Secure Mix-Zones for Privacy Protection of Road Network Location Based Services Users

Rubina S. Zuberi¹² and Syed N. Ahmad³

¹Department of Electronics & Communication Engineering, I.T.S. Engineering College, Greater Noida 201308, India
²Dr. A.P.J. Abdul Kalam Technical University (AKTU), Lucknow 226021, India
³Department of Electronics & Communication Engineering, Jamia Millia Islamia, New Delhi 110025, India

Correspondence should be addressed to Rubina S. Zuberi; rshahinz@gmail.com

Received 30 November 2015; Accepted 21 March 2016

Copyright © 2016 R. S. Zuberi and S. N. Ahmad. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Privacy has been found to be the major impediment and hence the area to be worked out for the provision of Location Based Services in the wide sense. With the emergence of smart, easily portable, communicating devices, information acquisition is achieving new domains. The work presented here is an extension of the ongoing work towards achieving privacy for the present day emerging communication techniques. This work emphasizes one of the most effective real-time privacy enhancement techniques called Mix-Zones. In this paper, we have presented a model of a secure road network with Mix-Zones getting activated on the basis of spatial as well as temporal factors. The temporal factors are ascertained by the amount of traffic and its flow. The paper also discusses the importance of the number of Mix-Zones a user traverses and their mixing effectiveness. We have also shown here using our simulations which are required for the real-time treatment of the problem that the proposed transient Mix-Zones are part of a viable and robust solution towards the road network privacy protection of the communicating moving objects of the present scenario.

1. Introduction

If we closely see the paradigm of Location Based Services (LBS), their wide popularity is facing a continuous impediment from the privacy concerns of its users. Most of the cellular phones, tablets, and other mobile communicating device users want to use the location based applications available. But the breach in the privacy of “where they are?” has severely restricted the boost in the growth of these services [1]. Location privacy has been enhanced continually from different perspectives by the Service Providers. Initially, researchers dwelled into the enhancement of particular location of the moving objects (MOs) [2–5]. Amongst these techniques, k-anonymity [6, 7] attracted considerable attention from the privacy enhancing techniques providers. The threat models suggesting location correlations by the attacker found easily that it becomes futile to consider anonymization of single location at a time. Temporal along with spatial privacy enhancements techniques were then considered. Hence, set of locations rather than location at a time were considered for privacy enhancement procedures. This was recognized as the location trajectory privacy [8–11]. Trajectory privacy techniques are also used in publishing location based data and for mobile data management systems [12–14]. Trajectory k-anonymity was first proposed by Chow and Mokbel [8] who extended the concept of the snapshot (single location) queries to continuous queries and formed this new domain of trajectory k-anonymity. In order to process continuous LBS requests, there are two main approaches: (a) an LBS request is submitted repeatedly for each time instance until it expires, thus requiring the evaluation of the results continuously, and (b) the query result is computed only once if the information on the future trajectory is provided. The first approach suffers from the drawback of sampling (if the sampling rate is too low the results will be incorrect) [15]. Hence there is no guarantee about the query results. Chow and Mokbel [8] made the algorithm for continuous queries which can achieve these goals: it (a) distinguishes between location privacy and query privacy, (b) employs the k-sharing region and memorization properties, and (c) supports continuous location based queries. They brought about the concept of continuous queries but were more focused on query privacy.
when location information is available. Tao et al. [16] were the first to think about the possibility of continuous queries. They ventured into the problem of finding nearest neighbors (NN) continuously on a traversed segment or trajectory. The search for k-NN for a moving point also became the subject of the data base community. Based on the provision of future trajectories by the user, there are some approaches which anonymize the trajectories. Shin et al. [17] showed that the longer the attacker can track the user’s trajectory, the stronger the possibility that the user’s sensitive information is revealed. They proposed partitioning of trajectory and dividing the continuous requests too. These approaches hence pertain to trajectory privacy protection where the trajectory could be offline/real-time. The role of imprecision in the location information of Geographic Information Systems (GIS) which lead to an uncertainty threshold \( \delta \) useful for trajectory privacy was also provided. They worked for achieving \((k, \delta)\) anonymity for trajectory publishing domain. But recently there is a good attention of researchers on the modified Mix-Zones which are part of a robust solution to the real-time trajectory privacy protection.

