Noname manuscript No.
(will be inserted by the editor)

Anonymizing Continuous Queries with Delay-tolerant Mix-zones over Road Networks

Balaji Palanisamy, Ling Liu, Kisung Lee, Shicong Meng, Yuzhe Tang and Yang Zhou

Abstract This paper presents a delay-tolerant mix-zone framework for protecting the location privacy of mobile users against continuous query correlation attacks. First, we describe and analyze the continuous query correlation attacks (CQ-attacks) that perform query correlation based inference to break the anonymity of road network-aware mix-zones. We formally study the privacy strengths of the mix-zone anonymization under the CQ-attack model and argue that spatial cloaking or temporal cloaking over road network mix-zones is ineffective and susceptible to attacks that carry out inference by combining query correlation with timing correlation (CQ-timing attack) and transition correlation (CQ-transition attack) information. Next, we introduce three types of delay-tolerant road network mix-zones (i.e., temporal, spatial and spatio-temporal) that are free from CQ-timing and CQ-transition attacks and in contrast to conventional mix-zones, perform a combination of both location mixing and identity mixing of spatially and temporally perturbed user locations to achieve stronger anonymity under the CQ-attack model. We show that by combining temporal and spatial delay-tolerant mix-zones, we can obtain the strongest anonymity for continuous queries while making acceptable tradeoff between anonymous query processing cost and temporal delay incurred in anonymous query processing. We evaluate the proposed techniques through extensive experiments conducted on realistic traces produced by GTMobiSim on different scales of geographic maps. Our experiments show that the proposed techniques offer high level of anonymity and attack resilience to continuous queries.

Balaji Palanisamy
School of Information Sciences, University of Pittsburgh
E-mail: bpalan@pitt.edu

Ling Liu, Kisung Lee, Yuzhe Tang, Yang Zhou
College of Computing, Georgia Institute of Technology
E-mail: lingliu@cc.gatech.edu, kslee@gatech.edu, yztang@gatech.edu, yzhou86@gatech.edu

Shicong Meng
IBM T. J. Watson Research Center
E-mail: smeng@us.ibm.com

1 Introduction

Continuous location-based queries are gaining growing interests and attentions in both mobile data management research and mobile service industry. Examples of continuous queries (CQs) include "informing me the nearest gas stations coming up along the highway I-85 south every 1 minute in the next 30 minutes" or "show me the restaurants within 2 miles every two minutes during the next hour". Many consider continuous spatial queries as a fundamental building block for continuous provisioning of location services to mobile users traveling on the roads.

Continuous query attacks (CQ-attacks) refer to the query correlation attacks. Concretely, a CQ represents a time series of query evaluations of the same query within a given validity time window. For example, if Alice is traveling on I-85 south and requested a CQ service: show me the restaurants within 2 miles every two minutes during the next hour, then the CQ server will process this CQ as a standing query for 1 hour period upon its installation and it will consists of a sequence of 30 evaluations along the trajectory of Alice. The CQ-attack refers to the risk that an adversary can perform inference attacks by correlating the semantic continuity in the time series of query evaluations of the same CQ and the inherent trajectory of locations. In this paper we show that such CQ-attacks can intrude location privacy of mobile users (i.e., exposing the identity of Alice), even though the locations of mobile users are anonymized through well-known location anonymization techniques.

A fair amount of research efforts have been dedicated to protecting location privacy of mobile travelers. We can broadly classify the state of art research and development results into two categories. The first category is represented by location cloaking techniques [20,7,17,26,32]. Spatial location cloaking typically adds uncertainty to the location information exposed to the location query services by increasing the spatial resolution of a mobile user's locations while meeting location k-anonymity and/or location l-diversity [7]. More specifically, the spatially cloaked region is constructed to ensure that at least k users (*location k anonymity*) are located in the same region, which contains l different static sensitive objects (locations). Spatial cloaking is effective for snapshot queries but vulnerable to CQ-attacks. The second class of location anonymization techniques is represented by mix-zone development [9,15,16, 11,29]. Mix-zones are spatial regions where a set of users enter, change pseudonyms in such a way that the mapping between their old and new pseudonyms is not revealed. Also inside a mix-zone, no applications can track user movements. Mix-zones break the continuity of location exposure by introducing uncertainty such that it is very hard to perform correlation attacks to link old pseudonym with new pseudonym of mobile users. However, neither spatial cloaking nor mix-zone techniques are resilient to CQ attacks as they are vulnerable to query correlation. Concretely, with spatial cloaking, an adversary can infer user movement by performing query correlation attacks over the adjacent or overlapped cloaking boxes. Similarly, mobile users requesting CQ services are vulnerable under CQ-attacks even though their movements on the road networks are protected by mix-zone anonymization. With these problems in mind, in this paper we present a delay-tolerant mix-zone framework for protecting location privacy of mobile users with continuous query services in a mobile environment. First, we describe and analyze the continuous query correlation attacks

(CQ-attacks) that perform query correlation based inference to break the anonymity of road network-aware mix-zones. We formally study the privacy strengths of the mix-zone anonymization under the CQ-attack model and identify that providing high initial anonymity in the mix-zone model is the key to anonymizing continuous queries in a mix-zone framework. We argue that spatial cloaking or temporal cloaking over road network mix-zones is ineffective and susceptible to attacks that carry out inference by combining query correlation with timing correlation (CQ-timing attack) and transition correlation (CQ-transition attack) information.

Next, we introduce three types of delay-tolerant road network mix-zones (i.e., temporal, spatial and spatio-temporal) that are free from CQ-timing and CQ-transition attacks and in contrast to conventional mix-zones, perform a combination of both location mixing and identity mixing of spatially and temporally perturbed user locations to achieve stronger anonymity under the CQ-attack model. In the delay-tolerant mix-zone model, users expose spatially or temporally perturbed locations outside the mix-zone area. However, on the exit of each temporal delay tolerant mix-zone, the mix-zone changes their perturbed locations by introducing a random temporal shift to their already perturbed locations. Similarly, the spatial delay tolerant mix-zones introduce a random spatial shift to the spatially perturbed locations when the users exit them. While conventional mix-zones only change pseudonyms inside them, the additional ability of delay-tolerant mix-zones to change and mix user locations brings greater opportunities for creating anonymity. Our third type of delay tolerant mix-zones, namely spatio-temporal delay-tolerant mix-zones effectively combine temporal delay-tolerant and spatial delay tolerant mix-zones to obtain the highest anonymity for continuous queries while making acceptable tradeoff between anonymous query processing cost and temporal delay incurred in anonymous query processing. To the best of our knowledge, this is the first work to systematically study the benefits of combining location perturbation based techniques with mix-zone based location anonymization schemes for protecting against continuous query attacks. We evaluate the proposed techniques through extensive experiments conducted using traces produced by GTMobiSim [28] on different scales of geographic maps. Our experiments show that the delay-tolerant mix-zone techniques are efficient and offer the desired level of anonymity for continuous queries.

2 Related work

Location privacy has been studied over the past decade along two orthogonal dimensions: spatial cloaking through location k-anonymity represented by [17,26,7,32,13, 6] and mix-zone based privacy protection and its variations represented by [9,15, 16,11,29,25]. However, these approaches are suitable only for snapshot queries and are inadequate and ineffective for protecting location privacy of mobile users with continuous query services.

In recent years, there had been research efforts that dealt with location privacy risks of continuous queries. [10] describes various attacks in Location-based systems, including the continuous query attacks and the challenges of supporting continuous query services. [12] proposes spatial cloaking using the memorization prop-

erty for continuous queries. This is further used in [30] for clustering queries with similar mobility patterns. However, this type of techniques may lead to large cloaking boxes resulting in higher query processing cost as users may not always move together. [14] identifies that location cloaking algorithms with only k-anonymity and l-diversity guarantee are not effective for continuous LBS and therefore propose query m-invariance as a necessary criterion when dealing with continuous location queries. However, m-invariance based approach is ineffective when the mobile users ask uniquely different CQ services as they move on the road. An alternative thread of research is represented by the *Personal information retrieval* techniques as an alternate to location cloaking for anonymous query processing [19]. PIR techniques guarantee privacy of mobile users regardless of which types of queries (continuous or snapshot) they ask. However PIR based solutions are known to be expensive in both computation and storage overheads, even with the recent new techniques such as hardware-assisted PIR techniques [34], developed to improve the scalability and efficiency of the PIR approach. Another general issue with PIR based solutions is its limitation in terms of what kinds of queries can be protected under PIR [33].

The concept of mix-zones was first presented in the context of location privacy in [9]. The idea of building mix-zones at road intersections is proposed in [15] and [11]. In [16], a formulation for optimal placement of mix-zones in a road map is discussed. Almost all existing mix-zone techniques follow a straight forward approach of using a rectangular or circular shaped zone and their construction methodologies do not take into account the effect of timing and transition attacks in the construction process. The MobiMix framework presented in [29] is the first road-network aware attack-resilient mix-zone that guarantees an expected value of anonymity by leveraging the characteristics of both the underlying road network and motion behaviors of users traveling on spatially constrained road networks. However, all existing road network mix-zone approaches, to the best of our knowledge, fail to protect mobile users from continuous query attacks.

