorization property in [2] and does not consider transportation mode homogeneity. In RSC the first set of mobile clients in the first snapshot is found for all subsequent snapshots to provide global privacy. (2) Static Transportation Mode Cloaking (*S-TC*) – In this algorithm we consider transportation mode homogeneity constraints for the local and global privacy. To satisfy local privacy, *S-TC* anonymizes with K_{local} *-1* other mobile clients traveling with the same transportation mode. To satisfy global privacy, *S-TC* ensures that at least K_{global} -1 other mobile users with similar transportation mode are present in the intersection of all the snapshots in the continuous query. S-TC finds the K_{global} closest users in the intersection of the first and second snapshot and continues to include these same users in all subsequent snapshots.

(3) Dynamic Transportation Mode Cloaking (*D-TC*) as described in section 6.

We used the mobile object generator from [12] and generated 10,000 mobile users moving on a map of the Chamblee region in the state of Georgia, USA. The mobile users were randomly assigned transportation modes from the set {*stationary, walk, run, car, bus, train}*. Furthermore, the mobile users randomly generate location based requests to the AS. On receiving each request the AS anonymizes that request, we then measured our evaluation criterion below against the anonymized request. Our experiments where conducted on a HP Notebook PC running Windows Vista and contained a P8400 Intel DUO 2.27 GHz processor with 4GB RAM. The algorithms were implemented in JAVA.

7.1 EVALUATION CRITERIA AND METRICS

We now discuss the evaluation criteria that we use to measure the efficiency of our algorithm. We evaluated the algorithm with three considerations: (1) Privacy Guarantee (2) Quality of Service (3) Performance measure.

(1) *Privacy Guarantee* – We evaluated both the local and the global privacy of the algorithms. We measured the *instantaneous local privacy*, *average local privacy* and *global privacy.* Below we explain how we formulate our privacy metric.

(a)Local Privacy Guarantee - We have highlighted the fact that without transportation mode homogeneity anonymization then it is possible to link query to mobile clients that submit requests to LBS. In this work we measure the local privacy with respect to transportation mode homogeneity. We define $P_{local}(R_i)$ as the local privacy level in percentage for a single region request R_i as shown below. Recall that $|R_i|^{mode}$ was defined in section 5 as the number users in region R_i with a specific transportation mode.

$\boldsymbol{P_{local}(R_i)=(|R_i|^{mode}/K_{local})*100}$, *where mode = transportation mode of request issuer*

$$\textbf{if } |R_i|^{mode} > K_{local} \textbf{ then } |R_i|^{mode} = K_{local}$$

The average local privacy percentage $P_{local}(CS)$ for a continuous query CS with *n* region requests (snapshots).

$$P_{local}(CS) = (\Sigma_{i=1 \text{ to } n} P_{local}(R_i)) / n$$

Although the dynamic layered approach may suppress some of the snapshots in order to preserve global privacy, here we are measuring the composition of each potential snapshot request regardless of whether it gets suppressed or not.

(b) Global Privacy Guarantee - Consider a user submitting requests. If for the first snapshot, we get (B,C, D, E) for K_{local} =5 and for the 2nd snapshot we get (B, C, G, H) for K_{local} =5. Globally, we only have an overlap of two common mobile clients B and C. So we can only satisfy K_{global}=2, if and only if B and C are using the same transportation mode as the request issuer. If a user specifies K_{global} = 4, the desired global privacy level would not have been achieved. The global privacy would have been ¾=0.75 or 75% *global privacy*. In our approach if a snapshot reduces the global privacy from 100%, we suppress that snapshot. This way

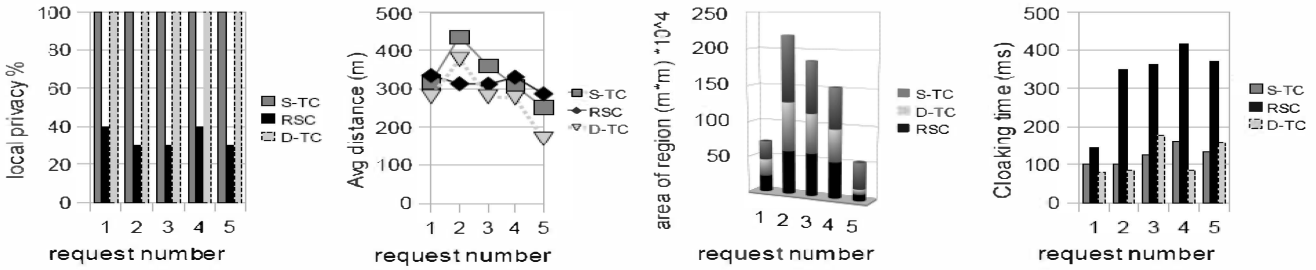


Fig 4 (a) (b) (c) (d)

