Road Network Mix-zones for Anonymous Location Based Services

Balaji Palanisamy, Sindhuja Ravichandran, Ling Liu, Binh Han, Kisung Lee and Calton Pu

College of Computing, Georgia Tech

{balaji, sindhuravi, lingliu, binh.han, kisung.lee, calton.pu}@cc.gatech.edu

***Abstract*—We present MobiMix, a road network based mix-zone framework to protect location privacy of mobile users traveling on road networks. An alternative and complementary approach to spatial cloaking based location privacy protection is to break the continuity of location exposure by introducing techniques, such as mix-zones, where no applications can trace user movements. However, existing mix-zone proposals fail to provide effective mix-zone construction and placement algorithms that are resilient to timing and transition attacks. In MobiMix, mix-zones are constructed and placed by carefully taking into consideration of multiple factors, such as the geometry of the zones, the statistical behavior of the user population, the spatial constraints on movement patterns of the users, and the temporal and spatial resolution of the location exposure. In this demonstration, we first introduce a visualization of the location privacy risks of mobile users traveling on road networks and show how mix-zone based anonymization breaks the continuity of location exposure to protect user location privacy. We demonstrate a suite of road network mix-zone construction and placement methods that provide higher level of resilience to timing and transition attacks on road networks. We show the effectiveness of the MobiMix approach through detailed visualization using traces produced by GTMobiSim on different scales of geographic maps.**

I. Introduction

Continuous exposure of location information, even with spatially cloaked resolution, may lead to breaches of location privacy due to statistics-based inference attacks. Location privacy is a system-level capability of location systems, which controls the access to this information at different spatial granularity and different temporal and continuity scale, rather than stopping all access to location information. An alternative and complementary approach to spatial cloaking based location privacy protection is to break the continuity of location exposure by introducing techniques, such as mix-zones, where no applications can trace user movements. Several factors impact on the effectiveness of mix-zone approach, such as user population, mix-zones geometry, location sensing rate and spatial resolution, as well as spatial and temporal constraints on user movements. None of the existing mix-zone approaches consider these factors effectively. Most existing mix-zone proposals fail to provide effective mix-zone construction algorithms that are effective for mobile users traveling on road networks and yet resilient to timing and transition attacks.

We present MobiMix, a road network based Mix-Zone framework to protect location privacy of mobile users traveling on road networks. In a road network, mix-zones can be constructed at road intersections where there is high uncertainty in the trajectories followed by the users. However, compared to the theoretic mix-zones [2], the road networks impose many challenges that limit the anonymity provided by the mix-zones constructed independently of the spatially constrained road networks. For instance, the timing information of users' entry and exit into the mix-zone and the non-uniformity in the transitions taken at the road intersection provide valuable information to the attacker to guess the mapping between the old and new pseudonyms [3]. In MobiMix [5], [6], we develop a general framework and a suite of algorithms for constructing mix-zones in road networks, taking into account the constraints and limitations imposed by the road networks, the timing of users entering and exiting a mix-zone, and the transitioning probability of users in terms of their movement trajectory. In this paper, we first introduce a visualization of the location privacy risks of mobile users traveling on road networks and show how mix-zone based anonymization breaks the continuity of location exposure to protect user location privacy. We then demonstrate the first prototype of the MobiMix location anonymization system and show its efficiency and privacy strengths through detailed visualization using traces produced by GTMobiSim [4] on different scales of geographic maps.

II. Mix-zones Overview

In this section, we review the theoretical mix-zone model and its assumptions and introduce the concept of road network mix-zones and their challenges.

A. The Mix-zone Model

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. A mix-zone is analogous to a mix node in anonymous communication systems [2], where each mix node collects n equal-length packets as input and reorders them randomly before forwarding them, thus providing unlinkability between incoming and outgoing messages. 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. These k users exit the mix-zone in an order different from their order of arrival, providing unlinkability between their entering and exiting events. The strong assumptions used by the model to ensure high privacy guarantee can be formally illustrated as follows:

Definition 1: A mix-zone Z is said to be k-anonymized with 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., an user entering through an entry point, $e \in E$, is equally likely to exit in any of the exit points, $o \in O$.

