Location Privacy with Road network Mix-zones

Balaji Palanisamy, Ling Liu, Kisung Lee, Aameek Singh† and Yuzhe Tang

College of Computing, Georgia Tech †*IBM Research - Almaden*

***Abstract*— Mix-zones are recognized as an alternative and complementary approach to spatial cloaking based approach to location privacy protection. Mix-zones break the continuity of location exposure by ensuring that users' movements cannot be traced while they reside in a mix-zone. In this paper we provide an overview of various known attacks that make mix-zones on road networks vulnerable and illustrate a set of counter measures to make road network mix-zones attack resilient. Concretely, we categorize the vulnerabilities of road network mix-zones into two classes: one due to the road network characteristics and user mobility, and the other due to the temporal, spatial and semantic correlations of location queries. For instance, the timing information of users' entry and exit into a mix-zone provides information to launch a timing attack. The non-uniformity in the transitions taken at the road intersection may lead to transition attack. An example query correlation attack is the basic continual query (CQ) attacks, which attempt to break the anonymity of road network aware mix-zones by performing query correlation based inference. The CQ-timing attacks carry out inference attacks based on both query correlation and timing correlation, and the CQ-transition attacks execute inference attacks based on both query correlation and transition correlation. We study the factors that impact on the effectiveness of each of these attacks and evaluate the efficiency of the counter measures, such as non-rectangle mix-zones and delay tolerant mix-zones, through extensive experiments on traces produced by GTMobiSim at different scales of geographic maps.**

I. Introduction

Advances in sensing and positioning technology, fueled by wide deployment of wireless local area networks (WLAN), have made many devices location-aware. Location-based services (applications that require geographic location information as input) are becoming increasingly common. The collection and transfer of location information about a particular subject can have important privacy implications. Concrete examples of location-based services (LBSs) include searching nearest points of interest (*"Where is the nearest gas station to my current location?*"), spatial alerts (*"Remind me when I drive close to the grocery store*") , location-based social networking (*"Is my colleague Tom currently at his office?*"). Such services require the Location-based Service Provider to track the location information of their mobile users in order to deliver location based services. Continuous location based services represent queries that are *continuously evaluated* along the trajectory of a mobile user either periodically or aperiodically. Examples of continuous queries (CQs) are "informing me the nearest gas stations coming up along the highway I-85 south every 3 minutes in the next 30 minutes" or "show me the restaurants on highway I85 north within 5 miles every two minutes during the next hour". Although LBSs offer users many interesting and life enhancing experiences, they also open doors for new security risks that can endanger the location privacy of mobile clients [13], [2].

Location privacy is a particular type of information privacy. According to [5], location privacy is defined as the ability to prevent other unauthorized parties from learning one's current or past location. In LBSs, there are conceivably two types of location privacy – personal subscriber level privacy and corporate enterprise-level privacy. Personal subscriber-level privacy must supply rights and options to individuals to control when, why, and how their location is used by an application. With personal subscriber-level privacy, each individual has liberties to "opt in" and "opt out" of services that take advantage of their mobile location. Corporate enterprise-level privacy is fundamentally different in that corporate IT managers typically control when, why, and how mobile location capabilities provide application benefits to the organization as a whole.

Location privacy threats refer to the risks that an adversary can obtain unauthorized access to raw location data, derived or computed location information by locating a transmitting device, hijacking the location transmission channel, and identifying the subject (person) using a mobile device. In the United States, privacy risks related to location information have been identified in the Location Privacy Protection Act of 2001 [3]. On one hand, public disclosure of location information enables many useful services such as improved emergency assistance. Mobile users can obtain a wide variety of location-based information services, and businesses can extend their competitive edges to mobile commerce and ubiquitous service provisions. On the other hand, without safeguards, extensive deployment of location based services may risk location privacy of mobile users and to expose LBSs to significant vulnerabilities for abuse. For example, location information can be used to spam users with unwanted advertisements or draw unwanted inferences from victims' visits to clinics, doctors' offices, entertainment districts, church activities or political events. In extreme cases, unauthorized disclosure of private location information can lead to physical harm, for example in stalking or domestic abuse scenarios.