we can guarantee a global privacy of 100%. We measure the number of snapshots that were submitted that meet the global privacy requirement divided by the *n* value generated by the AS as the c*ompleteness* of the privacy algorithm. For example, if the AS generated n=10 and only 8 of the 10 snapshots met the global privacy requirement and were submitted by the AS, the *completeness* would have been 8/10= 80%. However, the global privacy would still have been 100%, since only snapshots that satisfied the local and global constraints are submitted by the AS to the LBS.

(2) *Quality of service* – The quality of service in mobile location based systems depends upon the size of the region that is sent to the LBS [12]. If a large region is sent to the LBS by the AS, the LBS returns a more coarse grained location dependent information back to the mobile user. This deteriorates the quality of service. Also the LBS sends more than the required information to the mobile user, which requires more filtering and processing.

We evaluated the QoS using two metrics: (1) *Average distance metric,* and (2) *Area of bounded region metric*.

For the a*verage distance metric,* we sum the distances from the mobile user that submits the request to all the K_{local}-1 mobile clients in the cloaking region, then take the average. If the average distance to the other K_{local}-1 clients is low, it means better quality. We now explain how we computed average distance values in the experiments for a single instance of a region request that is an element of the continuous query that we are evaluating. Let $dist(m_i)$ be the distance from the mobile user that submitted the request to another mobile user m_i that was anonymized together. We compute the average distance $dist_{avg}(R_i)$ of a region request R_i as follows:

$$dist_{avg}(R_i) = (\Sigma_{i=1 \text{ to } Klocal-1} dist(m_i))/(K_{local}-1)$$

The *average distance* measure $dist_{avg}(CS)$ for the continuous query *CS* with *n* such region requests is computed as follows:

$$dist_{avg}(CS) = (\Sigma_{i=1 \text{ to } n} dist_{avg}(R_i))/n$$

In the *area of bounded region metric*, we measure the area of the bounding rectangle surrounding the K_{local} mobile users in the region where the request is submitted from. Similar to the average distance approach, a lower area will result in less coarse grain results from the LBS. Let $area(R_i)$ be the area of the bounding region that surrounds the request issuer and the other K_{local}-*1* users, measured as the length*width of the region. We compute the average area $area_{avg}(CS)$ of a continuous query *CS* with *n* request as follows:

$$area_{avg}(CS) = (\Sigma_{i=1 \text{ to } n} area(R_i))/n$$

(3) *Performance measure* – We evaluated the performance as the *cloaking time* to find the K_{local}-*1* closest users. A linear search was used to find the K_{local}-*1* other mobile clients. The cloaking time is the time the algorithm takes to perturb the request. Cloaking time is a measure of the temporal complexity of the algorithm. Average cloaking time ct_{avg} for a continuous query CS consisting of *n* distinct region requests is shown below, where $ct(R_i)$ is the cloaking time of the region request R_i:

$$ct_{avg} = (\Sigma_{i=1 \text{ to } n} ct(R_i))/n$$

7.2 RESULTS

In this section we present empirical evaluation of the algorithms using the metrics discussed in section 7.1. In Figures 4 (a,b,c,d) we evaluated the local privacy level, cloaking time and quality of service of all three algorithms (*D-TC, S-TC* and *RSC*). The quality of service is evaluated using the *average distance* metric and also the *bounding region* metric.

The number of mobile clients moving along the map was 10,000. These mobile clients all have randomly assigned transportation modes from the set *{stationary, walk, run, bus, car, train}*. We configure the AS with K_{local} to a value of 10 and K_{global} to a value of 3. The mobile clients randomly submit continuous queries to the AS for anonymization. The number of instantaneous region requests R_i in each continuous query *CS* was fixed at 5 (*n*=5), i.e., each continuous query is executed 5 times by the AS during its active period. Figure 4 depicts experimental evaluation of each of the 5 region requests in *CS*.

Figure 4 (a) plots the graph of *local privacy level* for each *n=5* region requests. We observe that *D-TC* and *S-TC* have a much higher local privacy level than *RSC*. The local privacy level of RSC is below 50%, while *S-TC* and *D-TC* achieve a local privacy level of 100% for the *n* region requests. This is related to the fact that we measure local privacy with respect to the number of requests with similar transportation mode as the request issuer. While *D-TC* and *S-TC* consider transportation mode homogeneity, RSC has no regard for transportation mode.