Inside the mix-zone, the location of users cannot be tracked. The ideal mix-zone model has two strong assumptions namely (i) users spend random time inside and (ii) users take uniform transitions while entering and exiting mix-zones. However, such assumptions may be violated in a road network scenario.

B. Road Network Mix-zones

Theoretical mix-zones assume mobile users move in an Euclidian space without any spatial constraints. In real world, mobile users always move on a spatially constrained space, such as road networks or walk paths. Each road network mix-zone corresponds to a road intersection on a road network. The decision of which intersections are suitable for building mix-zones is usually made based on a number of factors such as the number of road segments at the intersection, the travel speed and trajectory constraints of mobile users inside the mix-zone. 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 [7]. 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 1, 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 the timing attack that challenges the mix-zone construction. Similarly, in Figure 1, 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. Based on the constraints of the road network, a road network mix-zone can be formally defined as follows:

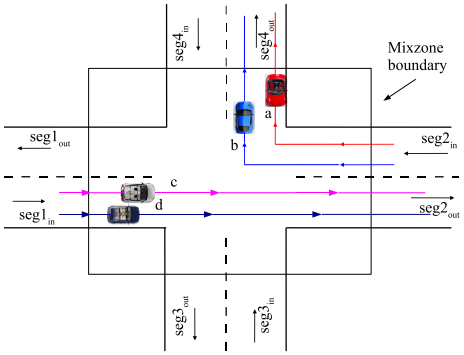


Fig. 1: Road Network Mix Zone

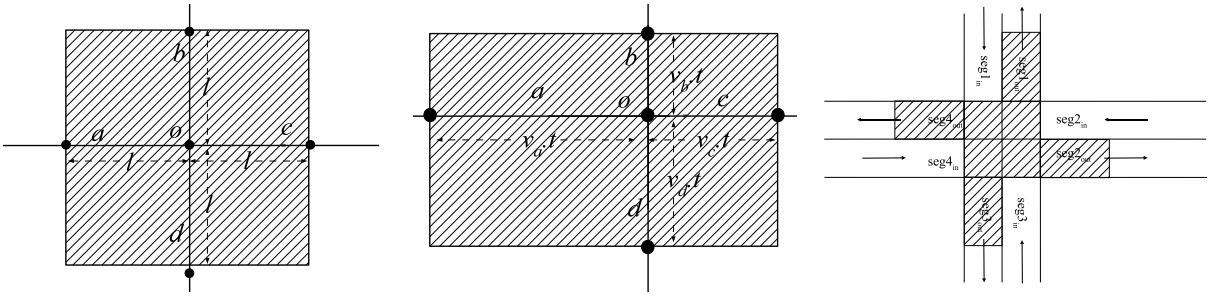
Definition 2: A road network mix-zone offers k-anonymity to a set A of users if and only if :

1) There are k or more users in the anonymity set A.
2) Given any two users $i, j \in A$, the pairwise entropy after timing attack $H_{pair}(i, j, t) \geq \alpha$.
3) For any two users $i, j \in A$, the pairwise entropy after transition attack $H_{pair}(i, j) \geq \beta$.

Here, the pairwise entropy between an user, i and j is the entropy obtained by considering i and j to be the only members of the anonymity set. In comparison, by Definition 1, a theoretical mix-zone ensures a uniform probability distribution for all possible mappings between old and new pseudonyms and a high pairwise entropy of 1.0 for all pairs of users in the anonymity set. An effective mix-zone should provide a pairwise entropy close to 1.0 for all possible pairs of the anonymity set. Next, we discuss the MobiMix techniques for attack-resilient mix-zone construction that effectively satisfy the above constraints.

III. MobiMix Approaches

We discuss the effectiveness of the MobiMix mix-zone construction approaches against timing attack and discuss how the mix-zone geometry and road characteristics impact on the attack-resilience. We first describe the weaknesses of the naive rectangular mix-zone approach and then discuss three MobiMix mix-zone construction techniques: (i) Time Window Bounded (TWB) Rectangular, (ii) Time Window Bounded (TWB) Shifted Rectangular and (iii) Time Window Bounded (TWB) Non-rectangular mix-zones. All of them perform better than the naive Rectangular mix-zones under timing attack.



