Preserving Trajectory Privacy in Participatory Sensing Applications

Gauri R Virkar¹, Sanchika A Bajpai²

¹, ²Department Of Computer Engineering, BSIOTR, Wagholi, Pune, Maharashtra, India

Abstract: With the advancement of technology in fields of wireless communication, different mobile communicating devices equipped with variety of embedded sensors and powerful sensing have been emerged. Participatory sensing is the process that enables individuals to collect, analyze and share local knowledge with their own mobile devices. Although the use of participatory sensing offers numerous benefits on deployment costs, availability, spatial-temporal coverage, energy consumption and so forth, it has certain threats which may compromise the participator’s location and their trajectory data. Henceforth, to ensure the participators’ privacy is the most urgent task. The existing proposals emphasized more on participators’ location privacy and very few of them consider the privacy of the trajectories. The theoretical mix zones model are improved by considering time factor from the viewpoint of the graph theory and mix zone graph model has been presented. Mix Zone is a area where no application can track the user. Pseudonyms are generally changed in this mix zone. The existing system has applied only a single mix zone. The privacy of the user can be further enhanced by applying Mix Zone Graph Model at multiple sensitive locations rather considering only single mix zone. So the main objective here is to apply Mix Zone Graph Model at multiple sensitive locations in order to preserve and enhance the user privacy. Further this model considers only sensitive trajectories for providing privacy thereby reducing overall storage space. The results shows that the proposed system is having better privacy level as compared to existing techniques and overall the system is time as well as space efficient.

Keywords: Location privacy, Mix zone graph model, Multiple mix zone, Participatory sensing, Trajectories

1. Introduction

The growth of mobile phones along with their pervasive connectivity leads to the development of a new sensing technology model called as participatory sensing [1] systems. Here mobile devices carried by the individual’s acts as a sensor thereby eliminating the need of deploying sensors at particular areas. Participatory Sensing facilitates the participator to sense, analyze; collect and share the sensed information from their surrounding environment using their mobile phones. For example mobile phones may report actual (continuously) temperature or sound level; likewise, vehicles may notify about traffic conditions.

The vast amount of trajectory data gets collected and progressively increase as the participators sense the data. Trajectories are defined as the path followed by the moving object which is generally represented by \((x, y, t)\) where \(x\) and \(y\) are the location coordinates and \(t\) denotes the timestamp. In typical participatory sensing applications, the data reports generated as an output may reveal participators’ spatial temporal information. Adversary can obtain some valuable results from the published trajectories. The collected data may be used to deduce private information about the user. So to ensure the participators’ privacy is the most urgent task. The gathered information is very crucial to the participatory sensing systems as their deficiency endangers the success of such systems. Therefore the need is to preserve the privacy of the participatory sensing users by protecting their trajectories.

Mix Zone Graph Model [2] is one of the existing approaches for providing privacy to the trajectories of the participators. A mix zone is a region where no applications can track user movements. It is the region where the users can change their pseudonyms without being observed by the adversaries. A pseudonym is a uniquely generated random number. Each participator enters a mix zone with a pseudonym and exits the mix zone with other pseudonym. The use of pseudonym breaks the continuity of a user’s location exposure thereby protecting the future locations of the users. However, existing mix zone model solutions mainly focus on the development of single mix zone. Henceforth, for providing more security multiple locations are selected for applying mix zone graph model. Thus, multiple mix zone model [3] is used for providing maximal privacy to the trajectories of the participatory sensing users.

1.1 Location Privacy

Location privacy is defined as the ability to prevent other unauthorized parties from learning one’s current or past location. Traditionally, privacy of personal location information has not been a critical issue but, with the advancement of location tracking systems capable of following user movement twenty-four hours a day and seven days a week, location privacy becomes crucial: records of everything from the particular rack a person visit in the library to the clinics a person visit in a hospital can represent a very invasive list of data. Numerous systems could figure out the location of person. One of several original systems designed for position following could be the Global Positioning System (GPS). This technique makes use of satellites to aid devices figure out their own position. Generally, automated digital devices obtain information either through communication, observation, or inference.

1.2. Trajectory Privacy

A trajectory is the path that a moving object follows through space as a function of time. Example of trajectories could be monitoring of wild animals, birds, people, a soccer player,
etc. Trajectories may be uni-dimensional or perhaps multi-dimensional. Participatory sensing systems primarily depend on the collection of information across large geographic areas. The sensor data uploaded by participators are usually tagged with the spatial-temporal information when the readings were recorded the published trajectories for decision making. For example, merchants may possibly decide where to build a food store that could produce maximum gain by analyzing trajectories associated with consumers in a selected spot and also the Department of Transportation can make an optimized vehicle scheduling strategy by monitoring the trajectories connected with motor vehicles. However, it will add considerable threats to the participators’ privacy. Adversary may perhaps examine the particular trajectories which contain abundant spatial-temporal background information to be able to link numerous reports that are collected. Hence, it is crucial to be able to unlink the particular participators’ identities from sensitive data collection locations.