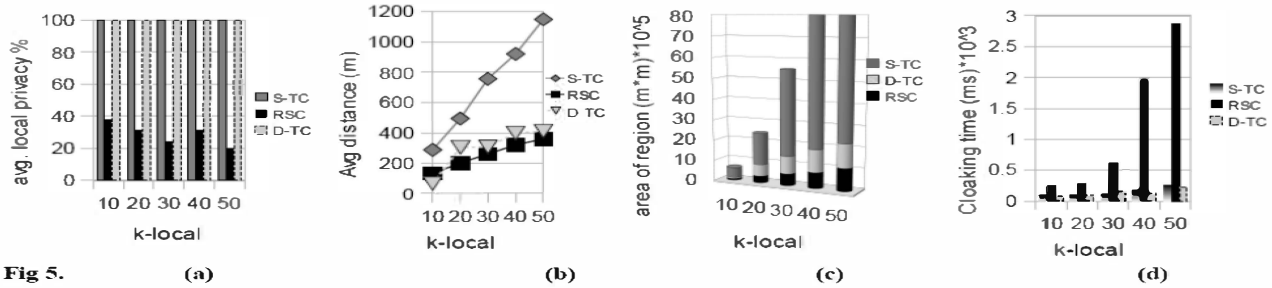


Fig 5. (a) (b) (c) (d)

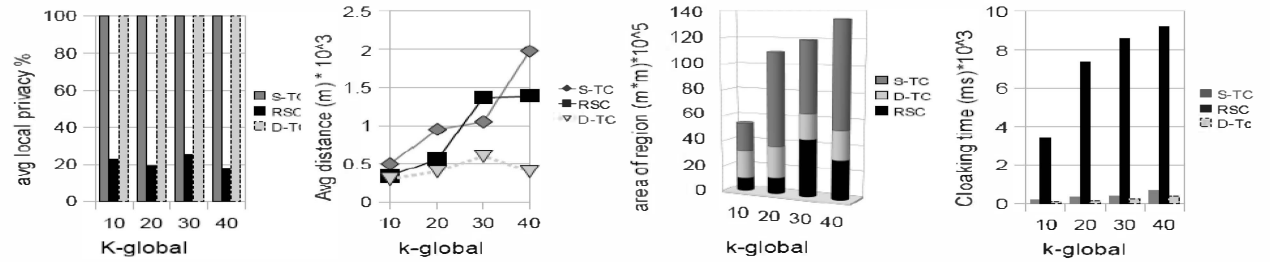


Fig 6. (a) (b) (c) (d)

Figures 4(b, c) highlight the quality of service evaluation. First, in Figure 4(b) we measure the QoS in terms of average distance to the K_{local}*-1* other mobile clients. We observe that the average distance for the three algorithms varies for the 5 snapshots. *D-TC* always performs better than *S-TC* and maintains the highest QoS for most of the 5 snapshots.

Secondly, in Figure 4(c), we evaluated the area of region bounding rectangle *BR.* Again it is clear that the area of the region containing the mobile clients in *S-TC* is much larger than both *RSC* and *D-TC*. Both figures, Figures 4(b and c), show that *RSC* and *D-TC* outperforms *S-TC* for QoS. The observation in Figures 4 (b,c) can be attributed to the fact that *D-TC* always selects the closest mobile clients, and also it is much easier to locate K_{local}*-1* users traveling regardless of transportation mode as in *RSC* than to find K_{local}*-1* users traveling by a specific transportation mode as in *S-TC.*

In Figure 4 (d) we evaluated the performance based on the cloaking time of the algorithms. From the graph it is clear that *S-TC* and *D-TC* have a much better performance than *RSC*. The *TC*s anonymize in a much faster time than *RSC*. This is due to the fact that a linear search is performed. In *D-TC* and *S-TC* we only consider requests with similar transportation mode as the requester for searching for anonymization candidates.. However, in *RSC* all the mobile clients are taken into consideration in the search regardless of the mode of transportation. Since *D-TC* and *S-TC* search a smaller domain than *RSC*, the *TC's* compute much faster.

We conclude our discussion of Figure 4 by claiming that *D-TC* and *S-TC* have a much higher local privacy percentage and cloaking time than *RSC*. Also *D-TC* achieves a very good quality of service. However, *S-TC* pays a price in the quality of service (*QoS*).