Mix-Zones were introduced lately by Beresford and Stajano [18] but, due to the temporal attacks, this work could not attract much attention. Later, Mix-Zones were applied to the traffic oriented vehicular networks (VNs) by many [19–22]. While most of these talk about adequate placement of Mix-Zones, Palanisamy et al. [21] revealed how only the spatial construction techniques of the Mix-Zone can greatly affect the privacy of the users going through it. In this paper we have tried to mathematically model the traffic flows of the road network based Mix-Zone in Section 2 which will also present the evaluation of mixing effectiveness of the Mix-Zones. Section 3 describes the construction technique of the Mix-Zone used here. We have also described how spatial as well as temporal features can be incorporated in the construction of Mix-Zones. If only spatial features are considered we term the Mix-Zone as a static Mix-Zone while incorporating both spatiotemporal constraints in the construction of the Mix-Zones we call them transient Mix-Zones. The transient Mix-Zones improvise the safeguard against timing and velocity based attacks. This has been presented in Section 4. These types of attacks are found to be crucial in Mix-Zones. It has been shown through our results and analyses of real-time simulations that this problem is almost eliminated in the transient Mix-Zones proposed here. Section 4 will also describe the experimental data on which analyses have been performed. Section 5 describes the results and Section 6 gives the conclusion and the work which can be pursued further.

2. Mix-Zones

2.1. The Mix-Zone Road Network Model. The anonymization techniques used for privacy enhancement particularly for Location Based Systems (LBS) generally have two types of architectures. The first one consists of a trusted third-party (TTP) server and the other one has a peer to peer approach [23]. A real-time anonymity system using the former architecture can be developed for using the concept of Mix-Zones [18]. A trusted third-party server is the trusted server of the telecom operator that is associated with the Location Based Services server and the end user. The TTP is supposed to be absolutely secure with the data which it is going to process and will be solely responsible for doing the anonymization procedure [24]. This system will work only when the moving object’s (MO) paths intersect. On the road networks, these intersections are generally the traffic signal junctions [25].

Hence, the traffic signal junctions can be identified as Mix-Zones. A Mix-Zone is a special area (depicted as gray colour area in Figure 1) defined by the trusted third-party server which is responsible for linking the MOs with the Location Based Service Providers. This trusted server is responsible for providing the identities of the corresponding MOs to all the connected LBS application servers. These identities should be the pseudonyms which will provide the first privacy cover. The Mix-Zone will then enhance the real-time privacy of the MO by anonymizing all the users inside it. This anonymization will be effective for a particular time interval \( \Delta t \) which is the time period of activation of the Mix-Zone by the trusted third-party (TTP) server. In addition, it will repseudonymize the exiting MOs to mix them and create confusion for the user’s privacy attacker. The location and duration of the Mix-Zone are communicated to the MOs prior to their joining network by the TTP server.

If \( M \) is a Mix-Zone (Figure 1) defined geometrically by the six segments, Seg1_{in}, Seg1_{out}, Seg2_{in}, Seg2_{out}, Seg3_{in} and Seg3_{out}, then all the ingress MOs will move randomly in the said segments for a time period \( \Delta t \). The egress MOs will be renamed with different pseudonyms. The timing information of the MO prior to entering into the Mix-Zone and after leaving the Mix-Zone creates the probability of an attack. The uncertainty in this attack on the \( m \)-th Mix-Zone can be given by the entropy [26]

\[
H_N(m) = -\sum_{i=1}^{L} p_i \log_2 (p_i),
\]

where \( N \) is the number of times the \( m \)-th Mix-Zone will be activated such that \( m \in Z \) (\( Z \) being the area under
consideration). $p_s$ is the probability of different assignments from ingress MO to egress MO and $L$ is the number of such hypothesized assignments formulated by the attacker. The average of all such entropies in the said area will give an insight into the location of MO to the attacker. This average is the same as the expected value which directly relates to the mixing effectiveness of the Mix-Zone $m$:

$$E[H(m)] = \frac{1}{n} \sum_{i=1}^{n} H_{\text{in}}(m).$$

(2)

Each Mix-Zone is traversed through by well defined boundaries indicating possible ingress/egress points of the MOs. The duration of activation of Mix-Zone must depend on the flow of MOs inside the Mix-Zone. Each Mix-Zone is traversed by flows $f_j \in F_m \subseteq F$ for $j$ number of flows in $m$th Mix-Zone. $F_m$ is the set of flows in the Mix-Zone $m$ and $F$ is a set of all flows inside the considered area $Z$.