A fair amount of research efforts have been dedicated to protecting location privacy of mobile travelers. While 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 [4], [10], [11], [12], [18], mix-zone based techniques anonymize user identity by restricting when and where the exposure of users' positions are allowed [5]. **Mix-zones** are regions in space where no applications can trace user movements. The anonymity in mix-zones is guaranteed by enforcing that a set of users enter, change pseudonyms and exit a mix-zone

in a way such that the mapping between their old and new pseudonyms is not revealed [5]. The idea of building mix-zones at road intersections was first proposed in [8] and [6]. An optimal placement of mix-zones on a road map was formulated in [9]. These earlier techniques for road network mix-zones follow a straight forward refinement of basic mix-zones [5] by using rectangular or circular shaped zones. Both the definition and the construction methodologies of these mix-zones fail to take into account the effect of timing and transition attacks. The MobiMix road network mix-zones [14] are the first to define and promote attack resilient road network mix-zones. The mix-zone construction process of MobiMix tries to minimize the effect of attacks based on the characteristics of the underlying road network and at the same time provide an expected level of anonymity guarantee by taking into consideration of the statistics of user arrival rates and other factors of the road network and mobility patterns. Finally, the delay-tolerant mix-zones proposed in [15] is the first effort that identifies the vulnerabilities of existing road network mix-zones in the presence of continuous query correlation attacks (CQ attacks for short) and introduces techniques to effectively anonymize continuous location-based queries with delay-tolerant road network mix-zones.

Several factors impact on the effectiveness of mix-zone approaches, such as user population, mix-zone geometry, location sensing rate and spatial resolution, spatial and temporal constraints on user movement patterns as well as semantic continuity of the information requested by the LBSs. Mix-zones constructed on the road networks are vulnerable to timing and transition attacks due to the inherent nature of the road network and the mobility patterns of the users. Concretely, the timing information of users' entry and exit into the mix-zone provides information to launch a timing attack and the non-uniformity in the transitions taken at the road intersection provides valuable information for a transition attack. Both of these attacks aid the attacker in guessing the mapping between the old and new pseudonyms.

Mix-zones are also prone to CQ-attacks when the mobile clients obtain continuous query services. 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. We note that neither spatial cloaking nor mix-zone techniques are inherently resilient to CQ attacks. In this paper, we introduce the various attacks that road networks are vulnerable to and illustrate the possible counter measures to deal with them. We first describe and analyze the timing and transition attacks on road network mix-zones and then study the continuous query correlation attacks (CQ-attacks) that perform query correlation based inference to break the anonymity of road network-aware mix-zones. We describe three types of the continuous query correlation attacks (CQ attacks for short): (i) the basic CQ attacks in which only query correlation based inference is performed, (ii) the CQ-timing attacks in which inference attack is performed based on both query correlation and timing correlation, and (iii) CQ-transition attacks in which inference attack is performed based on both query correlation and transition correlation. We show how the anonymity of mix-zone based techniques breaks under these attack models and review the existing counter measures to deal with them. We study the effectiveness of the mix-zones attacks through extensive experiments conducted using traces produced by GTMobiSim [17] on different scales of geographic maps.

The rest of the paper is organized as follows: We first introduce the concept and definition of mix-zones in Section II and provide an overview of various attacks to mix-zones in Section III. Then we introduce attack resilient mix-zone design and construction techniques and analyze the privacy strength of these attack resilient mix-zones against the set of attack models described in Section III, including non-rectangular road-network mix-zones and delay-tolerant mix-zones. We highlight the effectiveness of our attack resilient mix-zones through experimental evaluation in Section V and conclude in Section VI.

II. MIX-ZONES

In this section we introduce the concept and notations for basic mix-zones and road network mix-zones.