In Figures 5 (a, b, c, d) we evaluated the average local privacy level, average QoS and average cloaking time for a continuous query. The difference between this experiment and the experiments in Figure 4 is as follows. In the experiments depicted in Figure 4 we measured the instantaneous local privacy level, QoS and cloaking time for each individual request in the continuous query (*CS*). Now we focus on the average local privacy level, average QoS and average cloaking time for the entire CS , i.e., all the region requests in CS. We used 10,000 mobile clients traveling by various modes of transportation *{stationary, walk, run, bus, car, train}* randomly submitting requests to the AS. In these experiments we varied K_{local} from 10 to 50 in increments of 10 and observe the variation in the evaluated criterion. K_{global} is fixed at 3.

In Figure 5(a) we measured the *average local privacy level* of *D-TC, S-TC* and *RSC*. We observe that for all values of K_{local}, *D-TC* and *S-TC* have a 100% *average local privacy* percentage, while RSC is much lower. An increase in K_{local} shows little effect on the average local privacy level for RSC.

Figure 5(b) depicts the average QoS measurement of the continuous query via the average distance metric. It is quite obvious that *D-TC* achieves much better QoS than *S-TC.* As *S-TC* fixed the K_{global}*-1* clients for anonymization at the beginning, the QoS deteriorates over time as these clients drift apart. Furthermore, an increase in K_{local} decreases the QoS for *D-TC*, *S-TC* and *RSC* as expected.

In Figure 5(c), we measured QoS using the area of bounded region approach. Our observation is similar to Figure 5(b).

The graph of Figure 5(d) highlights the average cloaking time to anonymize all the requests in the CS. We observe that RSC has a very high cloaking time, and the cloaking time of *RSC* also increases rapidly as K_{local} increases. Furthermore, the cloaking time of *D-TC* and *S-TC* is much lower than RSC. An increase in K_{local} from 10 to 50 does not increase the cloaking time by a large factor for *D-TC* and *S-TC*.

We end this part of the evaluation discussion by concluding

that both *D-TC* and *S-TC* guarantee a higher local privacy level than *RSC*. Also, *D-TC* and *S-TC* also have a much better *cloaking time* than RSC. Furthermore, *D-TC* has a much higher quality of service than *S-TC* and close to RSC in most cases. In *D-TC* and *S-TC* we may have suppression of snapshots that did not meet the global privacy constraint.

In Figures 6 (a,b,c, d) we observe the effects of a variation of K_{global} on the evaluation criterion. Recall from sections 5 and 6 that K_{global} is used to reduce the possibility of the query linking attack by ensuring that the intersection of the region requests in CS does not reveal less than K_{global} amount of mobile clients. In this experiment the number of mobile users moving on the map was 10,000. We varied K_{global} from 10 to 40 in increments of 10 and K_{local} was fixed at 50.

Figure 6(a) shows that a change in K_{global} does not have any effect on the average local privacy of a mobile client for *D-TC, S-TC* and *RSC*. This is because in the evaluation criteria and metrics in section 7.1, the local privacy percentage measurement is not based on K_{global}.

Also, in Figures 6(b,c), a variation in K_{global} did not show a huge impact on the quality of service of *D-TC*. In *D-TC* we aggregate the closest mobile users regardless of K_{global} for the initial requests. On the other hand in Figures 6(b,c), both *S-TC* and *RSC* show a reduction in QoS as K_{global} increases.

Figure 6(d) depicts the graph of cloaking time, we observed that an increase in K_{global} implies an increase in cloaking time for *S-TC, D-TC* and *RSC.* Furthermore, an increase on K_{global} has a more profound impact on RSC.

Global Privacy

Recall that in our model global privacy is guaranteed because we suppress snapshots that did not meet the global requirement. In Figure 7, we evaluated the *completeness* of the algorithms as described in section 7.1. The *completeness* of the algorithms is the percentage of queries that satisfies the global privacy constraints including transportation mode homogeneity.

In *D-TC* and *S-TC* the *completeness* is at least *1/n* because at least the first snapshot will be submitted to the LBS. The AS was configured with K_{local} =10, the number of snapshot queries in the continuous set was 5 (*n=5*) and there were

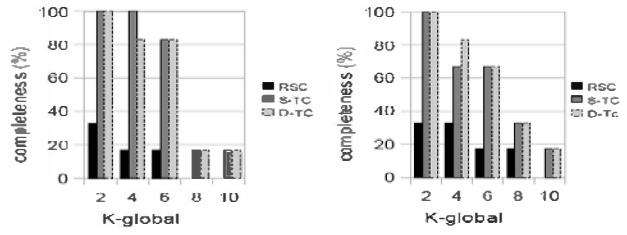


Fig 7

4000 mobile clients moving with transportation modes from the set *{stationary, walk, run, bus, car, train}*.