2.2. Mixing Effectiveness of the Mix-Zone. To find out the strength of the Mix-Zone in creating ambiguity in the location information of MO, we will try to find out the mixing effectiveness of the Mix-Zone in a manner which is unrelated to a set of events or the way MOs are actually moving in the Mix-Zones [27]. If we model the flow of the MOs for finding the mixing effectiveness of the Mix-Zones then, taking $f_j$ as homogenous Poisson process with intensity $\mu_j$, the distribution $\text{Poiss}(s)$ denotes the probability that $s$ MOs ingress the flow $f_j$ during a time period $T$ of $n$ steps ($t_n$). For every flow $f_j$ that traverses the Mix-Zone $m$ there is a probable time distribution for MO $h_{m,j}(\Delta t)$, where $\Delta t$ is the time MO has spent in the Mix-Zone. An estimation of this sojourn time distribution gives an insight for finding the activation time for a Mix-Zone making it transient.

Let us consider a set of only two flows $f_j$, $j = 1, 2$, converging at an egress point $g$. The probability that attacker $A$ makes an error in its assignment ($s$) depends on the number of MOs traversing a Mix-Zone $m$ (from (1)). This number in turn depends on the activation time of the Mix-Zone, the time spent by MO inside the Mix-Zone, and the interarrival times of the MOs. Simplifying the problem further, let us consider only one MO ingressing the Mix-Zone $m$ at $t = 0$ from flow $f_j$ and another MO from flow $f_z$ ingressing the Mix-Zone $m$ at $t = \xi$. Figure 2 clearly indicates that as the number of MOs predicted by attacker $A$ (at the egress points) increases, the probability of error of prediction increases too. This implies that it would always be difficult for the attacker to predict in a heavy traffic flow scenario. This is favourable phenomenon for privacy protecting Mix-Zones on the road network which could work as well on the nonroad networks too.

The appropriate mixing occurs when the probability of error in making these assignments by attacker $A$ increases substantially. To check that we can use the well known Bayes Decision Rule, this rule minimizes the probability of error by choosing the hypothesis with the largest a posteriori probability.

In accordance with the decision-theory problem, attacker $A$ must classify each egress event $g$ happening at time $t_g$ as coming from one of $f_j$ possible entering flows. According to Bayes’ theorem, the a posteriori probability that an observed event $x$ belongs to flow $f_j$ is

$$p(f_j \mid x) = \frac{p_j(x) \cdot p(f_j)}{\sum_i p_i(x) \cdot p(f_i)}$$

Here, $j = 1, 2,$

(3)

where $p_j(x) = p(x \mid f_j)$ is the conditional probability of observing $x$ knowing that $x$ belongs to $f_j$ and $p(f_j)$ is the a priori probability that an observed egress event belongs to flow $f_j$.

Bayes’ probability of error [28] is then given by

$$p_e(p_1, p_2, \ldots, p_n) = \min\{p(f_1) \cdot p_1(x), p(f_2) \cdot p_2(x), \ldots, p(f_n) \cdot p_n(x)\}.$$  

(4)

The a priori probabilities depend on the intensity $\mu$ of the flows and are equal to

$$p(f_j) = \frac{\mu_m}{\sum_i \mu_i};$$

(5)

this is the error an attacker will make when he/she is having all the a priori and a posteriori trajectory timing information of a particular MO. The higher the probability of this error, the more effective the $m$th Mix-Zone. This in turn is dependent on the intensity of flow $\mu$. The intensity of flow though is not in control but timing the Mix-Zone in accordance with the intensity of flow of traffic will result in increasing the probability of error $p_e$ of the attacker which is the desired result. This kind of transient Mix-Zone can be emulated using various methods, for instance, timing the Mix-Zone in tandem with the green signal of the traffic light. It increases the probability of acquiring a good intensity flow.

3. Construction of the Traffic Signal Mix-Zone

A Mix-Zone can be simply constructed by first identifying the intersecting flows of traffic and then defining a rectangle with its center on the intersection (Figure 2). It can also be constructed by considering the Mix-Zone geometry, the statistics of the user population, and the spatial and velocity constraints on the movement patterns of the MOs [21]. These techniques very well describe the spatial construction of
the Mix-Zones. But the resilience to the temporal attacks also needs to be incorporated in the Mix-Zones [29]. In our work we have considered the spatiotemporal aspect of the Mix-Zones to address some of the timing and velocity based attacks on the MOs privacy. For the descriptive purpose we have taken the time dependent and time independent aspect of the Mix-Zones separately as static and transient Mix-Zones, respectively.

