IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Trajectory Anonymity for Privacy in Location Based Services

Neha Jain[*1], **Kamlesh Lakhwani**[2]

[*1]M. Tech. Student, Suresh Gyan Vihar University, Jaipur, Rajasthan, India

[2]Asst. Professor, Suresh Gyan Vihar University, Jaipur, Rajasthan

kinshika@gmail.com

Abstract

The rapid development of location based services i.e. Navigation has enabled tracking of user more accurately than ever before and the tracking has increased. These technologies provide the services based on the geographical location of the current user. However, the personal location information generated by such technologies is at risk of being misused unless protection capabilities are built into the design of such technology. These concerns may ultimately prevent society from achieving the benefits from these technologies available to customer. The assumption of the increasing location-based industry is that corporations will own and control location and other information about individual customer. Tracking technology is continuously expanding and providing the services based on the social needs. In this paper, we have designed and implemented an approach for protecting privacy in the use of location-based services that anonymized the Data Set of the trajectories with k anonymization. After k anonymization, each possible location sample of each trajectory is further anonymized using the space and time grouping. The major goal of this paper is to create a system that takes the input data set of the trajectories and creates anonymized version of this data set. This Privacy protection technique will help to protect the whole trajectory of the user. Implementation is done in C language and tested on GPS trajectory dataset (Microsoft Research Asia) Geolife project on 64 bit dual core 2.0 GHz with 2GB RAM over a Ubuntu platform.

Keywords: Privacy-preserving, Data publication, Trajectory *k*-anonymity

Introduction

Location Based Services (LBS) have recently attracted the industries and research. Most popular services are Navigation and Point Of Interest. These services use the tracking devices like GPS. These tracking technologies will continue to expand in use and it is very important for the technologies to adapt to the social needs of the users and to the needs of society as a whole. The broader domain of tracking technologies, location-based services are of capabilities that allow users to access information relative to their own location on the earth (latitude and longitude). The personal location information generated by LBS devices is at risk of being misused by adversaries unless privacy protection techniques are built into the design of such systems. Location Based Services can be described as applications, which answers according to a geographic location. A Geographic location might be the input of a town name, zip code, street or the position of a mobile device user. Providing services based on knowledge where someone is or where they intend to go is the essence of LBS.

LBS (Location Based service) include large amount of spatial-temporal data and privacy related information in the queries. Although removing the user Identity from Trajectory data is not useful, spatial-temporal data is enough to re-identify the user that raise the privacy issues in historical trajectory data. If adversaries know the exact location of the user, he can identify the whole trajectory of that user. Our proposed solution can solve this problem. It will protect the whole trajectory of the user from adversaries. Location-detection devices such as cellular phones, Devices having GPS are growing rapidly. Thus, location privacy of users may be exposed when requiring for Location based services. Location privacy-preserving techniques have gained much attention in last a few years. Tracking devices are becoming available for use by third parties with whom the user does not have any contacts. The ability of the location-based service provider to collect location information of the users using Location Based Services has raised many personal information privacy issues. LBS may threaten user’s privacy. Malicious attacker may mine data with LBS provider to steal users’ location information. Subscribers are in the vulnerable position of not having control of their own location information privacy. Privacy threaten arises when the people trajectories are left behind and collected by the services provider. These trajectories are published for the new applications. For example analyzing the user's trajectory may help the business man in taking the financial decision of developing commercial shops or restaurants. Other example can be the traffic controlling system where trajectories of vehicles are collected and monitored by the government to optimize the traffic in the city. Although the publication of these trajectories database is beneficial for decision making processes, it create the serious threatens in users privacy. The trajectory data from these tracking systems are being made available for government, educational, commercial and commercial purposes for analysis or new applications. Now days, Location-detection devices are being used by society very rapidly. The goal is to develop an application which takes trajectory dataset as input and outputs the new dataset which is anonymized version of given dataset.Our aim is to modify the dataset in such a way that adversary having partial knowledge of the person .i.e. physical location of the user, will not able to detect disclose the whole Trajectory path of the user.

The rest of the paper is organized as follows. Section 2 of the paper gives the design of privacy preserving through graph partition. It also gives the previous work and techniques for privacy in Location Based Services and the current proposed algorithm with new enhancements (module 3 to module 6).Section 3 gives the detail of the test bed and analysis of the test bed under various trajectory dataset. The paper concludes with the Section 4 providing the conclusion and future work.

Related Work

Trajectory data privacy is a rather young research area that has received a lot of interests in past years. The Relational data base privacy protection methods are extended spatiotemporally to the community of trajectory data privacy. There are three types of techniques for Trajectory Privacy.

- Dummy trajectory confusion
- Suppression-based method
- Trajectory k-anonymity.