The most important observation from Figure 7(a) is the fact that as K_{global} increases and approaches K_{local} the *completeness* of all the algorithms is reduced. Also, as K_{global} approaches K_{local} the *completeness* of the *D-TC* and *S-TC* approaches *1/n.* Furthermore, *S-TC* and *D-TC* have a much higher completeness than *RSC*. More specifically, for *RSC,* as K_{global} increases and approaches K_{local} the *completeness* is zero, this means that none of snapshots met the global constraints.

In Figure 7(b) the configuration on the AS was similar to Figure 7(a), the only difference was that instead of generating 4000 mobile clients, in Figure 7(b) we generated 8000 mobile clients. The graph of Figure 7(b) highlights that the *completeness* is still low as K_{global} approaches K_{local} even with an increase in the number mobile clients from 4000 to 8000. In general, an increase in the number of mobile clients did not have a large effect on the difference in *completeness* of the algorithms. The completeness is similar for *S-TC* and *D-TC*as K_{global} approaches K_{local}.

From Figure 7, it is clear, for *D-TC*, the K_{local} should be chosen much larger than K_{global} to achieve high completeness, while the price on QoS is expected to be moderate as discussed before.

8. RELATED WORK

Most of the previous work done on privacy in mobile location based systems focused on snapshot queries [11, 12, 13, 14, 15, 16] and did not distinguish between *location privacy* [1] and *query linking privacy* [2] . Our work could be implemented as an extension to these snapshot systems in order to support privacy on continuous queries. The most popular privacy metric for location based mobile users is *K*-anonymity. Location *K*-anonymity was coined in [1] by Gruteser and Grunwald. Other privacy approaches include private information retrieval (*PIR*) mechanisms [17] and dummies [14,15,29]. In our previous work [29] we used realistic mobile dummies on the AS to improve the success rate. We did not consider dummies in this work. Gedik and Lui [12] introduced the concept of personalized *K*-anonymity where *K* was static for all users within the system.

The first work to distinguish between location privacy and query linking privacy is [2]. The authors in [2] addressed continuous queries but did not consider transportation mode anonymity. Furthermore in [2] the algorithm continues to anonymize the same set of mobile clients over the entire continuous set. This enlarges the request region and reduces QoS. Our work is different from [2] in the way we achieve global privacy. We guarantee global privacy at a reasonable QoS by searching for a subset (K_{global}) of the initial request. If we cannot satisfy K_{global} in any of the *n* snapshots, we may suppress that snapshot instead of submitting a query with a lower than required privacy level (*guaranteed global privacy*). If we suppressed a snapshot then we continue to verify future snapshot requests for the global constraint.

The work in [18] also focused on continuous queries, but it focused on entropy as anonymity measure. This work [18] is different from ours as the entropy takes into account not only the number of the entities inside, but also their anonymity probability distribution. Another research considering continuous queries in mobile LBS is [9]. It employs a privacy model and a distortion model to balance the tradeoff between the quality of service and the privacy level. In [22] the authors coined *l diversity* as a second

dimension of privacy beyond *K-anonymity*. Also, [21] uses concept of *segment l-diversity* over road networks. Our work is different; we aim for *transportation mode homogeneity* instead of diversity. In [13, 14, 15, 17], a client-server model is considered instead of using anonymization servers. Furthermore, the work in [28] is based on the peer to peer architecture. Our solution is the first to consider transportation mode anonymity as a metric to improve the privacy of mobile clients in location based mobile systems.

9. CONCLUSION

We presented a new approach called *transportation mode homogeneity anonymization* that improves the privacy of mobile clients in location based systems. We also proposed a novel dynamic layered approach on achieving *K-anonymity* by separating the local privacy on each snapshot and global privacy across snapshots with different privacy goals and exploiting the local privacy anonymization group as candidates to obtain global anonymization group candidates. Experimental evaluation shows that our model improves both the local and global privacy level, but also achieves an excellent performance measure and good quality of service.

REFERENCES

[1] M. Gruteser, D. Grunwald. *Anonymous usage of location based services through spatial and temporal cloaking. ACM/USENIX MobiSys, 2003.*

[2] C. Chow, M. Mokbel. *Enabling Private Continuous Queries For Revealed User Locations. International Symposium on Advances in Spatial and Temporal Databases, 2007.*

[3]USAToday. Authorities: GPS system used to stalk woman.http://www.usatoday.com/tech/news/2002-12-30-gpsstalkerx.htm.