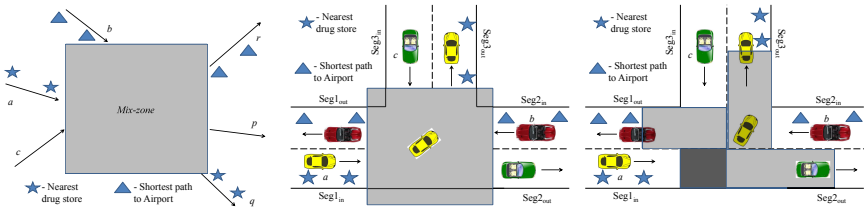
Inspired by the mix-zone concept, the *Cachecloak* algorithm [25] employs an alternate technique for path-mixing by using cache prefetching to hide the exact location of mobile user by requesting the location based data along an entire predicted path. Although these techniques are effective when all users obtain the same service, they are vulnerable to continuous query correlation attacks when the mobile users obtain uniquely different CQ services. Recently, content caching[5] has been proposed as an alternate solution to location privacy. However, caching large amounts of information on tiny mobile devices may not be effective. In addition, they may limit the usability of the services by restricting mobile clients to ask only services that are cached before-hand.

In this paper, we introduce delay-tolerant road-network mix-zones as an effective countermeasure against CQ attacks. We show that by performing a combination of both location mixing and identity mixing in the mix-zones, the delay-tolerant mix-zones offer greater level of anonymity that is sufficient to meet the anonymity levels of continuous queries under the CQ-attack model while maintaining acceptable quality of continuous query services. Though both location perturbation based techniques and mix-zone anonymization are well researched topics by themselves, to the best of our knowledge, this is the first work to systematically study the benefits of combining

location perturbation based techniques with mix-zone based location anonymization schemes to tackle sophisticated attacks such as the continuous query attacks. In general, we believe that an effective combination of location perturbation with mix-zone based techniques has the potential to provide the strongest defense against the sophisticated CQ attacks.

3 Mix-zones and CQ-attacks

In this section, we introduce the basic mix-zone concepts, illustrate the vulnerabilities of mix-zones to continuous query correlation attacks (*CQ-attacks*) and present a formal analysis of the continuous query anonymization problem.



(a) Theoretical mix-zone (b) Road network mix-zone (c) Non-rectangular mix-zone

Fig. 1: Mix-zone anonymization and its risks under CQ-attack

3.1 Mix-zone concepts

A mix-zone of k participants refers to a k-anonymizaton region in which users can change their pseudonyms such that the mapping between their old and new pseudonyms is not revealed. In a mix-zone, a set of k users enter in some order and change pseudonyms but none leave before all users enter the mix-zone. Inside the mix-zone, the users do not report their locations and they exit the mix-zone in an order different from their order of arrival, thus, providing unlinkability between their entering and exiting events. The properties of a mix-zone can be formally stated as follows:

Definition 1 A mix-zone Z is said to provide k-anonymity to a set of users A iff

1. The set A has k or more members, i.e., $|A| \geq k$.
2. All users in A must enter the mix-zone Z before any user $i \in A$ exits. Thus, there exists a point in time where all k users of A are inside the zone.
3. Each user $i \in A$, entering the mix-zone Z through an entry point $e_i \in E$ and leaving at an exit point $o_i \in O$, spends a completely random duration of time inside.
4. The probability of transition between any point of entry to any point of exit follows a uniform distribution. i.e., a user entering through an entry point, $e \in E$, is equally likely to exit in any of the exit points, $o \in O$.

Figure 1(a) shows a mix-zone with three users entering with pseudonyms a, b and c and exiting with new pseudonyms, p, q and r. Here, given any user exiting with a new pseudonym, the adversary has equal probability of associating it with each of the old pseudonyms a, b and c and thus the mix-zone provides an anonymity of $k = 3$. Therefore, the uncertainty of an adversary to associate a new pseudonym of an outgoing user i' to its old pseudonym is captured by Entropy, $H(i')$ which is the amount of information required to break the anonymity.

$$H(i') = -\sum_{j \in A} p_{i' \to j} \times log_2(p_{i' \to j})$$

where $p_{i' \to j}$ denotes the probability of mapping the new pseudonym, i' to an old pseudonym, j. Here note that when users change pseudonyms inside mix-zones along their trajectories, an adversary observing them loses the ability to track their movements.

Unlike the theoretical mix-zones, mix-zones constructed at road intersections (Figure 1(b)) may violate some conditions. For instance, in a road network mix-zone, users do not stay random time inside while entering and exiting the mix-zone [15,29] and also violate the assumption of uniform transition probabilities in taking turns. Such violations provide additional information to the adversary in inferring the mapping between the old and new pseudonyms. Precisely,the timing information of users entry and exit in a road network mix-zone leads to timing attack and the non-uniformity in the transition probabilities at the road intersection leads to transition attack respectively. The MobiMix road network mix-zone model (Definition 2) and the construction techniques [29] deal with the challenges of constructing road network mix-zones that are resilient to such timing and transition attacks. Accordingly a road network mix-zone is defined as:

Definition 2 A road network mix-zone offers k-anonymity to a set A of users if and only if:

1. There are k or more users in the anonymity set A.
2. Given any two users $i, j \in A$ and assuming i exiting at time t, the pairwise entropy after timing attack should satisfy the condition: $H_{pair}(i, j) \geq \alpha$.
3. Given any two users $i, j \in A$, the pairwise entropy after transition attack should satisfy the condition: $H_{pair}(i, j) \geq \beta$.

Here the pairwise entropy, $H_{pair}(i, j)$ between two users i and j represents the entropy obtained by considering i and j to be the only members of the anonymity set. In comparison, a theoretical mix-zone offers a high pairwise entropy of 1 for all pairs of users. Therefore, a good road network mix-zone would offer a pairwise entropy close to 1 for all pairs of users in the anonymity set. The non-rectangular road network mix-zone geometry (Figure 1(c)) proposed in [29] enables each user to travel along only one road segment inside the mix-zone and thereby avoids timing attack due to difference in speed distributions. Similarly, the MobiMix road network mix-zone model discusses the necessary criteria to ensure transition attack resilience in the mix-zone construction process. However as we discuss next, when users run continuous queries, any type of mix-zone is prone to CQ-attacks.

3.2 CQ-attack

When a user is executing a continuous query, even though her pseudonym is changed whenever she enters a road network mix-zone, an adversary may simply utilize the consecutive snapshots of the query to reveal the correlation between the old and new pseudonyms. Consider the example in Figure 1(a) where three users enter with pseudonyms a, b and c and exit with new pseudonyms p, q and r. The attacker finds that before entering the mix-zone, users a and b run continuous queries on obtaining nearest drug store and shortest path driving directions to the airport respectively. Upon their exits, the attacker again finds more instances of their corresponding continuous queries with different pseudonyms, q and r. Here, although users a and c change their pseudonyms to q and r, the continuous exposure of their CQ information breaks their anonymity. Similar attack can happen in a road network mix-zone as shown in Figure 1(b) where three users with pseudonyms, a, b and c enter and leave the mix-zone. As users a and b are running continuous queries, the attacker finds an instance of $a's$ continuous query before entering the mix-zone and when user a exits with a new pseudonym, say α and receives another instance of the same query, the attacker infers that the new pseudonym α must correspond to the old pseudonym a. To the best of our knowledge, no existing road network mix-zone technique is effective against CQ-attack. For instance, we find in Figure 1(c) that even the non-rectangular mix-zone [29] that is most effective against road network timing attack is also prone to the CQ-attack.

Anonymity under CQ-attack model: When a user executes a continuous query, it induces a trajectory corresponding to the movement of the user even though the user's pseudonym is changed whenever she crosses a road network mix-zone. When a mobile user starts a continuous query from a personal location like office or home address, then the user's trajectory induced by the continuous query can be easily linked with the user's personal location and hence to user's real identity even though the pseudonyms are changed time to time. Therefore, in the proposed mix-zone anonymization model, any mobile user who wishes to obtain CQ service first moves to the nearest mix-zone and starts to run the CQ service from the mix-zone. Consider the user with pseudonym, a in Figure 2 who wants to start a CQ service for obtaining k-nearest drug stores. User a first goes to the nearest mix-zone and starts the first instance of the query after exiting the mix-zone with a new pseudonym e. Here, the attacker becomes confused to associate this continuous query with the users $\{a, b, c\}$ who entered the mix-zone as each of them have equal likelihood of starting this continuous query. Hence the continuous query obtains an initial anonymity of $k = 3$. However, it should be noted that at the subsequent mix-zones, the query correlation reveals the mapping between the old and new pseudonyms of the continuous query. For example, in the second mix-zone the mapping between new pseudonym j and old pseudonym e is revealed and similarly in the third mix-zone, the mapping between new pseudonym m and old pseudonym j is revealed. Thus, the anonymity strengths of the mix-zones is weakened under the continuous query correlation attack model.