3.1. Static Mix-Zones. The construction of a static Mix-Zone can be modified by the specific geometry of the actual road junction [28]. One of the possible modifications has been shown in Figure 1. The boundaries of this Mix-Zone play an important role in defining the ingress and egress of MOs. Also, the MOs will be disconnected from the trusted third-party server responsible for providing LBS till they remain in the Mix-Zone. This will also enable 𝑘-anonymity of the location, 𝑘 being the number of users inside the Mix-Zone. If the Mix-Zone becomes static/time invariant then it would lead to strong timing attack possibilities for the MOs. The attacker can observe the time of ingress 𝑡𝑖𝑛(𝑖) and the time of egress 𝑡𝑒𝑔(𝑖) for each user ingressing and egressing the Mix-Zone. Let us consider an example of an anonymity set 𝐾 = {𝑎, 𝑏, 𝑐}. If the MO 𝑎 egresses with a new pseudonym 𝑑 and if the likelihoods of 𝑎, 𝑏, and 𝑐 exiting at time 𝑡𝑒𝑔(𝑑) are 0.3, 0.07, and 0.09, respectively, then, the mapping probabilities based on these likelihoods are

\[ P_{d→a} = \frac{0.3}{0.3 + 0.07 + 0.09} = 1.875, \]
\[ P_{d→b} = \frac{0.07}{0.3 + 0.07 + 0.09} = 0.437, \]
\[ P_{d→c} = \frac{0.09}{0.3 + 0.07 + 0.09} = 0.5625. \]

Thus, the attacker can compute that 𝑑 to 𝑎 is the most probable mapping and 𝑑 to 𝑏 is the least probable one.

3.2. Transient Mix-Zones. As stated in the last section, if the flow of the MOs is taken to be a Poisson process, given the mean arrival rate, \( \lambda_i \), on each incoming segment, \( l \), let \( \lambda_L \) denote the rate parameter which is the sum of the Poisson processes of each incoming segment, \( l \). Then,

\[ \lambda_L = \sum_{l} \lambda_i, \]

where \( l \in L \) represents the mean arrival rate of the entire road junction. If \( N(t) \) represents the number of users who had entered the Mix-Zone at time \( t \) since the beginning, then the probability of having \( N(t) = k \) is given by

\[ P[N(t) = k] = \frac{e^{-\lambda_L \tau} (\lambda_L \tau)^k}{k!}, \]

where \( \tau \) is the time period for which Mix-Zone will become effective. This timed Mix-Zone will be able to control \( k \) parameter and hence the number of users inside the Mix-Zone.

The problem to be addressed now becomes this timing which can assure privacy of the MO to a great extent. One of the possibilities is to have a time windowed Mix-Zone [21] of certain time duration. To apply this window to the road networks we have joined this time duration with the traffic signal time duration. The time interval till the signal is green is chosen as time duration of the window. This will be able to combat the timing attack mappings of the attacker. The reason is that most of the MOs will move with almost the same speeds. Hence the timing mappings will become futile for the attacker [30]. Even if this type of arrangement is not provided then also the traffic signals give another set of timing and possible movement information to the attacker.

Thus, in our system, the Mix-Zone will get activated for all the green signals at all the traffic junctions. MO location privacy will increase by the number of times it will pass through such Mix-Zones.

3.3. Secure Mix-Zones: The RSUs. The work presented here assumes that the provision of Mix-Zone is done only on the traffic signal enabled road junctions. Hence the boundaries defining the Mix-Zone are dependent upon the number of roads connected to the traffic signal junction and the flow statistics of the recent traffic through the junction as mentioned in our paper [31]. The defined boundaries are the gateways of the authentication process too of these Mix-Zones. The authentication process changes the message packets of the MO [32]. The packets will enter the Mix-Zone and this MO will carry this replaced messaging packet till it encounters another similar authentication process or Mix-Zone. The overview of the procedure of the authentication process is depicted by Figures 3 and 4.

4. Experimental Evaluation

The evaluation of our approach is being performed using an open source simulator gt-mobisim [33] which uses the maps from the US Geological Survey’s [34] TIGER maps [35]. The gtmobisim.jar file has been used in this evaluation which is being coded in the Java language.