(a) Rectangular Mix-zones (b) Shifted Rectangular Mix-zones (c) Non-rectangular Mix-zones

Fig. 2: Mix Zone Shapes

A. Naive Rectangular Mix-zones

A straight forward approach to construct mix-zones around the road junction is to define a rectangular region centered at the road junction as shown in Figure 2(a). The rectangle is defined based on some default size. 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. As any two users that were present together at any same given time, become members of each other's anonymity sets, the attacker can eliminate a large number of low probable members from consideration based on the timing information. For instance, if the arrival times of the two users differ by a large value, then eventhough they might be present together in the mix-zone at some time instance, it is more likely that the user who entered first is likely to exit first.

B. TWB Rectangular Mix-zones

In the time window bounded approach, the rectangle is constructed in the same way as in naive rectangular mix-zone, however, the anonymity set for each user, i is assumed to comprise of only 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. However, the right size of the time window should be decided based on a number of factors including the mix-zone size, the speed distribution of users on the road segments and the level of anonymity users expect. For road intersections that have segments with the same speed distribution, we can precisely guarantee a lower bound on the pairwise entropy for the members of the anonymity set by constructing the anonymity set with the right value of time window based on our MobiMix road network model. However, for road junctions that have segments with different speed distributions, the timing attack can be very effective.

C. TWB Shifted Rectangular Mix-zones

In the Time window bounded shifted rectangular approach, the rectangle 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 as shown in Figure 2(b). In the same way, from the centre of the junction, it takes the same time to reach any exit point when travelled at the mean speed of the road segments. Here, a set of users entering within the short time window, $|t_{in}(i) - \tau_1|$ to $|t_{in}(i) + \tau_2|$ are likely to exit the mix-zone at the same time. Hence, when user i exits as i' the attacker would find that i' is likely to be any of the members of the anonymity set, A_i. If t represents the average time to reach the centre of the road junction from an entry point which is the same as the average time to reach an exit point from the junction center, then the mix-zone lengths on the segments would be given by the product of their mean speed, say v and the average time, t as shown in Figure 2(b). Compared to naive rectangular and time window bounded rectangular mix-zones, shifted rectangular mix-zones provide good pairwise entropy for many cases, however, they do leak information when the speed of the users deviate from the mean speed.

D. TWB Non-Rectangular mix-zones

A more effective way to construct mix-zones would be to have the mix-zone region start from the centre of the junction only on the outgoing road segments as shown in Figure 2(c). The non-rectangular approach is free from timing attacks caused by the heterogeneity in the speed distributions on the road segments. As in the rectangular approaches, the anonymity set for each user, i comprises of users who had entered the mix-zone within a time window in the interval, $|t_{in}(i)-\tau_1|$ to $|t_{in}(i)+\tau_2|$. 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.

IV. Demonstration Toolkit

The demonstration of MobiMix comprises of three components. We first demonstrate the location privacy risks of mobile users traveling on road networks and illustrate how mix-zone based anonymization breaks the continuity of location exposure to protect user location privacy. We

then demonstrate the MobiMix road network mix-zone construction and deployment methods that provide higher level of resilience to timing and transition attacks on road networks. We visualize the rectangular, shifted rectangular and non-rectangular mix-zones over road networks and demonstrate the privacy obtained by individual users along their trajectories. The toolkit also incorporates visualization of the geographic maps and user mobility on the road network and demonstrates all MobiMix mix-zone techniques with close visualization.

A. *Visualizing Geographic Maps and user mobility*

The demonstration uses GT Mobile simulator [4] 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 [7]. The simulator extracts the road network based on three types of roads – *expressway*, *arterial* and *collector* roads. The interface allows to use any desired geographic map available from USGS. Based on the traffic volume information provided, it generates a set of 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 GUI also allows to specify the speed distribution of the cars based on the road class categories.