In contrast, if users in the system ask only snapshot queries, then user a's anonymity keeps on increasing as a traverses through more mix-zones as it does not run continuous queries. For example, when user a changes its pseudonym from e to j in the

second mix-zone, its anonymity increases from $k = 3$ (corresponding to $\{a, b, c\}$) to $k = 5$ (corresponding to $\{a, b, c, f, g\}$). Hence, even though the initial anonymity is low ($k = 3$), in the snapshot query model, the anonymity gained in the intermediate mix-zones add to the user's anonymity. However for a continuous query, its initial

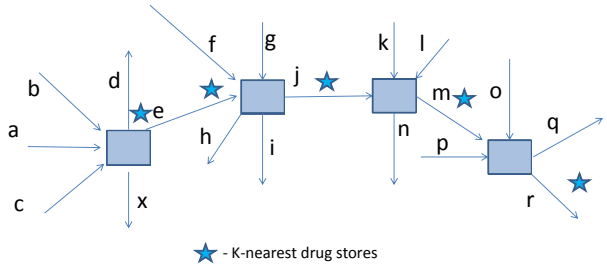


Fig. 2: CQ-induced trajectory

anonymity forms the major component and intermediate mix-zones add anonymity only when users in the intermediate mix-zones ask the same query. For instance, if m out of the k users traversing an intermediate mix-zone run continuous queries and if there are R number of unique continuous queries run by the m users and if A^r is the set of users running the continuous query, Q_r, $1 \leq r \leq R$, then the entropy of each continuous query user, i executing the query, Q_r and exiting the mix-zone with a new pseudonym, i' is given by

$$H(i) = -\sum_{j \in A^r} p_{i' \to j} \times log_2(p_{i' \to j})$$

where $p_{i' \to j}$ denotes the probability of mapping the user exiting with the new pseudonym, i' to an old pseudonym, j that runs the same continuous query. Therefore, for a user starting to execute a CQ from mix-zone m_1, if f_i users out of the m_i continuous query users in the i^{th} mix-zone execute the same CQ, then the entropy X of the user executing the continuous query is given by

$$X = log|k_1 - m_1| + \sum_{2 \leq i \leq n} log|f_i|$$

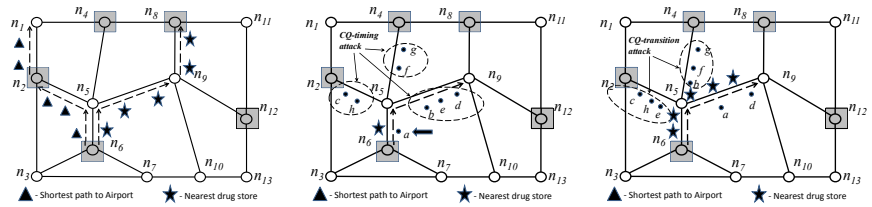
where $log|k_1 - m_1|$ represents the initial anonymity of the user while starting the continuous query in mix-zone, m_1.

The goal for designing CQ-attack resilient solutions is to increase the anonymity strengths of the mix-zones by considering the fact that the attacker has the continuous query correlation information at the intermediate mix-zones to infer and associate the CQ induced trajectory with its user. Note that the initial anonymity forms the major component of the anonymity under the CQ-attack model and therefore it is important that the mix-zones provide high initial anonymity for the continuous queries so that even when the attacker breaks the anonymity in the subsequent mix-zones, the initial anonymity remains sufficient to meet the required privacy level. For instance, if the first mix-zone in the above example provides a higher anonymity, say $k = 100$ instead of $k = 3$, then the initial anonymity may be sufficient to meet the

privacy requirements of the continuous query even though the attacker breaks the anonymity obtained in the intermediate mix-zones under the CQ-attack model. In the next sub-section, we discuss CQ-cloaking techniques (spatial cloaking or temporal cloaking of CQs) over road network mix-zones as a candidate approach for achieving higher query anonymity and show that it is ineffective and susceptible to attacks that combine CQ information with timing correlation (CQ-timing attack) and transition correlation (CQ-transition attack).

3.3 CQ-cloaking approach and its vulnerabilities

In the CQ-cloaking approach, the continuous queries are either temporally or spatially perturbed while the snapshot queries continue to be unperturbed. In the CQ-cloaking approach, the locations used by the CQ is perturbed such that a continuous query originating from a mix-zone is indistinguishable from at least k users traversing the mix-zone. While this technique does not make changes to the mix-zone model, we show that the location exposure of snapshot queries makes the CQ anonymization susceptible to CQ-timing attack and CQ-transition attack.



(a) CQ-cloaking (b) CQ-timing attack (c) CQ-transition attack

Fig. 3: Continuous Query: Timing and Transition attacks

Consider the example shown in Figure 3(a) where we have two CQs labeled as star CQ and triangle CQ respectively. The square nodes represent road network mix-zones. We observe two different CQ traces starting from the mix-zone at road junction n_6. The triangle trace crossed the junctions n_6, n_5 and n_2 and the star trace crossed the junctions n_6, n_5 and n_9 respectively and each star or triangle represents one snapshot execution of the corresponding CQ. Intuitively, if we delay the execution of the individual CQ snapshots of CQ users starting at mix-zone n_6 such that at least k_c users leave the mix-zone within the temporal delay, it will make it harder for an adversary to associate the CQ-induced trajectory with the corresponding CQ user. For instance, in Figure 3(a), if the continuous query on the shortest path to the airport (marked by stars) originating from the mix-zone n_6 is perturbed temporally in such a way that there are k or more users coming out of the mix-zone at road junction n_6 within the continuous query's temporal cloaking window, then from the attacker's perspective, the query could have originated from any of the k users who entered the mix-zone within the time window. CQ-spatial cloaking is similar to CQ-temporal

cloaking except that instead of delaying the snapshots of the continuous queries, the CQ exposes a larger spatial region such that there are k or more users within the spatial region.

CQ-timing Attack: As mentioned earlier, CQ-cloaking techniques are vulnerable to CQ-timing attack when users in the anonymity set violate the steady motion assumption, i.e., if all users do not travel at the imposed speed of the road segment. In the example shown in Figure 3(b), we find that users with pseudonyms a, b, c, d, e, f, g and h enter the mix-zone during the continuous query's temporal cloaking window, d_{tmax}. When the steady motion assumption fails, user a travels slowly and stays on segment $\overline{n_5 n_6}$ while other users move ahead of the segment, $\overline{n_5 n_6}$. If user a is the issuer of the continuous query, then the continuous query would stay on segment, $\overline{n_5 n_6}$ even though it is executed with a temporal delay while other users of the anonymity set move ahead. By observing this, the attacker can eliminate the low probable members and identify the issuer of the continuous query with high confidence. Concretely, we assume the attacker has knowledge of the maximum temporal delay, d_{tmax} used by the users. Let $A_{C(i)}$ represents the anonymity set of the continuous query, C_i, for each user, $j \in A_{C(i)}$ and let $M_{C_i,j}$ be the likelihood that the continuous query originates from user j. Since only continuous query's location is temporally perturbed, the attacker observes the movement of the users through their location exposure for snapshot queries. Let $loc(j,t)$ and $loc(C_i,t)$ represent the location of user j and the location of the temporally cloaked continuous query, C_i observed by the attacker at time t. If $X_j^{t-d_t}(loc(C_i,t))$ is a Boolean variable indicating if the attacker observed user j moving through the location, $loc(C_i,t)$ at time $t-d_t$, the likelihood $M_{C_i,j}$ is given by

$$M_{C_i,j} = \sum_{0 \le d_t \le d_{tmax}} q(C_i,d_t) \times X_j^{t-d_t}(loc(C_i,t))$$

where $q(C_i,d_t)$ denotes the probability that the continuous query, C_i uses a temporal lag d_t. Based on the proportion of the likelihoods of the users, the attacker can assign the probability, $Q_{C_i,j}$ that represents the probability of user j to be source of the continuous query, C_i.

$$Q_{C_i,j} = \frac{M_{C_i,j}}{\sum_{j \in A(C_i)} M_{C_i,j}}$$

Based on the probabilities, the attacker may either ignore the low probable members from consideration or narrow down the search to the high probable members. We note that similar attack is also possible in the case of CQ-spatial cloaking when the steady motion assumption fails.

CQ-transition Attack: Additionally, CQ-cloaking techniques are also prone to CQ-transition attack. When the transitions taken by a subset of the users in the anonymity set differ from that of the user executing the query, then those members can be eliminated based on transition correlation. For example, in Figure 3(c), at road intersection n_5, users c and e and h take a left turn on to the road segment, $\overline{n_2 n_5}$ whereas users b, f and g move straight on segment, $\overline{n_4 n_5}$ and users a and d turn right on to segment, $\overline{n_5 n_9}$. When the continuous query uses CQ-temporal cloaking or CQ-spatial cloaking and follows the querying user after a temporal delay or using a spatial cloaking region, from the transition taken by the continuous query from

segment, $\overline{n_6 n_5}$ to $\overline{n_4 n_5}$, the adversary will be able to eliminate the users, c, e, h, b, f and g from consideration as their transitions differ from that of the continuous query. Concretely, if Y_j is a Boolean variable indicating if the path taken by user j is the same as that of the continuous query, C_i, then the probability $Q_{C_i,j}$ of user j being the source of the continuous query, C_i is given by

$$Q_{C_i,j} = \begin{cases} \frac{1}{\sum_{j \in A(C_i)} Y_j} & \text{if } Y_j = 1 \\ 0 & \text{otherwise} \end{cases}$$

In the next section, we introduce the concept of delay-tolerant mix-zones that use a modified mix-zone model to perform both location mixing and identity mixing to achieve higher anonymity for continuous queries as compared to only identity mixing in the conventional mix-zone model. Delay -tolerant mix-zones perturb both continuous queries and snapshot queries and are free from CQ-timing and CQ-transition attacks.

4 Delay-tolerant Mix-zones

Delay-tolerant mix-zones combine mix-zone based identity privacy protection with location mixing to achieve high anonymity that is otherwise not possible with conventional mix-zones. In the delay-tolerant mix-zone model, users expose spatially or temporally perturbed locations outside the mix-zone area. However, on the exit of each delay tolerant mix-zone, the mix-zone changes their perturbed locations by introducing a random temporal or spatial shift to their already perturbed locations. While conventional mix-zones only change pseudonyms inside them, the additional ability of delay-tolerant mix-zones to change and mix user locations brings greater opportunities for creating anonymity.[1] Therefore, the anonymity strength of delay-tolerant mix-zones comes from a unique combination of both identity mixing and location mixing.