A. 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:

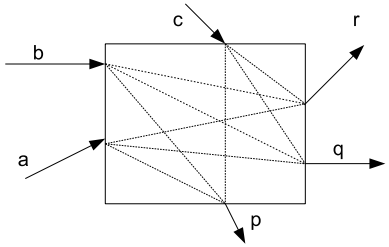
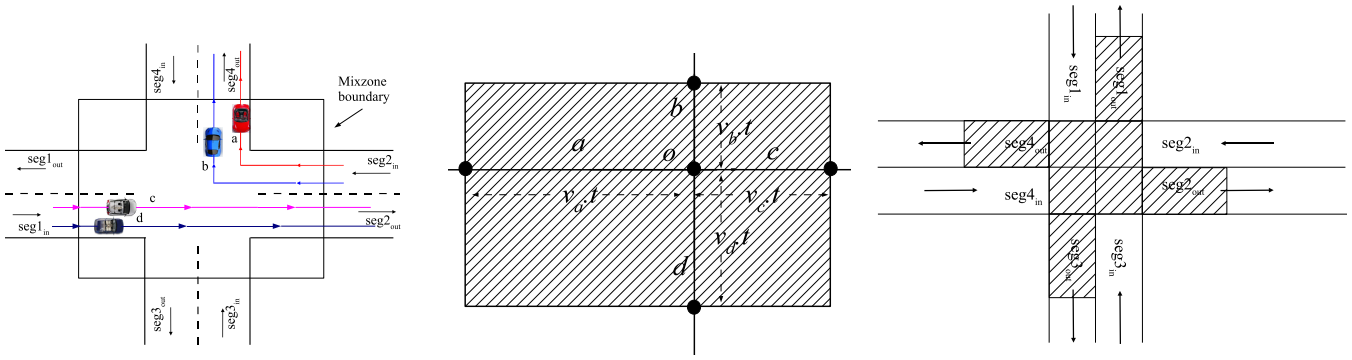


Fig. 1: Mix-zone Model

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) TWB Rectangular mix-zone (b) TWB Shifted Rectangular Mix-zones (c) TWB Non-rctangular mix-zone

Fig. 2: Road Network Mix-zone Approaches

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 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.

B. Road Network Mix-zones

Unlike the theoretical mix-zones, mix-zones constructed at road intersections (Figure 2(a)) 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 [8], [14]. Such violations provide additional information to the adversary in inferring the mapping between the old and new pseudonyms. Mix-zones constructed at road intersections have a limited number of ingress and egress points corresponding to the incoming and outgoing road segments of the intersection. Furthermore, users in a road network mix-zone are also constrained by the limited trajectory paths and speed of travel that are limited by the underlying road segments and the travel speed designated by their road class category [1]. Thus, users are not able to stay random time inside a road network mix-zone and no longer follow uniform transition probability when entering and exiting the mix-zone.

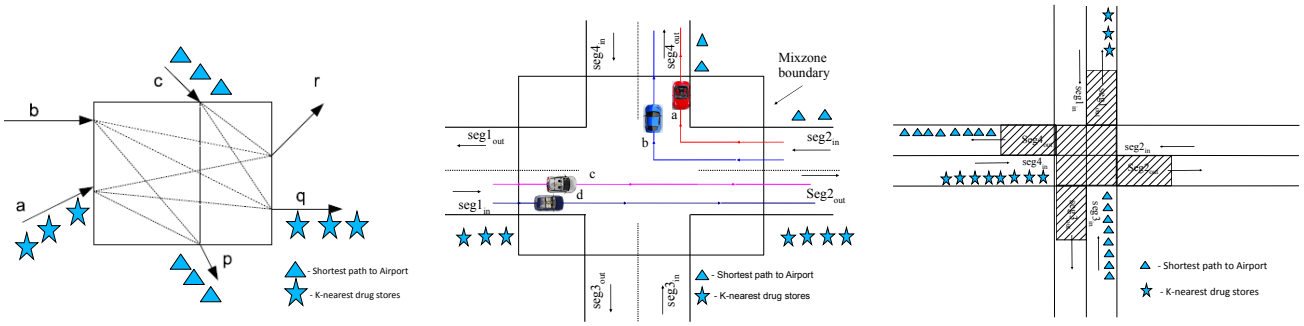
For example, in Figure 2(a), users a and b enter the road intersection from segment 2 and turn on to segment 4. Users c and d enter from segment 1 and leave on segment 2. When user a and b exit the mix-zone on segment 4 with their new pseudonyms, say α and β, the attacker tries to map their new pseudonyms α and β to some of the old pseudonyms a,b, c, and d of the same users. The new pseudonym α is more likely to be mapped to two of the old pseudonyms, a or b, than the other pseudonyms because users a and b entered the mix-zone well ahead of users c and d and it is thus less probable for c and d to leave the mix-zone before users a and b given the speed and trajectory of travel. Here, the limited randomness on the time spent inside a road network mix-zone introduces more challenges to construct efficient mix-zones. Similarly, in figure 2(a), in order for the attacker to map α and β to c and d, the old pseudonyms, users c and d should have taken a left turn from segment 1 to segment 4 and users a and b should have taken a $U-turn$ on segment 2. Based on common knowledge of inference, the attacker knows that the transition probability of a $U-turn$ is small and the mapping of α and β to c and d is very less probable. Thus, an efficient road network mix-zone should be resilient to such transition attacks. Next, we introduce the attack models related to the limitations of the road network mix-zones.