1.3 Existing Technique Limitation

TrPF, Trajectory Privacy Preserving Framework for Participatory Sensing Applications, is an existing approach which preserves the trajectories of the participators by applying Mix zone Graph Model at a single sensitive location. The problem here is that if an adversary is successful to guess the pseudonym of this single location Mix Zone Graph Model, the whole trajectory can be inferred.

1.4 Our Observation

Instead of applying Mix Zone Graph Model at single sensitive location, multiple locations can be considered as the candidates for applying Mix Zone Graph Model. As the number of locations increases, the number of pseudonyms to be cracked by an adversary increases. Thus the probability of successful attack by an adversary is reduced. An attack is said to be successful if an adversary is able to crack all the pseudonyms used in the corresponding mix zones. Consider a scenario where Mix Zone Graph Model is applied at three locations. The adversary will be able to deduce the whole trajectory only when he/she will be able to crack the pseudonyms at all three locations. Hence, as the number of mix zones increases the number of pseudonyms to be identified increases eventually increasing the privacy level.

1.5 Our Solution

In this paper, we propose an approach for preserving trajectories of the participators by applying Mix Zone Graph Model at multiple locations thereby enhancing the privacy level of the participators. Further, due to cost constraints, not all point of interests can be considered as the candidates for applying Mix Zone Graph Model. So the solution for selection and placement problem of the number of mix zones to be considered is been addressed here.

1.6 Our Contribution

Our contribution in this paper is as follows:-

- To secure location and trajectory privacy of the participatory sensing user by applying Mix Zone Graph Model.
- To secure multiple sensitive locations of the participatory sensing user.
- To prove that privacy of the user can be enhanced by protecting multiple sensitive locations instead of single sensitive location.

2. Related Work

Several work exists where location of the users’ as well as their trajectories are given privacy. Dummy location [4] is a mechanism of creating fake alias location of the user’s location in order to confuse the adversary. Location k anonymity is defined in [5] as a privacy approach designed to protect identification of an individual against a specific datasets. Another technique used for location privacy is obfuscation [6] where the user’s location is purposefully altered to lower the precision of the user’s spatial temporal information. This can be achieved using generalization or perturbation. Pseudonym [7] is a randomly generated unique identifier provided to each user before entering the sensitive area called as mix zones [8]. Mix zone is the area where a participators movement cannot be tracked by anyone. Pseudonym is generated to break any link present between the user’s identity and their events. Mix networks [9] are used to anonymize the channels used between the links between the reports submitted by the user to the system.

It is been observed that once a user’s trajectory has been identified, then it becomes easy to derive the locations of the users. Trajectory privacy schemes exist in the literature. Some of them are- dummy trajectories [10] where fake user location trajectories of the users are created. This technique provides privacy to the trajectories however the main problem is how to generate the exact look alike fake trajectories. Another technique proposed is suppression based [11] technique where the whole trajectories are generally suppressed with the assumption that the adversary would not be able to infer the user’s information since the whole trajectories are not exposed. The main threat to this approach is that essential data may get lost during the process of suppression. Trajectory k-anonymity [12] technique proposes a scheme where every trajectory is generated such that a user finds it indistinguishable to guess the other k-1 trajectories. All these techniques deal with the whole trajectory and thus increases the storage space cost. Not all locations are sensitive, so providing privacy around these sensitive locations can only be considered instead of whole trajectories [13]. To overcome the defects above, a new scheme has been proposed to preserve the privacy of the trajectories at multiple sensitive locations.

3. Existing System

Most of the existing techniques focus on providing location privacy of the participators while few approaches consider preserving the trajectories of the user. An approach called as Trajectory Privacy Preserving Framework TrPF for participatory sensing applications has been proposed. The participators known as data collector sense the spatial-
temporal information through their mobile device. This information is stored by the Report Server which generates data reports that are eventually stored on Application Server. Any authorized end user or participator can view these reports. Trusted third party servers are used for maintaining security to end users or the data collectors. Fig.1 shows the overall architecture of TrPF system.

In this approach the trajectories of the participators are preserved using Mix Zone Graph Model. Not all, but only sensitive trajectories are considered for while applying mix zone graph model. Firstly, a sensitive location, o, is taken as centre and a sensitive area is constructed around it. The trajectories intersecting the sensitive area are said to be as sensitive trajectory segments. Mix zone graph model is then applied on these segments. Thus, the trajectories of the participators are preserved.