Before presenting the detailed analysis of the privacy strengths of the delay-tolerant mix-zones, we first illustrate the concept of delay-tolerant mix-zones with an example temporal delay-tolerant mix-zone. Table 1 shows the entry and exit time of users in a conventional rectangular road network mix-zone. We find that user a enters the mix-zone as $t = 100$ and exits at time $t = 104$. Similarly the other users enter and exit as shown in Table 1. Here the adversary may know that the average time taken by the users to cross the mix-zone is 4 sec. Therefore when user a exits at a' at time $t = 104$, the attacker can eliminate users e and f from consideration as they have not even entered the mix-zone by the time user a exits[2]. Similarly, the adversary can eliminate users o and n from consideration based on timing inference that users o and n have exited the mix-zone by the time a and b enter the mix-zone. Therefore

[1] This model is analogous to anonymous delay-tolerant routing in mix networks where network routers have the additional flexibility to create anonymity by delaying and reordering incoming packets before forwarding them [31].

[2] For the sake of example simplicity, we assume that the users take the average time of 4 sec to cross the mix-zone, in a real road intersection, it could actually take slightly longer or shorter time to cross based on the speed of travel.

when a exits as a', the attacker has uncertainty only among the users $\{a, b, c, d, m\}$. Also, among the users $\{a, b, c, d, m\}$, the attacker can eliminate more users through sophisticated reasoning based on timing inference described later.

User	t_{in}	t_{inside}	t_{out}
o	94	4	98
n	96	4	100
m	98	4	102
a	100	4	104
b	101	4	105
c	103	4	107
d	103	4	107
e	106	4	110
f	108	4	112

Table 1: Conventional Road Network Mix-zone

User	Observed t_{in}	t_{inside}	d_{told}	d_{tnew}	Observed t_{out}
w	81	4	4	20	105
v	84	4	7	20	108
u	84	4	7	20	108
s	87	4	10	20	111
r	89	4	12	20	113
q	90	4	13	20	114
p	92	4	15	20	116
o	94	4	18	20	118
n	96	4	19	20	120
m	98	4	20	18	120
a	100	4	4	16	120
b	101	4	4	15	120
c	103	4	7	13	120
d	103	4	7	13	120
e	106	4	10	10	120
f	108	4	12	8	120
g	109	4	13	7	120
h	111	4	15	5	120
i	113	4	18	3	120
j	115	4	19	1	120
k	117	4	20	0	120
l	118	4	20	0	121

Table 2: An example temporal Delay-tolerant mixing

However, in the delay-tolerant mix-zone model, each user uses a temporal delay, d_t within some maximum tolerance, d_{tmax}. Inside the mix-zone, the temporally perturbed location of each user is assigned a random temporal shift. In the delay-tolerant mix-zone example shown in Table 2, we find that user a initially uses a temporal delay, d_{told} of 4 sec and inside the mix-zone it is shifted randomly to 16 sec. Here d_{tmax} is assumed as 20 sec. Therefore when user a exits as a', it becomes possible that many users can potentially exit in the exit time of user a. The example in Table 2 shows one possible assignment of new temporal delays, d_{tnew} for other users in order for them to exit at the same time as a'. Thus, during the exit of user a as a', the attacker is confused to associate the exiting user a' with the members of the anonymity set, $\{a, b, c, d, e, f, g, h, i, j, k, m, n\}$. In principle, users' new temporal delays, d_{tnew} are randomly shifted inside the mix-zone ensuring the possibility of each of the users to exit at the exit time of each other and thus the delay-tolerant mix-zone model provides significantly higher anonymity compared to conventional mix-zones. Such high

anonymity provides the initial anonymity required for the continuous queries under the CQ-attack model.

Timing attack in delay-tolerant mix-zone: Before we proceed to analyze the privacy strengths of the delay-tolerant mix-zone, we formally define the timing attack described above. The attacker observes the time of entry, $t_{in}(i)$ and time of exit $t_{out}(i)$ for each user entering and exiting the mix-zone. When the attacker sees an user i' exiting, he tries to map i' to one of the users of the anonymity set, A_i. The attacker assigns a probability, $p_{i' \to j}$ that corresponds to the probability of mapping i' to j, where $j \in A$ based on the likelihood of user j exiting at the exit time of i', denoted by $t_{out}(i')$. Once the mapping probabilities are computed, the attacker can utilize the skewness in the distribution of the mapping probabilities to eliminate some low probable mappings from consideration and narrow down his inference to only the high probable mappings. Consider the example shown in Table 1 for the conventional mix-zone case, when user a exits as a', we have the anonymity set, $A = \{a, b, c, d, e, f, o, n, m\}$. Clearly the probability of mapping a' to the users $\{e, f, o, n\}$ is zero as they either enter the mix-zone after a' exits or leave the mix-zone before a even enters. However, between the users $\{a, b, c, d, m\}$, the attacker can assign a non-zero likelihood of resembling a' as they have been inside the mix-zone while a was present. Let these likelihoods be 0.37, 0.35, 0.1, 0.1 and 0.12 respectively. In this case, we show that it is easy to analyze the anonymity based on pairwise entropy. The pairwise entropy, $H_{pair}(a, b)$ between two users a and b during the exit of a' is the entropy obtained by considering a and b to be the only members of the anonymity set. In this case, we have two mapping probabilities, $p_{a' \to a}$ and $p_{a' \to b}$ which are computed as: : $p_{a' \to a} = \frac{0.37}{0.37+0.35} = 0.513$, $p_{a' \to b} = \frac{0.35}{0.37+0.35} = 0.486$. Hence the pairwise entropy $H(a, b)$ is given by

$$H_{pair}(a, b) = -(p_{a' \to a} log p_{a' \to a} + p_{a' \to b} log p_{a' \to b})$$

Therefore, $H_{pair}(a, b)$ = 0.99. In general, a mix-zone defines a short mix-zone time window, τ such that a set of users entering within the short time window, τ have high pairwise entropy with each other. In the above example, if we take $\tau = 1$ sec, then we find users a and b enter within 1 sec and hence they have high pairwise entropy close to 1. However, users a and c do not enter within the mix-zone time window of 1 sec and therefore we find that their pairwise entropy is lower($H_{pair}(a, c)$ = 0.75). In comparison, a theoretical mix-zone (recall Definition 1) ensures a uniform distribution for all possible mappings between old and new pseudonyms and therefore ensures a high pairwise entropy of 1.0 for all pairs of users in the anonymity set. Thus, the effective anonymity set of a road network is assumed to comprise of only those users that have high pairwise entropy with each other.

Definition 3 A delay-tolerant road network mix-zone offers k-anonymity to a set A of users if and only if:

1. There are k or more users in the anonymity set A.
2. Given any two users $i, j \in A$ and assuming i exiting at time t, the pairwise entropy after timing attack should satisfy the condition: $H_{pair}(i, j) \geq \alpha$.
3. Given any two users $i, j \in A$, the pairwise entropy after transition attack should satisfy the condition: $H_{pair}(i, j) \geq \beta$.

In the above example, only users $\{a, b\}$ will belong to the effective anonymity set of a under the conventional mix-zone model as only they have high pairwise Entropy with each other. However as discussed earlier, under the delay-tolerant mix-zone model more number of users ($\{a, b, c, d, e, f, g, h, i, j, k, m, n\}$) will have high pairwise entropy with each other and belong to the anonymity set of user a. Here, in addition to the pairwise entropy with respect to timing attack, the pairwise entropy with respect to transition attack is also considered for cases where the road intersections do not have uniform transition probabilities to different segments (Condition 3). In such cases, the effective anonymity set will contain only those members who enter from road segments such that their transition probability to the exit segment of user a is similar to the transition probability of their exit segments. As the delay-tolerant mix-zones primarily influence the impact of timing attack, we focus our discussion based on timing attack. However, in our experiments using real road networks, we take into account the fact that road intersections do not have uniform transition probability and accordingly construct the anonymity set with only those members which have similar transition probability as discussed in [29]. Next, we present the design and formal analysis of three proposed delay-tolerant mix-zone techniques namely (i)temporal delay-tolerant mix-zones, (ii) spatial delay-tolerant mix-zones and (iii) spatio-temporal delay-tolerant mix-zones.

4.1 Temporal Delay-tolerant Mix-zones

As discussed before, in a temporal delay tolerant mix-zone, every mobile user delays the location exposure with a randomly chosen delay, d_t within the maximum temporal tolerance, d_{tmax}. The temporal time window, d_{tmax} is chosen based on the arrival rate of the users in the road junction so as to ensure an expected number of users arriving into the mix-zone within the temporal tolerance, d_{tmax}. Note that the random delay used by a mobile client does not change during its travel between mix-zones. Only when the mobile client enters a new mix-zone, its temporal delay is randomly shifted to a new value within the temporal window, d_{tmax}. Based on the delayed exposure of users' location information, the attacker knows their current temporally cloaked location.

Based on the temporally cloaked location exposed by the users, the adversary observes each user i entering the mix-zone at a temporally cloaked time $tcloak_{in}(i)$ and exiting at a temporally cloaked time $tcloak_{out}(i')$ with a new pseudonym i'. The speed followed by the users in a road segment is assumed to follow a Gaussian distribution with a mean μ and standard deviation σ, where μ and σ are specific to each road class category. For user i, the set of all other users who had entered the mix-zone during the time window defined by $|tcloak_{in}(i)-\tau-d_t|$ to $|tcloak_{in}(i)+\tau+d_t|$, forms the anonymity set of i, denoted as A_i where τ is a small value and represents the mix-zone time window. Let t be the temporally cloaked exit time of user i, which is also $tcloak_{out}(i')$.