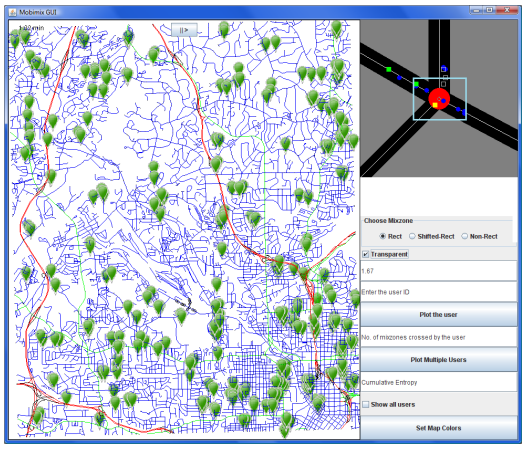


Fig. 3: MobiMix toolkit: Mix-zone deployment using rectangular mix-zones

B. *Visualizing Mix-zones*

The demonstration toolkit provides a close visualization of the functioning of mix-zones in the road intersections. The interface lets us visualize any mix-zone deployed in the road network by closely rendering the user mobility in and out of the mix-zone. It renders a closer visualization of the road network and the geometry of the mix-zones deployed on them. We demonstrate all the three types of MobiMix mix-zones namely (i) time window bounded rectangular mix-zones, (ii) time window bounded shifted rectangular mix-zones and (iii) time window bounded non-rectangular mix-zones. Figure 3 shows the mix-zone deployment on the Northwest Atlanta region of Georgia and a close visualization of a rectangular mix-zone with visible traffic. Here, only the mix-zone boundary is rendered by the visualizer and therefore all user movements inside the zone are visible. Similarly in Figure 4, a TWB non-rectangular mix-zone is closely visualized where the mix-zone area spans only the outgoing segments of the road junction.

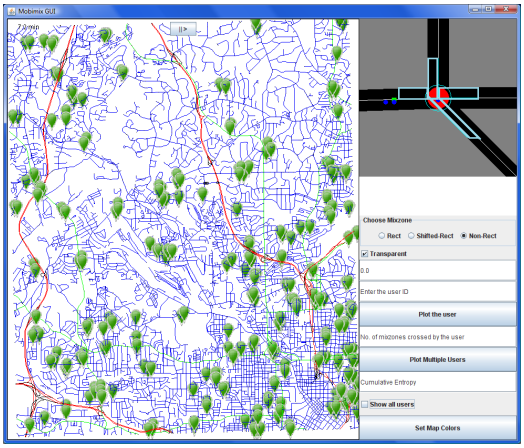


Fig. 4: MobiMix toolkit: Non-rectangular mix-zones

C. *Privacy Comparison and Performance Evaluation*

The final part of the demonstration toolkit includes the comparison of the privacy strengths and performance of various MobiMix techniques that explains user location privacy levels under different schemes. Here, the privacy strength is compared in terms of (i)anonymity set size, (ii) obtained Pairwise Entropy and (iii) overall Entropy of the users. Similarly, the performance of the mix-zones is compared in terms of the success rate which represents the fraction of cases where the mix-zone provides the expected level of anonymity and relative-k which is defined as the ratio of the anonymity obtained in the mix-zones to the anonymity expected from them.

V. Acknowlegement

This work is partially sponsored by grants from NSF CISE NetSE program, SaTC program and I/UCRCs program, an IBM faculty award and a grant from Intel ISTC on Cloud Computing.

References

[1] B. Bamba, L. Liu, P. Pesti, and T. Wang. Supporting Anonymous Location Queries in Mobile Environments with PrivacyGrid. In *WWW*, 2008.

[2] A. Beresford and F. Stajano. Location Privacy in Pervasive Computing. *Pervasive Computing, IEEE*, 2003.

[3] J. Freudiger, M. Raya, M. Flegyhazi, P. Papadimitratos, and J.-P. Hubaux. Mix-Zones for Location Privacy in Vehicular Networks. In *WiN-ITS*, 2007.

[4] 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/*.

[5] B. Palanisamy and L. Liu. MobiMix: Protecting Location Privacy with Mix-zones over Road Networks. In *ICDE*, 2011.

[6] B. Palanisamy, L. Liu, K. Lee, A. Singh and Y. Tang. Location Privacy with Road Network Mix-zones. In *MSN*, 2012.

[7] U.S. Geological Survey. http://www.usgs.gov