III. Threat Models and Attacks

This section is dedicated to illustrate the vulnerabilities of basic mix-zones, such as road network based timing and transition attacks and continuous query correlation attacks (*CQ-attacks*) and present a formal analysis of the mix-zone anonymization problem.

A. Attacks based on Road Network Characteristics

We describe three attack models based on the characteristics of road networks: (1) Timing Attack, (2) Transition Attack and (3) Combined timing and transition attack. The effect of attacks based on real world constraints had been studied ever since Beresford et.al. [5] proposed the mix-zone model. Freudiger et. al. [8] studied the effectiveness of the mix-zones on road networks and identified that the timing information and transition probability at the road intersection provide valuable information to the attacker for mapping the new pseudonym to the old pseudonym. Similarly, Buttyan et. al. [6] showed that the privacy obtained in the road network mix-zones is impacted by attacks related to the timing of the entry



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

Fig. 3: Mix-zone anonymization and its risks under CQ-attack

and exit events in a road network. The MobiMix road network mix-zone framework [14] developed a formal model of these attacks in road network mix-zones which are described below.

1) Timing Attack: In timing attack, 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 $j \in A$. The mapping probabilities are computed through inference based on the likelihoods of the rest of the users to exit 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 an example anonymity set, $A = \{a, b, c\}$, let user a exit with a new pseudonym a' at $t_{out}(a')$ and let the likelihoods of a, b and c exiting at time $t_{out}(a')$ be 0.1, 0.09 and 0.05 respectively. In this case, we show that it is easy to compute the mapping probabilities based on these likelihoods: $p_{a' \to a} = \frac{0.1}{0.1+0.09+0.05} = 0.416$, $p_{a' \to b} = \frac{0.09}{0.1+0.09+0.05} = 0.375$ and $p_{a' \to c} = \frac{0.05}{0.1+0.09+0.05} = 0.208$. Thus, with the timing information, the attacker is able to find that $a' \to a$ is the most probable mapping and $a' \to c$ is least probable.

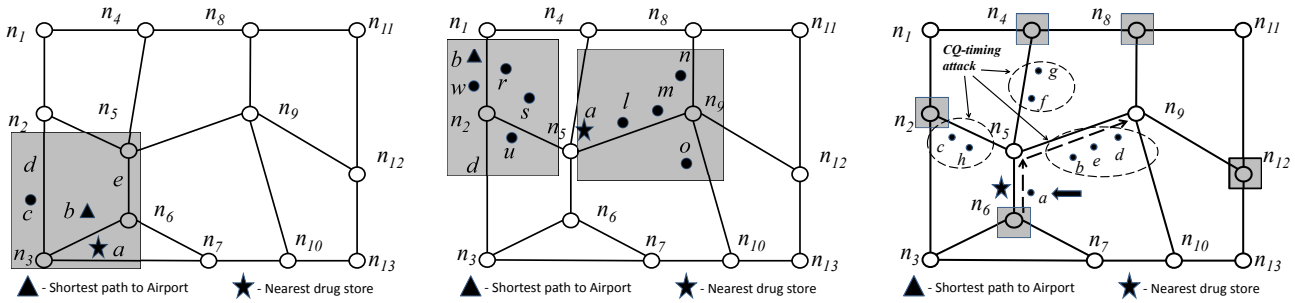
2) Transition Attack: In transition attack, the attacker estimates the transition probability for each possible turn in the intersection based on previous observations. On seeing an exiting user, i', the attacker assigns the mapping probability $p_{i' \to j}$ for each $j \in A$ based on the conditional transitional probabilities $T((ingress(j), egress(i'))$. Transition attack can equally affect the effectiveness of road network mix-zones as timing attack if not handled with care.

3) Combined Timing and Transition Attack: In the combined timing and transition attack model, the attacker is aware of both the entry and exit timing of the users and as well the transition probabilities at the road intersection for a given road network mix-zone. One can estimate the mapping probabilities $p_{i' \to j}$ for each $j \in A$ based on both the likelihoods of every user j exiting at time $t_{out}(i')$ and the conditional transition probabilities $T(ingress(j), egress(i'))$. This combined attack is often more powerful than the timing and transition attacks in isolation as it utilizes the information leaked by both timing and transition attacks.