For each user, j in the anonymity set, A_i, we compute $p_{i'\rightarrow j}$, the probability that the exiting user i' at temporally cloaked time, $tcloak_{out}(i')$ is j and $p_{i'\rightarrow i}$ be the probability that the exiting user is i. Let $P(j, t)$ define the probability that user j exits

the mix-zone at the cloaked time, t. Here, the observed movement of the temporally cloaked user location in the mix-zone is controlled by two factors: the speed of the user inside the mix-zone and the change in the temporal delay used by the user. Let $P_v(j,x)$ be numerically equal to the probability that user j takes x units of time to traverse the mix-zone region. It is computed based on the speed distribution on the road segments. Let $P_t(j,t)$ be the probability that the temporal delay of user j is shifted by t seconds after exiting the mix-zone. In order for user j to exit at temporally cloaked time, $tcloack_{out}(i')$, it depends on both the temporal shift introduced to user j in the mix-zone as well as the time j takes to cross the mix-zone. Thus, $P(j,t)$ is given by

$$P(j,t) = \int_0^\infty P_v(j,x) \times P_t(j, t - tcloack_{in}(j) + x)dx$$

Here we note that our temporal location mixing algorithm assigns a new temporal delay, d_{tnew} based on the current temporal delay, d_t and the maximum temporal delay, d_{tmax}.

$$d_{tnew} = |d_{tmax} - d_t|$$

and hence it ensures a uniform distribution of shift values, $P_t(j,t)$ while also ensuring that the temporal lags of the users are uniformly distributed. Similar to $P(j,t)$, we can also obtain $P(i,t)$ based on the temporally cloaked arrival time of i, $tcloack_{in}(i)$. We have

$$P(i',t) = P(i,t) + P(j,t)$$

Here, $P(i',t)$ is the cumulative likelihood of both i or j exiting as i' as i' is either of them. Therefore, the probability of i' being j when i' exits at time t, denoted as $p_{i' \to j}(t)$ is given by the following conditional probability

$$p_{i' \to j}(t) = P((j,t)/(i',t))$$

Similarly, the probability of i' being i, $p_{i' \to i}(t)$ is given by

$$p_{i' \to i}(t) = P((i,t)/(i',t))$$

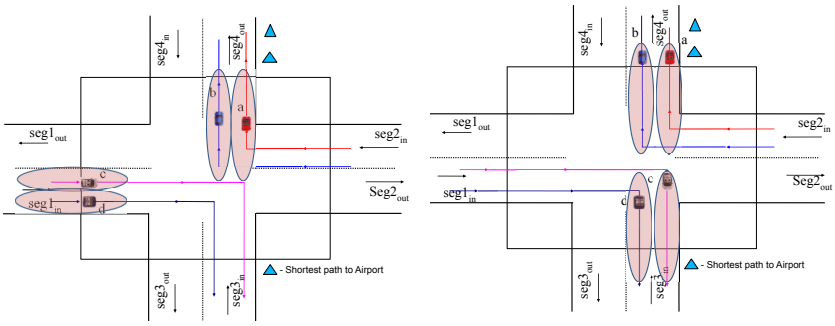
and the pair-wise entropy after timing attack, $H_{pair}(i,j)$ between users i and j when i exits as i' can be obtained as

$$H_{pair}(i,j) = -(p_{i' \to i}(t) log p_{i' \to i}(t) + p_{i' \to j}(t) log p_{i' \to j}(t))$$

As it is intuitive, for an exiting user, i', the number of users in the effective anonymity set (i.e., those members that have high pairwise entropy with each other and with i') is directly proportional to the temporal tolerance, d_{tmax}, ie., the greater the temporal tolerance value, d_{tmax}, the more the number of users that could possibly resemble i during the exit of i' with a high pairwise Entropy. Thus by varying the temporal tolerance, d_{tmax} the temporal delay tolerant-mix-zones can offer any desired level of anonymity to the users.

4.2 Spatial Delay-tolerant Mix-zones

We now present our second class of delay-tolerant mix-zones namely spatial delay-tolerant mix-zones. Unlike the temporal delay-tolerant mix-zones, in the spatial delay-tolerant mix-zone approach, users' locations are instantaneously sent out using a spatial region instead of the exact point location. Here, the spatial region masks the exact time of traversal of the user inside the mix-zone ensuring the possibility that the user could be located at any point within the spatial region. This ensures that the adversary can not infer the exact time of traversal of the user. The spatial region is constructed by first identifying the temporal window size, d_{tmax} based on the arrival rate of the users in the mix-zone and by translating the user's current location into a spatial region based on the temporal window size, d_{tmax}. A spatial region corresponding to a temporal window size d_{tmax} includes all road segments that can be reached within d_{tmax} units of time (i.e., the corresponding d_l units of length) from the center of the region when travelled at the mean speed of the road segments. The delay-proportional spatial cloaking algorithm described in Algorithm 1 computes the spatial region in such a way that the distance from the center of the spatial region and the location of the mobile user exactly corresponds to the spatial distance, d_l proportional to the temporal delay, d_t of the user in the temporal delay-tolerant approach. Inside the delay-tolerant mix-zone, the spatial regions of the users are randomly changed by introducing a spatial shift.



(a) Before spatial location-mixing (b) After spatial location-mixing

Fig. 4: Illustration of Delay-tolerant mix-zones

We illustrate the principle of spatial delay-tolerant mix-zone through an example in Figure 4(a) and Figure 4(b). Figure 4(a) shows the entry of the spatial regions of users a, b, c and d into the mix-zone and Figure 4(b) illustrates the spatial location mixing process. We find that the location mixing process changes the spatial regions of the users, a, b, c and d in such a way that the distance of the users from the center of their spatial region is randomly shifted (Notice changed distance in the spatial regions in Figure 4(b)). Therefore, after the spatial location mixing process, the spatial regions of users a, b, c and d all have similar probability to exit at a given time. Note that without this spatial mixing, the attacker would still infer that the

regions of a and b entered well ahead of c and d and hence a and b will exit before c and d.

Algorithm 1 Delay-proportional Spatial Cloaking

```
1: d_tmax: continuous query temporal window
2: v: vertex v corresponds to the road junction v
3: pathtime_v: mean travel time to vertex v from starting mix-zone, if v is the mix-zone junction, then
   pathtime_v = 0
4: region: is a global list of road segments representing the cloaking region. It is empty at beginning
   and represents the cloaking region when the algorithm terminates
5: procedure FINDCLOAKREGION(d_tmax, pathtime_v, v)
6:     for all segments(v, u) ∈ segs(v) do
7:         pathtime_u = pathtime_v + length(v,u)/speed(v,u)
8:         if (pathtime_u < d_tmax) then
9:             if (region.contains(v, u) == false) then
10:                region.add(v, u)
11:                FindCloakRegion(d_tmax, pathtime_u, u)
12:            end if
13:        end if
14:    end for
15: end procedure
```

Here we assume that each user, i's spatial region enter the mix-zone at time, $t_{in}(region(i))$ and exits at time $t_{out}(region(i))$ with a new pseudonym, i'. For each user, i, the set of other users whose spatial regions entered the mix-zone during the time window defined by $|t_{in}(region(i)) - \tau|$ to $|t_{in}(region(i)) + \tau|$ forms the anonymity set of i, denoted by A_i. The anonymity obtained in the mix-zone is dependent on the pairwise mapping probabilities, $p_{i' \to i}$ and $p_{i' \to j}$ where $j \in A_i$ and $p_{i' \to j}$ denotes the probability that the exiting user i' at time, $t_{out}(region(i'))$ is j and $p_{i' \to i}$ represents the probability that the exiting user is i.

The movement of a user's spatial region is governed by two factors in a spatial delay-tolerant mix-zone namely the randomness of the user movement in terms of its velocity inside the mix-zone and the spatial shift introduced in the spatial region after exiting the mix-zone. Let us assume that user j takes x units of time in the mix-zone and let $P_v(j, x, l)$ be the probability that user j travels with a velocity such that it takes x units of time to cross l units of distance of the mix-zone region. Let l_{mix} be the distance to cross the mix-zone and let us assume that j's spatial region is shifted by y units of length in the location mixing process. Let $P_s(j, y)$ denote the probability that the spatial shift introduced in the mix-zone is y units of length. Therefore, for user j to exit at time, $t_{out}(region(i'))$, it should travel a distance of $l_{mix} + y$ in the mix-zone instead of the usual distance l_{mix}. Therefore, we have

$$P(j, t) = \int_0^\infty P_v(j, x, l_{mix} + y) \times P_s(j, y) dx$$

where $x = t - t_{in}(region(j))$. Here, the spatial location mixing process assigns a new spatial region to each user ensuring a uniform distribution of spatial shift values,

$P_s(j, y)$. i.e., inside the mix-zone, every user's distance from the center of its spatial region is shifted by a random length, y.

Similar to $P(j, t)$, we can obtain $P(i, t)$ and hence $p_{i' \to j}(t)$, the probability of i' being j when i' exits at time t can be obtained from the conditional probability, $p_{i' \to j}(t) = P((j, t)/(i', t))$. Similarly, as discussed in section 4.1, the probability of i' being i, $p_{i' \to i}(t)$ can be computed and thus the pair-wise entropy, $H_{pair}(i, j, t)$ between users i and j when i exits as i' can be obtained.

We note that the spatial delay-tolerant mix-zone approach does not incur temporal delays, however they lead to higher query processing cost that is directly proportional to the size of the spatial regions. In the next subsection, we discuss spatio-temporal delay-tolerant mix-zones that yield suitable tradeoffs between the incurred delay and the cost of query processing.

4.3 Spatio-temporal delay-tolerant Mix-zones

In the spatio-temporal delay-tolerant mix-zone approach, user locations are perturbed using both a temporal delay as well as a spatial region instead of the exact point location and the mix-zone introduces both random temporal and spatial shifts to the spatio-temporally perturbed user locations. Therefore, to an adversary observing an user, the user could have been located at any point in the spatial region at any instance of time during the temporal time window.