3.1. Limitations

- Only single sensitive location is considered.
- Requires more time to process query as only raw trajectories are considered.

4. Proposed System

4.1 Architecture

The existing solution considers only a single sensitive location while constructing mix zone graph model. This leads to the lack of a systematic approach for global privacy protection. Henceforth to overcome this drawback, the proposed system defines multiple sensitive locations around which multiple mix zone graph model will be applied. Not all point of interests can be considered as the candidates for applying mix zone graph model. The main reason for this the available cost constraints which eventually limits the number of mix zones that one could deploy. So the problem is to address the multiple mix zone graph model’s placement. This is an optimization problem.

The proposed system can be explained as follows which is shown in Fig.2 – Firstly, the data collectors sense and provide their spatial temporal information to the Server using their mobile phones. Consider a participator provide their current location(x,y) using GPS embedded in mobile phones. As the participator moves his/her locations get stored on the server, eventually forming location traces i.e trajectories. These trajectories must be preserved from an adversary in order to the preserve the privacy of the participators. Mix Zone Graph Model technique has been used for providing privacy to the participator. In proposed system we consider multiple locations as the candidates for applying Mix Zone Graph Model. Due to cost constraints, not all locations can be considered as point of interests where the model can be applied. Hence selection and placement of multiple locations to be considered for applying Mix Zone Graph Model is the problem to be addressed whose solution is given next. After receiving multiple locations as an output of Multiple Mix Zone Placement Model, Mix Zone Graph Model is applied at all these locations. Meanwhile, an end user may query on this data store on the server and server may provide appropriate result.

For instance, consider a scenario of Online Car Booking System where end users i.e the customers at any time can book a car online through the system. The administrator depending on the availability of the drivers assigns a driver to the customer. The customer at any time can track their assigned driver. Considering the privacy of the driver, not all location trajectories of the driver should be visible to the customer. The driver who is the participator in this system provide their locations to the administrator using their mobile phones. Trajectories of driver’s are stored on the server which can be viewed by the administrator and driver itself. No the other party should be able to view the whole trajectory of a driver. Hence, Mix Zone Graph Model is applied at multiple sensitive locations of the driver thereby not allowing the customer to view the whole trajectory of a driver. The sensitive locations of the driver like his house, hospital, gym, work place, etc. must not be able to be known by the customer or an adversary. Applying Mix Zone Graph Model at multiple locations prevents an adversary or an end user from inferring the whole trajectory of the participator. Thus preserving the trajectory privacy of the participatory sensing user.

4.1.1 Solution - Multiple Mix Zone Placement Model

This approach generally determines the number of positions where Mix Zone Graph model has to be applied. Basically this approach first finds the points (vertices), whose removal makes the graph disconnected. Such points are called as articulation points. This partitions the graph into
disconnected components thus eliminating the need of pairwise connections between them. To refine the quality of solution further, the set of independent vertices are found. These are the vertices that are not adjacent to each other. Finally, the number of mix zones are limited by the given cost constraint.

Consider the following Graph $G = (V,E)$ where vertices $V$ represents Points Of Interests of a participator and $E$ represents the road segments connecting POIs. The first step is built on the observation that partitioning $G$ into several disconnected components is helpful to eliminate the pairwise connections across these components. Therefore, we are seeking for vertices whose removal disconnect the graph. Such vertices are typically referred to as articulation points in graph theory. Take the area graph in Figure 3 as an example. Any route from 1 to 9 or from 1 to 12 needs to go through vertices 6 and 10. Therefore, 6 and 10 are articulation points in this graph. If a mix zone is deployed at vertex 6 or 10, a pseudonym appears at any vertex in the bottom part of the graph cannot appear at vertices 9, 12, and 11. Hence, the total number of pairwise associations is reduced.

After $G$ is partitioned into disconnected components, the mix zone deployment in each component is further refined to improve the solution quality. In graph theory, an independent set refers to a set of vertices that are not adjacent to each other. Hence, if all vertices that are not in an independent set are selected as mix zones, there will be no pair-wise association between the vertices in the independent set. Again, refer to the bottom part of Figure 3 as an example. Circle highlighted vertices, $\{1, 8, 3, 5\}$, form a maximal independent set for the lower part of the graph. If vertices $\{2, 4, 6, 7\}$ are selected as mix zones, a user Alice’s pseudonym appears at vertex 1 will not appear at any other vertex in the independent set. As a result, Alice’s past and future locations on her trajectory are protected, even though her identity gets exposed at vertex 1. Finally, there is a need to control the number of mix zones to meet the cost and service constraint. At the last step of our algorithm, we iteratively remove the vertex that introduces the least number of pairwise association increment from the mix zone candidate set selected by previous steps until cost constraint is met.