4.1. Simulation Setup. This simulator can be used for generating mobility traces and query traces for large numbers of mobile agents moving in a road network. For our purposes we have used only the mobility traces along with the identification of traffic intersections. At the intersections the mobility flows are stopped to assume a traffic signal scenario. The Mix-Zones are activated at every start of the mobility flow of MOs and deactivated when the flow stops. A total of 40 mobility flows were deployed over the area, generating typically 1210 MOs in a light traffic scenario and 2000 MOs in a heavy traffic scenario. The radius of Mix-Zone is a constant \( R = 100 \) m. We simulate a mobile network for 20 minutes with MOs moving at a maximum speed of 50 km/h.

4.2. Simulation Metrics. We consider an attacker that can construct a mobility profile of each Mix-Zone by measuring the time at which the MOs enter/exit Mix-Zones. The attacker accumulates the trajectory data of the MOs moving in
Figure 3: Poisson distributions of the uncertainty entropies of \( j \)th flow and \( \Delta t \) time of MO (moving object) spent in the Mix-Zone. Number of MOs predicted by attacker \( A \) at egress points varies from (a) to (d).

Figure 4: Registration, subscription, and the process of pseudonymization of the MOs in the Mix-Zones.

the said area using the matrix \( T_n \) for each MO. He/she also has a record of the directions of these trajectories in the matrix \( D_n \). These directions’ matrix however contains the probabilities measured by the attacker using the information of ingress and egress points \((l, m)\), respectively. Hence each Mix-Zone is associated with a direction matrix \( D_{l,m}^j \). The matrix may be made using the relationship between ingress and egress points [18]. The mixing effectiveness \( ME_i \) for each Mix-Zone is the ratio between the number of correct matches of ingress and egress events of attacker \( C_i \) and the number of MOs \( N_j \) under consideration.
5. Results and Discussion

The uncertainty in the attacker’s calculated directional trajectories of MOs at each Mix-Zone is measured using the entropy measure as given in (1). An analysis of this unpredictability corresponding to the number of vehicles or the anonymity level each time the Mix-Zone becomes effective has been shown in Figure 6(b). The tracking success is the percentage of MOs that can be tracked over \( j \) consecutive Mix-Zones. If \( M_r \) is the number of MOs successfully tracked and \( M \) is the total number of MOs that have traversed \( j \) consecutive Mix-Zones, then, the tracking success \( TS(j) = \frac{M_r(j)}{M(j)} \) measured as percentage as shown in Figure 6(a).

The cumulative entropy of a particular MO \( m \) on the other hand is

\[
H(m, J) = \sum_{i=1}^{J} H_i(m) \times m. \tag{9}
\]

Here, \( J \) is the total number of Mix-Zones traversed by MO \( m \) over the said road network area. Figure 5(b) gives an account of this cumulative entropy over all the traversed Mix-Zones. Figure 5(a) again provides a clear picture of the tracking success and in turn the robustness of the road network over all the traversed Mix-Zones.

The heavy traffic scenario also shows a similar trend but the good decline is observed after more than seven traversed Mix-Zones and for complete elimination this will be going well beyond the decade [36]. Hence, in this regard transient Mix-Zones can easily increase the number of Mix-Zones and hence the privacy of the user. These specialized Mix-Zones increase their number not just through the change in their location but also through the change in their appearance. As they are appearing only for the time of traffic flow and get deactivated otherwise. Thus, even for the case of heavy traffic flow high privacy levels can be achieved without increasing the area or the spatial Mix-Zones.

6. Conclusion

The presented work tries to work out more robust Mix-Zone technique for location privacy protection. It has been shown here that the potential privacy threat with location based privacy protection techniques is the knowledge of the temporal metric to the possible privacy attacker. Here we have incorporated this temporal knowledge of the moving object (MO) to enable the Mix-Zones. Our analyses and their results show that this temporal shift improves the privacy of MO on the road network a great deal. This work is very close to the real world road networks. Hence, application to the real traffic signals for enabling transient Mix-Zones with collaboration of the Location Based Services Provider could be foreseen as the future work. The time duration for the activation of the transient Mix-Zones can also be analysed and modified for the real world. Not only does this work aim to be user friendly but its simplicity will also provide its good applicability for the LBS provider.

Competing Interests

The authors declare that they have no competing interests.
References


Submit your manuscripts at http://www.hindawi.com