The anonymity strength of this model is analyzed as follows. We assume that each user uses a temporal time window of d_{tmax} and a spatial cloaking region of length, d_{lmax}. Based on the delayed exposure of the spatial regions, the attacker estimates the current temporally perturbed position of the spatial region of each user's location. Here we assume that each user, i's spatial region, $region(i)$, enters the mix-zone at a temporally cloaked time, $tcloak_{in}(region(i))$ and exits at a temporally cloaked time, $tcloak_{out}(region(i))$ with a new pseudonym, i'. For user i, the set of all other users whose spatial regions entered the mix-zone during the mix-zone time window defined by $|tcloak_{in}(region(i)) - \tau - d_t|$ to $|tcloak_{in}(region(i)) + \tau + d_t|$, forms the anonymity set of i, namely A_i.

In a spatio-temporal delay-tolerant mix-zone, the movement of a user's spatio-temporally perturbed location is governed by three factors: (i) the randomness of user's movement inside the mix-zone determined by velocity, (ii) the random temporal shift introduced inside the mix-zone and (iii) the random spatial shift introduced inside the mix-zone. For all $j \in A_i$, let us assume that user j's spatial region, $region(j)$ takes x units of time to cross the mix-zone and let $P_v(j, x, l)$ be the probability that user j travels with a velocity such that it takes x units of time to cross l units of distance of the mix-zone region and let j's spatial region be shifted by y units of length and its temporal delay be shifted by z units of time. Let $P_s(j, y)$ denote the probability that the spatial shift introduced in the mix-zone is x units of length and $P_t(j, z)$ be the probability that the temporal delay of user j is shifted by z seconds after exiting the mix-zone. Therefore, for user j's spatial region, $region(j)$ to exit at temporally cloaked time, $t_{out}(region(i'))$, it should travel a distance of $l_{mix} + y$

in the mix-zone instead of the usual l_{mix} where l_{mix} is the length of the mix-zone region. Therefore,

$$P(j,t) = \int_0^\infty P_v(j,x,l_{mix}+y) \times P_s(j,y) \times P_t(j,t-tcloack_{in}(j)+x)dxdy$$

Here we note that the spatio-temporal delay-tolerant mix-zone introduces both random temporal and spatial shifts inside and thus it ensures a uniform random distribution of $P_s(j,y)$ and $P_t(j,t-tcloack_{in}(j)+x)$ in the above equation.

Thus, the pair-wise entropy between users i and j when i exits as i' can be obtained after knowing the probabilities $p_{i'\to i}(t)$ and $p_{i'\to j}(t)$ which are deduced from $P(i,t)$ and $P(j,t)$ similar to the analysis on temporal delay-tolerant mix-zones in section 4.1

5 Experimental Evaluation

We divide the experimental evaluation of our techniques into three components: (i) the effectiveness of the proposed techniques under the CQ-attack model, (ii) performance in terms of query processing cost, incurred temporal delays and success rate of anonymization and (iii) evaluation of the spatio-temporal tradeoffs between incurred temporal delays and query processing cost. Before reporting our experimental results, we first describe the experimental setup, including the road-network mobile object simulator used in the experiments.

5.1 Experimental setup

We use the GT Mobile simulator [28] to generate a trace of cars moving on a real-world road network, obtained from maps available at the National Mapping Division of the USGS [2]. The simulator extracts the road network based on three types of roads – *expressway*, *arterial* and *collector* roads. Our experimentation uses maps from three geographic regions namely that of Chamblee and Northwest Atlanta regions of Georgia and San Jose West region of California to generate traces for a two hour duration. We generate a set of 10,000 cars on the road network that are randomly placed on the road network according to a uniform distribution. Cars generate random trips with source and destination chosen randomly and shortest path routing is used to route the cars for the random trips. The speed of the cars are distributed based on the road class categories as shown in Table 3.

5.2 Experimental results

Our experimental evaluation consists of three parts. First, we evaluate the effectiveness of the proposed techniques in terms of their anonymization effectiveness. We compare the delay-tolerant mix-zone techniques with CQ-cloaking techniques and conventional road network mix-zones in terms of the obtained entropy that captures

Road type	Expressway	Arterial	Collector
Mean speed(mph)	60	50	25
Std. dev.(mph)	20	15	10
Speed Distribution	Gaussian	Gaussian	Gaussian

Table 3: Motion Parameters

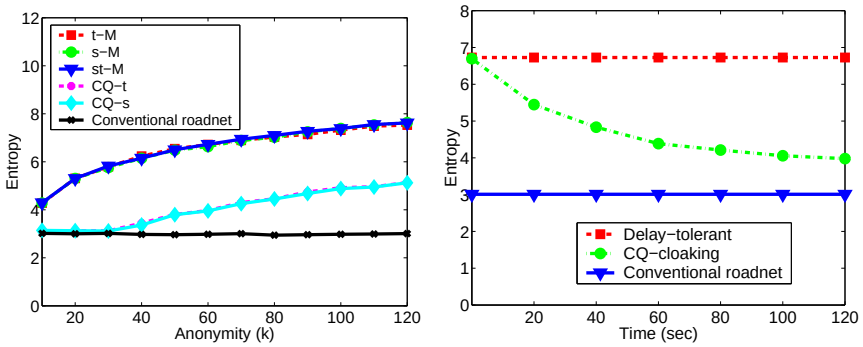
Parameter	Value
Map	Northwest Atlanta region
Mobility Model	Random Roadnet Router
Total number of vehicles	10000
Number of Road junctions	6831
Number of Road segments	9187

Table 4: Simulation Parameters and Setting

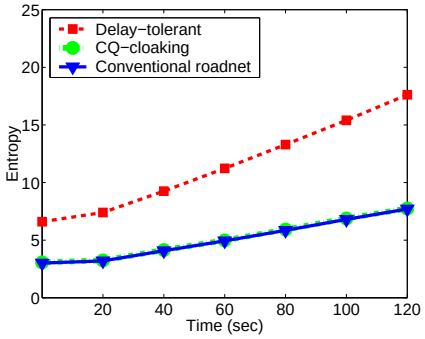
the amount of information required to break the anonymity and then compare the various delay-tolerant mix-zones in terms of the average temporal delays incurred and the cost of query processing in terms of query processing time. Next, we evaluate the spatio-temporal tradeoffs of delay-tolerant mix-zones that helps understand the best tradeoffs in terms of the incurred temporal delay and the query processing time. Our final set of experiments evaluate the effectiveness of the delay-tolerant mix-zones in terms of the average temporal delays, query execution time and success rate in providing the desired value of k. Our default setting uses the map of Northwest Atlanta that has 6831 road junctions and 9187 road segments as shown in table 4. Among the 6831 junctions in the road network, 1025 (15 %) road junctions are chosen as the candidate mix-zones and the experimental results are averaged among these mix-zones. By default, each delay-tolerant mix-zone is constructed over a non-rectangular road network mix-zone with pairwise Entropy lowerbound after timing attack, α taken as 0.9 so that the effective anonymity set of the mix-zone comprises only of users who have pairwise Entropy greater than 0.9 with each other. Similarly for comparison, each conventional mix-zone is also constructed using the non-rectangular geometry with $\alpha = 0.9$. For both delay-tolerant mix-zones and conventional mix-zones, the pairwise Entropy lowerbound β after transition attack is taken as 0.9. We assume that all continuous queries are unique and by default 10% of users in the system run continuous queries.

5.2.1 Comparison with Conventional Mix-zones and CQ-cloaking

This set of experiments compares the delay-tolerant mix-zone approaches with the conventional mix-zones and CQ-cloaking techniques in terms of their anonymity measured by Entropy. Here, the delay-tolerant mix-zones are constructed over a conventional road network mix-zone whose size is chosen to offer an anonymity of 4. In Figure 5(a), we compare the average entropy of the temporal, spatial and spatio-temporal delay-tolerant mix-zone approaches (t-M, s-M, and st-M) with the conventional mix-zone approach and the temporal and spatial CQ-cloaking approaches (CQ-t and CQ-s) for various values of required anonymity, k. Here, the temporal



(a) Entropy after CQ-attack (b) Entropy Timeline - CQs

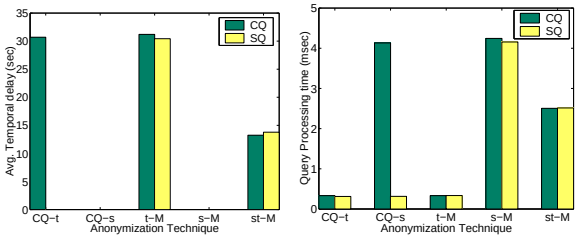


(c) Entropy Timeline - SQs

Fig. 5: Comparison with Conventional Mix-zones

window and spatial region size are chosen based on the arrival rate of the users in the mix-zones to ensure the required number of users, k with a high probability, $p = 0.9$. For the spatio-temporal delay-tolerant mix-zones, the spatial region size is fixed as 800 m and the temporal window is varied according to the required value of k. We find that the average entropy of the conventional mix-zone approach is significantly lower than that of the delay-tolerant mix-zones as they can not adapt to higher levels of anonymity but the delay-tolerant mix-zones always provide the required anonymity level for all values of k as shown by the high Entropy. Here, we also note that the CQ-cloaking approaches (CQ-t and CQ-s) have low level of Entropy due to the effect of CQ-timing and CQ-transition attacks. In Figure 5(b) and Figure 5(c), we plot the timeline of the Entropy obtained by continuous queries (CQ) and snapshot queries (SQ) respectively. Here, we use the spatio-temporal mix-zone as the candidate delay-tolerant mix-zone and temporal CQ-cloaking as the candidate CQ-cloaking technique. We find that with conventional mix-zones, the continuous queries obtain low initial anonymity and it stays constant throughout the timeline. With the CQ-cloaking approach, the queries obtain higher anonymity in the beginning but their anonymity is gradually reduced due to the impact of CQ-timing and CQ-transition attacks. However, the delay-tolerant mix-zones offer very high anonymity to meet

the privacy requirements of the continuous queries under the CQ-attack model. For snapshot queries, we find that the techniques have a different trend as shown in Figure 5(c). The conventional mix-zone model shows an increasing Entropy timeline where users gain more anonymity at the intermediate mix-zones as CQ-attack has no impact on snapshot queries. We also find that the delay-tolerant mix-zone offers greater anonymity to snapshot queries with a much steeper Entropy timeline but the CQ-cloaking technique offers only similar anonymity as the conventional mix-zone.