B. Continuous Query Correlation Attacks

Road network mix-zones are also prone to CQ attacks when mobile users obtain continuous query services. 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. To the best of our knowledge, all road network mix-zones are prone to CQ-attacks.

1) Basic CQ-attack: We first illustrate the basic CQ-attack which uses only query correlation between the consecutive snapshots of the continuous query to infer the mapping between the old and new pseudonyms. Consider the example in Figure 3(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 3(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 free from CQ-attacks. For instance, we find in Figure 3(c) that even the non-rectangular mix-zone [14] that is most effective against road network timing attack is also prone to the CQ-attack. Next we briefly discuss the CQ-attacks on Spatial cloaking based solutions and describe how Spatial and temporal cloaking based techniques



(a) Spatial cloaking at time t_1 (b) Spatial cloaking at time t_2 (c) CQ Cloaking on Mix-zones

Fig. 4: Continuous Query: Spatial cloaking based techniques

over Mix-zone networks lead to CQ-timing and CQ-transition attacks.

2) Spatial Cloaking: In the spatial cloaking technique, the granularity of location exposure is reduced by exposing a larger spatial region containing the locations of k mobile users instead of the user's actual location [10], [11], [4]. In other words, the exposed location is indistinguishable from the locations of k or more users. However, for continuous queries, the spatial cloaking technique is vulnerable to query correlation based on the information across different snapshot instances of the continuous query. For example in Figure 4(a), we find users a and b asking continuous queries for nearest drug store and shortest path to airport at some time instance t. The conventional spatial cloaking algorithm finds a cloaking box encompassing the locations of a and b with the locations of three other users, c, d and e so that the users obtain *location k-anonymity* corresponding to a k value of 5. As the spatial cloaking algorithm lacks knowledge of the continuous query correlation attack, at a later point in time, say time $t+\delta t$, when users a and b have moved far from each other, it may cloak users a and b using different cloaking boxes that only ensures location k anonymity (Figure 4(b)). Therefore, we find user a's location is cloaked with the locations of l, m, n and o and similarly, the location of user b is cloaked using the locations of users r, s, u and w. Using query correlation, the attacker will be able to compare consecutive cloaking boxes and infer that user a is issuing the continuous query to nearest drug store as a is the only common user between the two cloaking boxes. Similarly the adversary can infer that b is the user executing the continuous query on the shortest path driving directions to the airport. Next we discuss a similar cloaking based approach for anonymizing continuous queries in a mix-zone framework and show how it leads to CQ-timing and CQ-transition attacks.

3) Mix-zone based CQ Anonymization: In the CQ-cloaking approach in a mix-zone framework, the continuous queries are either temporally or spatially perturbed while the snapshot queries continue to be unperturbed. 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 it makes the CQ anonymization susceptible to CQ-timing attack and CQ-transition attack.

Consider the example shown in Figure 4(c) where we have a CQ labeled as star. The square nodes represent road network mix-zones. We observe the star CQ trace starting from the mix-zone at road junction n_6. Each star 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 4(c), if the continuous query on the shortest path to the airport (marked by stars) originating from the mix-zone n_6 is perturbed spatially or 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 spatial cloaking region (or 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. Both CQ-spatial cloaking (CQ-s) and CQ-temporal cloaking (CQ-t) are similar in principle, however in CQ-spatial cloaking, instead of delaying the snapshots, the CQ exposes a larger spatial region such that there are k or more users within the spatial region.

CQ-timing Attack: 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 4(c), 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.

CQ-transition Attack: CQ-cloaking techniques are prone to another vulnerability namely 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 from consideration. For example, in Figure 4(c), at road intersection n_5, users c and h take a left turn on to the road segment, $\overline{n_2n_5}$ whereas users f and g move straight on segment, $\overline{n_4n_5}$ and users b, e and d turn right on to segment, $\overline{n_5n_9}$. 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, $\overline{n_6n_5}$ to $\overline{n_9n_5}$, the adversary will be able to eliminate the users, c, h, f and g from consideration as their transitions differ from that of the continuous query.