4.2 Mathematical Model

Let $S = \{I, P, O\}$

$I$ = Input

$O$ = Output

$P$ = Process.

$I= \{SI, GQ\}$

$SI$ = Sense Information

$GQ$= Generate Query

$P = \{TR, MMPM, MZGM\}$

$TR$ = Generate trajectories.

$MMPM$ = Determine multiple locations.

$MZGM$ = Apply Mix Zone Graph Model.

$O = \{PR\}$

$PR$ = Provide results of the generated queries.

Fig. 4 represents the mathematical model of overall proposed system.

4.3 Algorithms Used

4.3.1 Graph Construct Algorithm

This algorithm is used to construct mix zone graph model which is represented by a graph $G(V,E)$. A mix zone graph model has been proposed such that Directed Weight Graph (DWG) is represented by $G = (V, E)$, where $V$ represents set of vertices that are constructed as the pseudonyms.

$E$ represents set of edges that represent the participants’ trajectory mapping from the ingress to the egress in the sensitive area.

Algorithm 1 Graph Construct

Input :- Trajectory $T_r$ and pseudonym set $P$.

Output :- Directed Weight Graph ($G$).
1: Procedure
2: Define sensitive location and construct sensitive area around it such that $S_i = \{o, r\}$ where $o$ is sensitive location and $r$ is the radius.
3: Determine the set of sensitive trajectory segments $T_f$.
4: Randomly select ingress pseudonym $P_i$ and assign it to the vertex $V_i$.
5: Randomly select egress pseudonym $P_j$ such that $P_j \neq P_i$ and assign it to the vertex $V_j$.
6: Construct Edge $E_{ij}$ such that $E_{ij} \rightarrow (V_i, V_j)$
7: Assign weight $W_{ij}$ to each edge using Weight Construct algorithm.

### 4.3.2 Algorithm 2: Weight Construct

This algorithm is used to find weights of the edges formed in the graph of the mix zone graph model. Here,

- $V_i$ represents participator entering the mix zone.
- $K$ represents total number of participators entering mix zone.
- $P_i$ represents ingress pseudonym of a participator.
- $P_j$ represents egress pseudonym of a participator.
- $t_{ingress}(V_i)$ represents time at which participator enters the mix zone.
- $t_{j} \text{ to } t_{j+1}$ represents time interval during which participator exists from the system.
- $P(V_i, t)$ represents the probability that a single participator exits the mix zones between time interval $[t_j, t_{j+1}]$.
- $P(V_i', t)$ represents the probability that all participator exits the mix zone.
- $\Delta_t$ represents time interval during which participator exists from the mix zone.
- $f(\Delta_t)$ is the probability density function (PDF) of data collection time in mix-zones.
- Therefore,

$$P(V_i, t) = \int_{t_j}^{t_{j+1}} f(t_{ingress}(V_i)) f(\Delta_t) dt$$  \hspace{1cm} (1)

The above mentioned equation represents probability of a single participator exiting from the mix zone. Thus, the probability for all the participators exiting from the mix zone is given by (2)

$$P(V', t) = \sum_{i=1}^{K} P(V_i, t)$$  \hspace{1cm} (2)

However only one of them is a real participator. Hence, the probability that the participator $V_i$ exits in the time interval $[t_j, t_{j+1}]$ is denoted by $P(V_i [t_j, t_{j+1}])$ is given by the following conditional probability:

$$P(V_i [t_j, t_{j+1}]) = \frac{P(V_i [t_j, t_{j+1}])}{P(V_i [t_j, t_{j+1}])} , i = 1, 2, ..., k$$  \hspace{1cm} (3)

$W_{ij}$ is given by $W_{ij} = P(V_i [t_j, t_{j+1}])$ such that $w_{ij}$ is between 0 to 1 and $i \in [1, k]$ and

$$\sum_{j=1}^{k} W_{ij} = 1$$  \hspace{1cm} (4)

The Weight Construct algorithm is given as follows:-

#### Algorithm 2: Weight Construct

**Input**: $t_{ingress}$ and $\Delta t = [t_j, t_{j+1}]$ and $\Delta'$

**Output**: Edge Weight $W$

1: Procedure

### 4.3.3 Algorithm 3: Optimization Of Multiple Sensitive Locations.

This algorithm generally determines the number of positions where mix zone graph model has to be applied. Basically this algorithm first finds the points (vertices), whose removal makes the graph disconnected. Such points are called as articulation points. This partitions the graph into disconnected components thus eliminating the need of pair wise connections between them. To refine the quality of solution further, the set of independent vertices are found. These are the vertices that are not adjacent to each other.