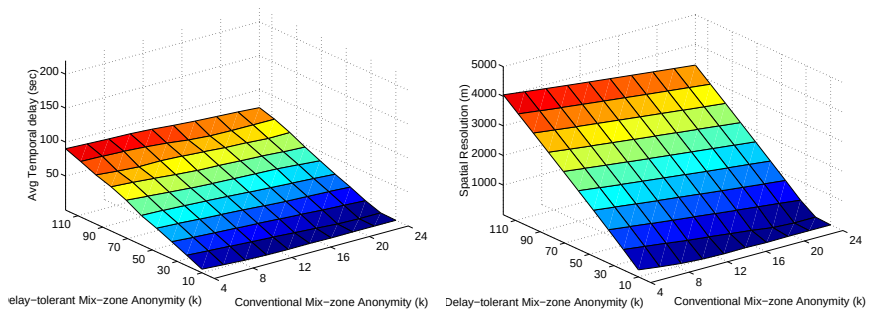


(a) Average temporal delay (b) Query Cost

Fig. 6: Performance of Continuous and SnapShot queries

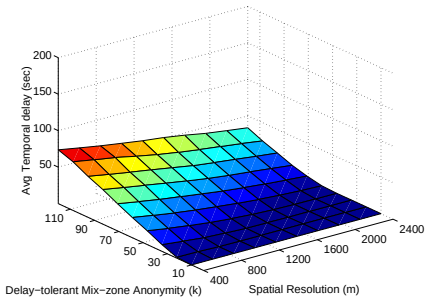
5.2.2 Performance of Continuous and Snapshot queries

Next, we study the performance impact of the proposed approaches for continuous and snapshot queries individually. We measure the average temporal delay incurred and the average query execution time of the techniques in figure 6(a) and 6(b). Here, all queries are anonymized corresponding to a $k = 50$. The spatio-temporal delay-tolerant mix-zone uses its default spatial region size of 800 m. The query execution time represents the average time to process a snapshot of a k-NN query for a k-nearest neighbor value of ($k_q = 7$) over 14000 uniformly distributed objects on the road network using the road network based anonymous query processor described in [32]. We find that with the CQ-temporal cloaking (CQ-t), only the continuous queries incur delay before getting processed and in CQ-spatial cloaking (CQ-s), neither of the queries incurs any delay. With the spatial cloaking approach, we obtain the results of the query for all possible locations within the cloaking region, however the continuous queries in the CQ-spatial cloaking approach result in higher query execution time. In temporal delay-tolerant mix-zones (t-M), both continuous and snapshot queries incur temporal delays but have low query execution time. Conversely, the spatial delay-tolerant mix-zones (s-M) do not incur any delays for the queries but have increased query execution time for both snapshot and continuous queries. The spatio-temporal delay-tolerant mix-zones technique (st-M) finds a tradeoff between these approaches and has more than 55% lower average temporal delay compared to the temporal cloaking case as well as a 40% lower query execution time compared to the spatial delay-tolerant mix-zones.



(a) t-M

(b) s-M



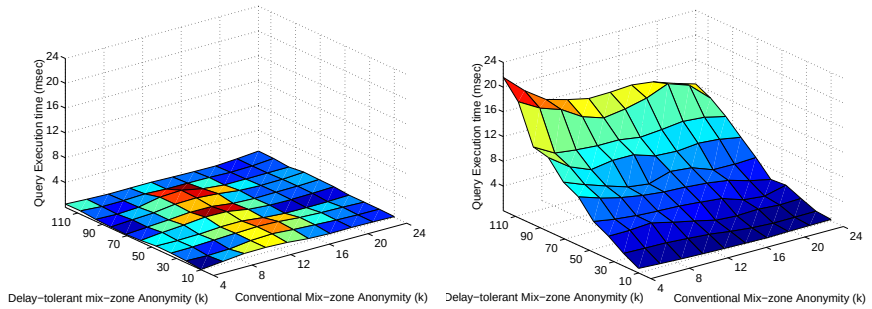
(c) st-M

Fig. 7: Performance of delay-tolerant mix-zones

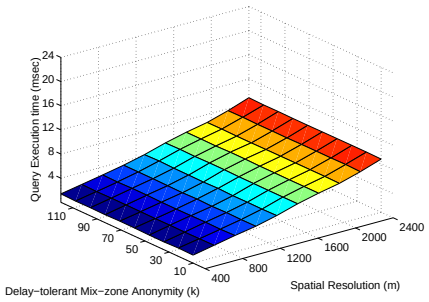
5.2.3 Evaluating Spatio-temporal tradeoff

Our next set of experiments compares the performance of the delay-tolerant mix-zones in terms of the average temporal delays incurred and the query execution time under various values of required anonymity, k. Figure 7(a) shows the average temporal delay required to anonymize users for various anonymity levels. The X-axis shows the anonymity offered by the conventional mix-zone and the Y-axis shows the anonymity level offered by the delay-tolerant mix-zones constructed over the conventional mix-zones and the Z-axis represents the average temporal delay, d_t used by the delay-tolerant mix-zones. We find that the temporal delay increases with increase in the anonymity level of the delay-tolerant mix-zone. Also, we find that there is only a small decrease in temporal delay with increase in mix-zone anonymity as the temporal window has a significant impact on the obtained anonymity.

Similarly, we study the performance of spatial delay-tolerant mix-zones in Figure 7(b). The X-axis represents the anonymity offered by the conventional road network mix-zone and the Y-axis represents the anonymity of the delay-tolerant mix-zones. We plot the required spatial resolution along Z-axis. We observe a similar trend with spatial resolution values as with the temporal delay in Figure 7(a). We notice that higher query anonymity levels require larger spatial cloaking regions for location perturbation. In the context of delay-tolerant mix-zones, we measure the data quality in terms of the size of the spatial region (spatial resolution) instead of the number



(a) t-M (b) s-M



(c) st-M

Fig. 8: Comparison of Query Execution time

of objects as those in the location k-anonymized cloaking [13]. This is because the spatial perturbation in a delay-tolerant mix-zone is focused on changing the pseudo-identity with a higher anonymity rather than obtaining a perturbed location with a k-anonymized cloaking region. However, we refer the interested readers to [13] for additional object distribution based metrics for measuring data quality of the location perturbation process.

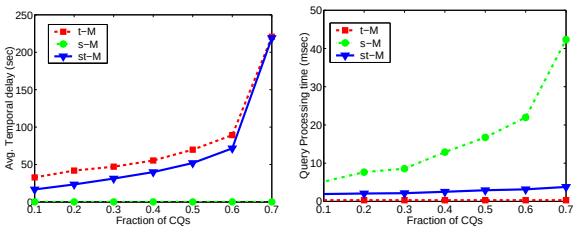
Next, we study the spatio-temporal tradeoff between the spatial resolution that determines query processing cost and the temporal delay incurred before processing the query requests. In Figure 7(c), the spatial resolution value is varied along the X-axis and the Y-axis represents the delay-tolerant mix-zone anonymity level. Here, the conventional mix-zone anonymity is set as 12. We find that the average temporal delay along Z-axis is much smaller with the effect of spatio-temporal perturbation compared to the temporal and spatial delay-tolerant mix-zones. Higher spatial resolution greatly reduces the temporal window size required to provide the required anonymity. However, the right trade-off between the temporal window size and the spatial resolution can be made based on the acceptable delay in processing the user queries and the desired cost of query processing.

Our next set of experiments compares the delay-tolerant mix-zones in terms of their average query execution time. In Figure 8, the average query execution time to process a snapshot of a k-NN query with ($k_q = 7$) over 14000 uniformly distributed objects on the road network is shown. In Figure 8(a), observe that the temporal delay-

tolerant mix-zones incur very low query execution time (less than 1 msec) as these queries use the point location of the mobile client with a temporal delay. Whereas, the query execution time of spatial delay-tolerant mix-zones in Figure 8(b) increases with increase in the required anonymity as larger anonymity requires larger spatial regions for processing. The query execution time of spatio-temporal delay-tolerant mix-zone approach in 8(c) shows the reduction in the query processing cost at the expense of the incurred temporal delay in processing the queries (Figure 7(c)).

5.2.4 Impact of fraction of Continuous Query users

The anonymity of the delay-tolerant mix-zones also depends on the fraction of total users who run continuous queries. For instance, if a number of continuous queries exist in the system, each uniquely identifying the querying user, the anonymization process might result either in long temporal delays or large spatial regions leading to higher query processing cost. The continuous query fraction denotes the fraction of the total users who currently execute a continuous query. Here, each query is anonymized with an anonymity, $k = 50$. In Figure 9(a), we measure the average temporal delay incurred by the approaches for varying values of fraction of continuous queries. We find that the average temporal delay of both (t-M) and (st-M) approaches increases steadily with increase in the proportion of continuous queries till a point, (0.6 in the figure) and then increases steeply indicating that it could be expensive to anonymize continuous queries if more than 60% of the users execute continuous queries. A similar trend is exhibited in Figure 9(b) for query processing cost suggesting that the anonymization cost could be higher when the fraction of continuous queries approaches 0.7. As a large number of mobile users are passive in general (i.e., execute no queries while traveling), we believe that a 60% continuous query fraction of the entire user population is expected to be rarely crossed in practice.