IV. Attack-resilient Mix-zone techniques

In this section, we discuss the state of the art techniques for mix-zone construction that are resilient to the above mentioned attacks. Before presenting the details of the construction techniques, we define the road network mix-zone as proposed by the MobiMix road network model [14]:

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: $H_{pair}(i, j, t) \geq \alpha$.
3) For any two users $i, j \in A$, the pairwise entropy after transition attack should meet the condition: $H_{pair}(i, j) \geq \beta$.

The pairwise entropy, $H_{pair}(i, j)$ between users i and j is the entropy obtained by considering i and j to be the only members of the anonymity set. 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. However, mix-zones on road networks may not achieve ideal pairwise entropy of 1.0 due to road network constraints and motion constraints for traveling on road networks. Thus we consider an effective road network mix-zone as the one that provides a pairwise entropy close to 1.0 for all pairs of users in the anonymity set.

A. Resilience to Timing attack

Many existing mix-zone proposals adopt a straight forward approach to construct mix-zones around the road junction using a rectangular or circular region centered at the road junction as shown in figure 2(a). However, such a straight forward approach has detrimental effects on the obtained privacy. In a naive rectangular mix-zone, for each exiting user i', the set of users that were inside the mix-zone at any given time during user i''s presence in the mix-zone forms its anonymity set, A_i. Therefore, although the anonymity set size is typically large in these mix-zones, a large number of members of the anonymity set become low probable under the timing attack if the arrival times of the users differ by a large value. To address this problem, the notion of mix-zone time window has been adopted in existing mix-zone proposals including [8], [6], [9] where a default value of time window is assumed for the junctions and accordingly, the anonymity set for each user, i is assumed to comprise of users who had entered within a time window in the interval, $|t_{in}(i) - \tau_1|$ to $|t_{in}(i) + \tau_2|$. Here, $t_{in}(i)$ is the arrival time of user i and τ_1 and τ_2 are chosen to be small values so that the time window ensures that the anonymity set of i comprises only of the users entering the mix-zone with a closely similar arrival time as that of i. The Time window bounded (TWB) rectangular approach in MobiMix [14] adopts a similar time window, however, the size of the time window is decided based on the arrival rate of users so that k or more users enter within the time window. The Time window bounded (TWB) shifted rectangular mix-zone (Figure 2(b)) is similar to the TWB rectangular mix-zone, however, the rectangle in this case is not centered at the centre of the junction, instead it is shifted in such a way that from any point of entry into the mix-zone, it takes the same amount of time to reach the centre of the road junction when travelled at the mean speed.

The Mobimix construction algorithms[14] proposed an effective way of constructing mix-zone by having the mix-zone region start from the centre of the junction only on the outgoing road segments as shown in Figure 2(c). We refer to this technique as the non-rectangular approach. The TWB non-rectangular approach is free from timing attacks caused by the heterogeneity in the speed distributions on the road segments. We note that in the time window bounded non-rectangular approach, the anonymity set for each user, i is assumed to comprise only of users who had entered within a time window in the interval, $|t_{in}(i) - \tau_1|$ to $|t_{in}(i) + \tau_2|$. Here, $t_{in}(i)$ is the arrival time of user i and τ_1 and τ_2 are chosen to be small values so that the time window ensures that the anonymity set of i comprises only of the users entering the mix-zone with a closely similar arrival time as that of i. Hence, when i exits out as i', the attacker would be unable to differentiate i' from all members of i's anonymity set, A_i as they are all likely to exit at the same time when i exits. The length of the mix-zone along each outgoing segment is chosen based on the mean speed of the road segment, the size of the chosen time window and the minimum pairwise entropy required. Thus the TWB non-rectangular mix-zones make guarantees on the lower bound pairwise Entropy between the users.

B. Resilience to Transition attack

As discussed earlier, in a road network, it is possible to launch transition attack to guess the linking between the pseudonyms. For each exiting user, i' the attacker observes the exiting segment of i' and tries to maps i' to one of the users, j in the anonymity set based on the conditional transitional probability of exiting in the outgoing segment, $oseg(i)$ given that j entered from the incoming segment, $iseg(j)$. In order to protect against transition attack in cases where the transition probability is skewed, the transition attack resilient technique in [16] proposed that the mix-zone time window should be chosen in such a way that for each outgoing segment, l, there are enough number of users (k or more) entering the mix-

zone from the road segments that have similar transitioning probability to the outgoing segment, l, and hence have a higher pairwise entropy, say greater than or equal to β. Therefore, the attacker will have at least k users in the anonymity set that he cannot ignore from consideration.