**Input**: A graph $G$ and $Z$.

**Output**: A set of at most NP selected mix zone positions

1: Procedure
2: Find articulation points in the given graph.
3: Find maximal independent set.

### 5. Evaluation and Results

Here, the real time data is taken as an input for the system. As explained prior, the participator is providing their location $(x, y)$ and timestamp $(t)$ to the Server using their mobile phones. The trajectories are stored in the form of $(tid, x, y, t)$ on the server where $tid$ represents the trajectory ID. Sensitive locations are considered around which Mix Zone Graph Model is applied. Meanwhile the end user can access the relevant data on the server. The Online Vehicle Booking System is built as a website using C# ASP.Net whereas the participator provides its spatial temporal data to the Server using Android mobile devices. Participator side module is developed using Android Programming in Java. The server has two modules- Admin module where administrator at any instance can view all the trajectories of the driver along with the driver’s sensitive locations as shown in Fig.5. The sensitive locations are indicated using flag. The module is the Sub-Admin’s module where sensitive locations are applied with Mix Zone Graph Model. Thus Sub-Admin cannot view sensitive locations of the driver as shown in Fig.6. The customer i.e the end user at any time can query to see the trajectories of the driver. The customer can view only the trajectories of the driver allocated to him.
This work aims in proving that the privacy level of a participator can be enhanced by applying Mix Zone Graph Model at multiple sensitive locations instead on single sensitive location. This can be proved by measuring the rate of successful attacks on single mix zone as compared to multiple mix zones. An attack is successful if the adversary finds out the corresponding pseudonym used by a user in the side information. The success rate of an adversary is the ratio of number of successful attacks over total number of attacks.

Fig. 7 shows the attack success rate when different number of mix zones is applied where X axis represents number of mix zones to be deployed at various sensitive locations and Y axis represents the rate of successful attack.

The graph shows that the rate of successful attack is high when number of mix zones is less. It shows that as the number of mix zones increases eventually the rate of successful attack decreases thereby improving the level of privacy. The reason for this is on increase in number of mix zones successful attack rate decreases because the adversary has to crack the corresponding number of pseudonyms in order to deduce the whole trajectory. This becomes sustainably simpler for an adversary with single mix zone as only one pseudonym has to be cracked. So as the number of locations where mix zone graph model has to be applied increases, the privacy preservation of trajectories increases. Thus, the proposed scheme offers better privacy as compared to the existing systems.

Fig. 8 shows the graph indicating comparison between the existing system and proposed system in terms of time computation. The number of sensitive locations are taken on X axis and Y axis represents time required in seconds for the computation of the algorithm. The graph shows that as the number of sensitive locations increases the existing system having single mix zone graph model requires considerably more time to compute and provide the results as compared to the proposed system. Hence we can say that proposed system is time efficient when compared with the existing system.

Another advantage of the proposed work is that it requires less storage space as compared to the existing techniques. Previous work like Dummy trajectories and trajectory k-anonymity stored all trajectories for providing protection. Given t trajectories and each trajectory contains N segments then the storage space required will be O( N* t ) to store total t trajectories. Whereas trajectory mixes zone graph model approach requires only pseudonym to be stored. Only sensitive trajectory segments are considered here and not all trajectories. Hence storage space required for this approach is quite less as compared to the previous work. Further, the increase of trajectories may not affect the number of pseudonyms too much. By comparison, our proposal has lesser storage memory than that of the other proposals. Thus, the results show that the proposed system is time as well as storage efficient and provides better privacy as compared to the existing privacy preserving techniques.

### 6. Conclusion and Future Work

Participatory sensing leverages the ubiquity of mobile phones to open new perspectives in terms of sensing. The analysis has revealed that virtually all applications capture location and time information. The collected data is been stored in form of the trajectories. The privacy of these trajectories needs to be preserved. Trajectory Mix zone Graph model is been used here for providing privacy to the trajectories of the participators’. This approach proposes multiple sensitive locations to be considered for applying Mix Zone Graph Model.
Model as opposed to single sensitive location. The results proves that applying mix zone graph model at multiple sensitive locations as compared to single sensitive location increases the privacy level of the participator. Hence the proposed system provides better results as compared to the existing techniques in terms of increased privacy level and reduced storage space as well as time. In future, mix zone graph model can be applied on multiple sensitive locations of semantic trajectories.

References