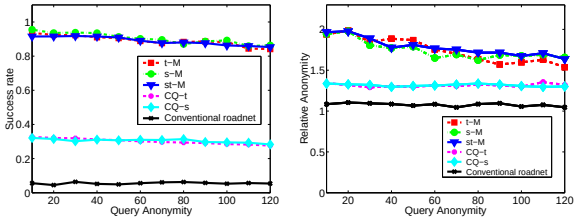
(a) Average Temporal delay (b) Query processing time

Fig. 9: Fraction of Continuous Query users

5.2.5 Success Rate and Relative-k

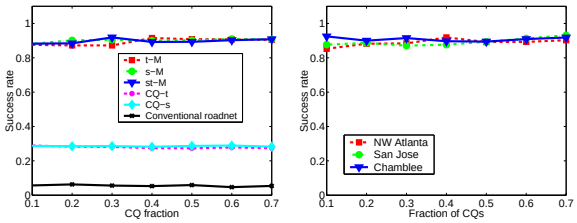
Our final set of experiments evaluates the performance of the delay-tolerant mix-zones in terms of their success rate in providing the desired level of anonymity. The

success rate represents the fraction of the cases where the proposed framework is able to provide an anonymity equal or greater than the requested value, k. In Figure 10(a), the query anonymity level is varied along the X-axis and the Y-axis represents the obtained success rate. Based on the arrival rate of the users in the mix-zone, the expected success rate is chosen as 0.9 so that the delay-tolerant mix-zones provide an anonymity of k or higher in more than 90% of the cases. We find that all the delay-tolerant mix-zone techniques obtain a success rate close to the expected success rate of 0.9, however the success rate of the CQ-cloaking approach is much lower (less than 0.3) and the conventional mix-zone approach has a even lower success rate of less than 0.06. Similarly, we study relative-k which is defined as the ratio of the anonymity obtained by the queries to the query anonymity requested. In Figure 10(b), we find that the relative anonymity level of delay-tolerant mix-zones ranges from 1.5 to 2.0 showing that the queries on an average obtain an anonymity which is 1.5 to 2.0 times the requested value. The successfull cases of CQ-cloaking and conventional mix-zone approaches have a lower relative anonymity as the mix-zones have lower successrate and provides lower value of k in general. In order to evaluate



(a) Success Rate (b) Relative k

Fig. 10: Success rate and Relative k



(a) Varying CQ fraction (b) Varying Maps

Fig. 11: Success rate

the success rate under different fractions of continuous queries in the system, we study the approaches by varying the fraction of users executing continuous queries. Here, each query is anonymized with an anonymity of 50. Figure 11(a) shows that the obtained success rate is close to the expected success rate for the delay-tolerant

mix-zones across different fractions of CQs in the system. However, the CQ-cloaking and conventional mix-zone techniques have much lower success rate. Similarly, the success rate of the techniques is compared across different scales of geographic maps described in section 5.1. We compare the success rate of spatio- temporal delay-tolerant mix-zones in Figure 11(b) that shows that the technique performs well across different geographic maps.

6 Conclusion

We presented a delay-tolerant mix-zone framework for protecting location privacy of mobile users against continuous query correlation attacks. First, we described and formally analyzed the mix-zone anonymization problem under the CQ-attack model and showed that spatial cloaking or temporal cloaking over road network mix-zones is ineffective and susceptible to CQ-timing and CQ -transition attacks. We introduced three types of delay-tolerant road network mix-zones that are free from CQ-timing and CQ-transition attacks and in contrast to conventional mix-zones, perform a combination of both location mixing and identity mixing of spatially and temporally perturbed user locations to achieve stronger anonymity for the continuous queries under the CQ-attack model. The temporal delay tolerant mix-zones perform temporal location-mixing by introducing a random temporal shift and the spatial delay tolerant mix-zones perform spatial location-mixing through random spatial shifts. We also showed that by combining temporal and spatial delay tolerant mix-zones, we can make acceptable tradeoffs between anonymous query processing cost and temporal delay incurred in anonymous query processing. Extensive experiments using traces generated by GTMobiSim on different scales of geographic maps show that the delay-tolerant mix-zones are effective under the CQ attack model and offer the required level of anonymity to the continuous queries. In future, we plan to investigate the mix-zone anonymization under more sophisticated attack models that combine prior background knowledge of the users travel patterns and personal locations along with CQ information.

References

1. J.R. Cuellar, J.B. Morris, D.K. Mulligan, J. Peterson and J. Polk. Geopriv requirements. *IETF Internet Draft*, 2003.
2. U.S. Geological Survey. http://www.usgs.gov.
3. USAToday. Authorities: Gps systems used to stalk woman. http://www.usatoday.com/tech/news/2002-12- 30-gps-stalker x.htm.
4. C. Aggarwal. On k-Anonymity and the Curse of Dimensionality. In *VLDB*, 2005.
5. S. Amini, J. Lindqvist, J. Hong, J. Lin, E. Toch and N. Sadeh Cache': Caching Location-Enhanced Contentto Improve User Privacy In *Mobisys*, 2011.
6. C. Ardagna, M. Cremonini, S. Vimercati, P. Samarati. An Obfuscation-Based Approach for Protecting Location Privacy. In *IEEE TDSC*, 2011.
7. B. Bamba, L. Liu, P. Pesti, and T. Wang. Supporting Anonymous Location Queries in Mobile Environments with PrivacyGrid. In *WWW*, 2008.
8. R. Bayardo and R. Agrawal. Data Privacy Through Optimal k-Anonymization. In *ICDE*, 2005.
9. A. Beresford and F. Stajano. Location Privacy in Pervasive Computing. *Pervasive Computing, IEEE*, 2003.

10. C. Bettini, S. Mascetti, X. Wang, D. Freni, and S. Jajodia. Anonymity and Historical-Anonymity in Location-Based Services In Privacy in Location-Based Applications: Introduction, Research Issues and Applications, Lecture Notes of Computer Science 5599, Springer, 2009.
11. L. Buttyan and T. Holczer and I. Vajda. On the effectiveness of changing pseudonyms to provide location privacy in VANETs In *ESAS*, 2007.
12. C. Chow and M. Mokbel Enabling Private Continuous Queries For Revealed User Locations. In *SSTD*, 2007.
13. C. Chow, M. Mokbel, J. Bao and X. Liu. Query-aware location anonymization for road networks. In *Geoinformatica*, July 2011.
14. R. Dewri, I Ray, I. Ray and D. Whitley. Query m-Invariance: Preventing Query Disclosures in Continuous Location-Based Services. In *MDM*, 2010.
15. J. Freudiger, M. Raya, M. Flegyhazi, P. Papadimitratos, and J.-P. Hubaux. Mix-Zones for Location Privacy in Vehicular Networks. In *WiN-ITS*, 2007.
16. J. Freudiger, R. Shokri and J.-P. Hubaux. On the Optimal Placement of Mix Zones. In *PETS*, 2009.
17. B. Gedik and L. Liu. Location Privacy in Mobile Systems: A Personalized Anonymization Model. In *ICDCS*, 2005.
18. G. Ghinita, P. Kalnis, and S. Skiadopoulos. PRIVE: Anonymous Location-Based Queries in Distributed Mobile Systems. In *WWW*, 2007.
19. G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, K. Tan Private Queries in Location Based Services: Anonymizers are not Necessary. In *SIGMOD*, 2008.
20. M. Gruteser and D. Grunwald. Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. In *MobiSys*, 2003.
21. U. Hengartner and P. Steenkiste. Protecting access to people location information. In *Security in Pervasive Computing*, 2003.
22. J. Hong and J. Landay. An Architecture for Privacy-Sensitive Ubiquitous Computing. In *Mobisys*, pages 177–189, 2004.
23. P. Karger and Y. Frankel. Security and privacy threats to its. In *World Congress on Intelligent Transport Systems*, 1995.
24. A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam. l-Diversity: Privacy Beyond k-Anonymity. In *ICDE*, 2006.
25. J. Meyerowitz and R. Choudhury. Hiding Stars with Fireworks: Location Privacy through Camouflage In *MOBICOM 2009*
26. M. Mokbel, C. Chow, and W. Aref. The New Casper: Query Processing for Location Services without Compromising Privacy. In *VLDB*, 2006.
27. K. Mouratidis and M. Yiu Anonymous Query Processing in Road Networks In *TKDE*, 2010.
28. P. Pesti, B. Bamba, M. Doo, L. Liu, B. Palanisamy, M. Weber. GTMobiSIM: A Mobile Trace Generator for Road Networks. College of Computing, Georgia Institute of Technology, 2009, *http://code.google.com/p/gt-mobisim/*.
29. B. Palanisamy and L. Liu. MobiMix: Protecting Location Privacy with Mix-zones over Road Networks. In *ICDE*, 2011.
30. X. Pan, X. Meng and J. Xu Distortion based Anonymity for Continuous Queries in Location Based Mobile Services. In *GIS*, 2009.
31. V. Shmatikov and M. Wang. Timing analysis in low-latency mix networks: attacks and defenses. In *ESORICS*, 2006.
32. T. Wang and L. Liu. Privacy-Aware Mobile Services over Road Networks In *VLDB 2009*
33. T. Wang and L. Liu. Execution Assurance for Massive Computing Tasks In *IEICE Transactions on Information and Systems*, Vol. E93-D, No. 6 (June 2010), Special session on Info-Plosion.
34. P. Williams, R. Sion Usable PIR. In *NDSS*, 2008.