C. Resilience to CQ-attacks

CQ-attack by far is the most challenging attack in a road network mix-zone. To the best of our knowledge, no road network mix-zone is completely free from CQ-attacks. However, 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 as the attacker breaks the anonymity obtained in the intermediate mix-zones 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. The Delay-tolerant mix-zones proposed in [15] 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 shift (temporal delay-tolerant mix-zones) or a random spatial shift (spatial delay-tolerant mix-zones) 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. Therefore, the anonymity strength of delay-tolerant mix-zones comes from a unique combination of both identity mixing and location mixing. Such high anonymity provides the initial anonymity required to anonymize the continuous queries so that the queries obtain the required anonymity even under the CQ-attack model.

V. Experimental Evaluation

In this section, we present our experiments on the effectiveness of the various attack resilient mix-zone techniques on road networks and discuss the level of privacy provided by them. We first describe the experimental setup and the road-network mobile object simulator used in the experiments.

A. Experimental setup

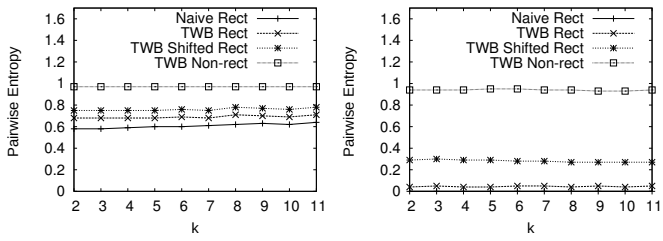
We use the GT Mobile simulator [17] to generate a trace of cars moving on a real-world road network, obtained from maps available at the National Mapping Division of the USGS [1]. The simulator extracts the road network based on three types of roads – *expressway*, *arterial* and *collector* roads. Our 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 I.

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

TABLE I: Motion Parameters

B. Effectiveness of timing attack

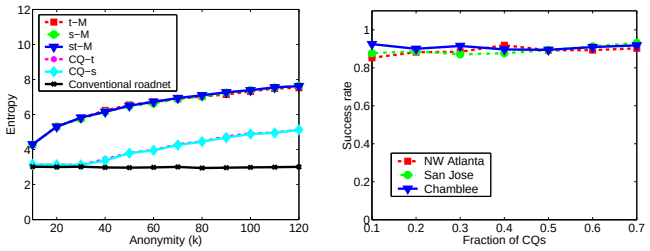
We first measure the resilience of the various techniques to timing attacks based on road network characteristics. Figure 5 shows the average and worst-case pairwise entropy of the mix-zones for various values of k, the size of the anonymity set. In Figure 5(a), we observe that the effect of timing attack is different across various approaches: we find that the TWB non-rectangular mix-zones perform the best under timing attack with the average pairwise entropy close to 1.0. Here, the length of the non-rectangular mix-zone is computed so as to ensure a lower bound pairwise entropy of $\alpha = 0.9$ for the chosen time window size, τ which is computed based on the user arrival rate in the road junction to ensure the expected value of k with a high probability of p = 0.9. In order to compare the effectiveness of the other mix-zone approaches with the TWB non-rectangular approach, the TWB rectangular and TWB shifted rectangular mix-zones are also constructed with the same length and time window as used by the non-rectangular mix-zone. Similarly, the size of the naive rectangular mix-zone is fixed in such a way that the mean time to cross the mix-zone equals the time window of the TWB non-rectangular mix-zone.



(a) Average Pairwise Entropy (b) Worst-case Pairwise Entropy

Fig. 5: Resilience to timing attack

In Figure 5(a), we also find that the naive rectangular and time window bounded rectangular mix-zones have low pairwise entropies after timing attack but the pairwise entropy of the TWB shifted rectangular approach is relatively higher,



(a) Entropy after CQ-attack (b) Success rate

Fig. 6: Effectiveness of Delay-tolerant mix-zones

close to 0.8 as its geometry is more resilient to timing attack. However, a high pairwise entropy of 0.9 or higher may be often required to ensure strong anonymity. In such cases, the time window bounded non-rectangular approach becomes the most efficient choice. The worst case pairwise entropy in figure 5(b) represents the lowest possible pairwise entropy obtained by the users after timing attack. Here also, only the TWB non-rectangular approach offers a high value for the worst case pairwise entropy.