[4]Fox News. Man Accused of Stalking Girlfriend With GPS.http://www.foxnews.com/story/0,2933,131487,00.html

[5] L. Sweeney. *K-Anonymity- A model for protecting privacy. International Journal of Uncertainty Fuzziness and Knowledge Based Systems, 2002.*

[6] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W. Ma. *Understanding mobility based on GPS data. In Ubiquitous Computing, ACM , 2008.*

[7] S. Reddy, M. Mun, J. Burke, D. Estrin, M Hansen, and M. Srivastava. *Using Mobile Phones to Determine Transportation Modes. ACM Transactions on Sensor Networks, 2010.*

[8] Chicago Transit Authority (CTA) Bus Tracker http://www.ctabustracker.com/bustime/home.jsp

[9] X. Pan, X Meng, J. Xu. *Distortion-Based Anonymity for Continuous Queries in Location-Based Mobile Services. ACM GIS, Nov. 2009.*

[10] G. Ghinita, P Kalnis, A. Khoshgozaran. C. Shahabi, K. Tan. *Private queries in Location Based Services: Anonymizers are not Necessary.* SIGMOD, 2008.

[11] M. Mokbel, C. Chow, W. Aref. *The New Casper: Query Processing for Location based Services without Compromising Privacy. 32nd International Conference on VLDB, September 2006.*

[12] B. Gedik, L. Lui. *Location Privacy in Mobile Systems: A Personalized Anonymization Model. ICDS, 2005*

[13] Man Lung Yiu, Christian S. Jensen, Xuegang Huang, Hua Lu. *SpaceTwist: Managing the Trade-Offs Among Location Privacy, Query Performance, and Query Accuracy in Mobile Services.* 24th *International Conference on Data Engineering, 2008.*

[14] H. Kido, Y. Yanagisawa, T. Satoh. *An Anonymous Communication Technique using Dummies for Location Based Services. Second International Conference on Pervasive Services, 2005.*

[15]T. You, W. Peng, and W. Lee. *Protecting Moving Trajectories Using Dummies. Proc. International Workshop Privacy-Aware Location-Based Mobile Services, 2007.*

[16] B. Bamba, L. Liu, P. Pesti, T. Wang. *Supporting Anonymous Location Queries in Mobile Environments with PrivacyGrid . World Wide Web, 2008.*

[17] G. Ghinita, P Kalnis, A. Khoshgozaran. C. Shahabi, K. Tan. *Private queries in Location Based Services: Anonymizers are not Necessary .SIGMOD, 2008.*

[18] T. Xu, Y. Cai. *Location Anonymity in Continuous Location Based Services.* GIS, 2007.

[19] P. Samarathi and L Sweeney. *Protecting Privacy when disclosing Information: k-Anonymity and its Enforcement through Generalization and Suppression. SRI-CSL-98-04*

[20] D. Patterson, L Liao, D. Fox, H. Kautz. *Inferring High Level Behavior from Low Level Sensors. In Proc. Of UBICOMP – 03 Springer Press, 2003.*

[21] T. Wang, L. Liu. *Privacy-Aware Mobile Services over Road Networks. VLBD, 2009.*

[22]A. Machanavajjhala, J Gehrke, D. Kifer and M Venkitasubramaniam. "*L-diversity": Privacy beyond k anonymity. ICDE, 2006.*

[23] TransitGenie Website- *Your Personal Transit Navigator* (November 2009) www.transitgenie.com

[24]NextBus-Website www.nextbus.com/predictor/agencySelector.jsp(2009)

[25] Google Maps For Mobile Website - http://www.google.com/mobile/maps/

[26] L. Liao, D. Fox, H. Kautz. *Learning and Inferring Transportation Routines. In Proc of AI ,2004.*

[27] L. Liao, D. Fox, H. Kautz. *Building Personal Maps from GPS Data. IJCAI MOO05. Springer Press, 2005*

[28] C. Chow, M. Mokbel, X. Liu. *A peer to Peer Spatial Cloaking Algorithm for Anonymous Location Based Services. ACM GIS, 2006.*

[29] L. Stenneth, P. Yu, O. Wolfson. *Mobile Systems Location Privacy:"MobiPriv" a Robust K-Anonymous System. IEEE WiMob, 2010.*