C. Resilience to CQ-attacks

This set of experiments compares the delay-tolerant mix-zone approaches with the conventional mix-zones and CQ-cloaking techniques in terms of their resilience to CQ-attacks. 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 6(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) described in Section III-B.3 for various values of required anonymity, k. Here, the temporal window and spatial region of the delay-tolerant mix-zones 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, $p = 0.9$. 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 cannot 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 temporal cloaking and CQ spatial cloaking approaches (CQ-t and CQ-s) have low level of Entropy due to the effect of CQ-timing and CQ-transition attacks. Thus, the delay-tolerant mix-zones are more effective to meet the privacy requirements of the continuous queries under the CQ-attack model. The success rate of the spatio-temporal delay-tolerant mix-zone techniques is compared across different scales of geographic maps in Figure 6(b). We find that the delay-tolerant mix-zones provide a high success rate and performs consistently across different geographic maps.

VI. Conclusion

This paper promotes the use of mix-zones as an effective alternative approach to location privacy protection, complementary to spatial cloaking [4], [10], [11], [12], [18]. We discussed the vulnerabilities and challenges of constructing attack-resilient mix-zones on road networks. For example, on a road network, the timing information of users' entry and exit into the mix-zone may lead to timing attacks and the non-uniformity in the mobility patterns taken at the road junctions may lead to transition attacks. When the location based services are continuous in nature, mix-zones will face the challenge of CQ-attacks, which perform query correlation based inference on continuous queries to break the anonymity of the road network mix-zones. We described our approach to developing attack resilient mix-zones on road networks and presented some highlights on the effectiveness of our approach through extensive experiments using traces generated by GTMobiSim [17].

Acknowledgement: This work is partially sponsored by grants from NSF CISE SaTC program, NetSE program, I/UCRC, an IBM faculty award, an IBM PhD fellowship and a grant from Intel ICST on Cloud Computing.

References

[1] U.S. Geological Survey. http //www.usgs.gov.
[2] USAToday. Authorities: Gps systems used to stalk woman. http://www.usatoday.com/tech/news/2002-12- 30-gps-stalker x.htm.
[3] Location Privacy Protection Act of 2001. http://www.techlawjournal.com/cong107/privacy/location/s1164is.asp
[4] B. Bamba, L. Liu, P. Pesti, and T. Wang. Supporting Anonymous Location Queries in Mobile Environments with PrivacyGrid. In *WWW*, 2008.
[5] A. Beresford and F. Stajano. Location Privacy in Pervasive Computing. *Pervasive Computing, IEEE*, 2003.
[6] L. Buttyan and T. Holczer and I. Vajda. On the effectiveness of changing pseudonyms to provide location privacy in VANETs In *ESAS 2007*
[7] C. Chow and M. Mokbel Enabling Private Continuous Queries For Revealed User Locations. In *SSTD*, 2007.
[8] J. Freudiger, M. Raya, M. Flegyhazi, P. Papadimitratos, and J.-P. Hubaux. Mix-Zones for Location Privacy in Vehicular Networks. In *WiN-ITS*, 2007.
[9] J. Freudiger, R. Shokri and J.-P. Hubaux. On the Optimal Placement of Mix Zones. In *PETS*, 2009.
[10] B. Gedik and L. Liu. Location Privacy in Mobile Systems: A Personalized Anonymization Model. In *ICDCS*, 2005.
[11] G. Ghinita, P. Kalnis, and S. Skiadopoulos. PRIVE: Anonymous Location-Based Queries in Distributed Mobile Systems. In *WWW*, 2007.
[12] M. Gruteser and D. Grunwald. Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. In *MobiSys*, 2003.
[13] P. Karger and Y. Frankel. Security and privacy threats to its. In *World Congress on Intelligent Transport Systems*, 1995.
[14] B. Palanisamy and L. Liu. MobiMix: Protecting Location Privacy with Mix-zones over Road Networks. In *ICDE*, 2011.
[15] B. Palanisamy, L. Liu, K. Lee, S. Meng, Y. Tang and Y. Zhou Anonymizing Continuous Queries with Delay-tolerant Mix-zones over Road Networks *Georgia Tech Technical Report*
[16] B. Palanisamy and L. Liu Attack-resilient Mix-zones over Road Networks: Architecture and Algorithms *Georgia Tech Technical Report*
[17] 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/*.
[18] T. Wang and L. Liu. Privacy-Aware Mobile Services over Road Networks In *VLDB 2009*