Dummy Trajectory Confusion

The first Trajectory Protecting privacy in a data publication with a simple approach to create dummy trajectories for confusion proposed in [11]. The authors proposed to generate dummy trajectories with existing in order to confuse the adversaries. For creating the confusion between original and dummy trajectories, dummy trajectories are generated under two principles.

- The movement patterns of dummies should be similar to actual trajectory
- The intersections should be as more as possible in trajectory data

Following the above two rules, dummy trajectories are generated trajectories with existing in order to confuse the adversaries.

Suppression-Based Method

Suppression based method is based on the assumption that different adversaries may have different part of users' trajectories. It is proposed in [6]. In case of suppression-based method it's reduce the probability of disclosing the whole Trajectory path of the user. In this method Trajectory parts are suppressed that reduce the probability of disclosing the whole Trajectory.

Trajectory *k*-Anonymity

Location k–anonymity approaches [2, 7, 8], protect user privacy by utilizing the current location of each user in the system. The concept of location k-anonymity for Location Based services was first introduced in [7] and later extended in [2] to deal with different values of k for different requests. The work described in [5] proposed a privacy system that takes into account only the spatial dimension. The k-anonymity concept was taken for relational database where database are published after removing the identity. It was realized that removing the identity is not enough to protect the privacy of individual user. So there are quasi-identifier fields that were suppressed or generalized to convert into k-anonymity dataset. Similar concept were applied in [8], Trajectory k-anonymity, here k trajectory are grouped and converted into generalized region having k anonymity among the k or larger group. Each location sample on trajectories is generalized to a region containing at least k trajectory samples.

Design of Trajectory Privacy-preserving through Graph Partition

There are different privacy protection models in trajectory dataset. We have enhanced the existing model [12], which solved the k anonymity problem using Graph partition. The proposed solution of this paper consists of three modules which are described below.

Existing Work in k-Anonymity

The method proposed in [12] consists of three main steps:

1. PRE PROCESSING: pre-processing of trajectories to form equivalent classes with the same time span;

ISSN: 2277-9655

2. GRAPH CREATION: constructing trajectory graphs for each equivalent class based on the notion of trajectory s-overlap;
3. PARTITIONING: partitioning trajectory graphs, connected components of size k(or larger than k) are retained to form trajectory anonymity sets. At last, each location sample on trajectories is generalized to a region containing at least k samples.

Proposed Solution

We enhanced approach suggested in [12]. The proposed method consists of Six steps:

1. **PREPROCESSING**: Pre-processing of trajectories to form equivalent classes with the same time span;
2. **GRAPH CREATION:** Constructing trajectory graphs for each equivalent class based on the notion of trajectory s-overlap;
3. **GRAPH PARTITION AND BEST K CALULATION:** It will detect the best k value for given trajectory. It will apply the graph partition method on different k and will calculate the loss of trajectory in each partition. Based on the k and the loss of trajectory it will declare the Best k;
4. **SPACE GROUPING:** In this module space grouping is done and the trajectories are combined in space domain having nearby time stamp on nearby location;
5. **USER INTERACTION:** User interface is implemented for interacting with user. The user is asked for time grouping option. It will show the list for which data set is required. Based on the input it will perform the time grouping;
6. **TIME GROUPING:** In this module the time grouping on the trajectory is done based on the user input;

Overall System Design

Fig.1 shows overall design having three layered architecture with flow of information within each modules of the system. Input layer consists of trajectory datasets and user interaction, Second layer consists of different modules of the system such as prepossessing, graph creation, partitioning with best k value calculation, space grouping, user interaction and time grouping modules. Output layer consists of anonymized trajectory data sets with time and space grouping with the intermediate results of module 3.

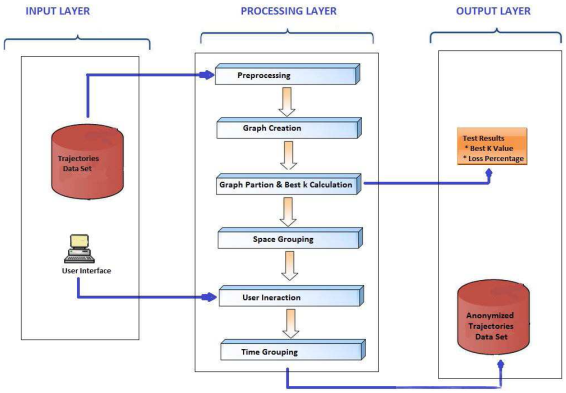

Fig 1: Design Diagram

Description of Proposed Solution

Pre Processing

In this module, the trajectories must have same number of locations sample with same time period. These collected trajectories form an equivalent class. If a trajectory has fewer samples, more intermediate samples are taken between the existing samples. It is done by inserting new sample locations between two samples by assuming that target is moving straight between two points. Fig 2. shows the original Trajectory in the left side and the Trajectory formed after pre processing in the right side.

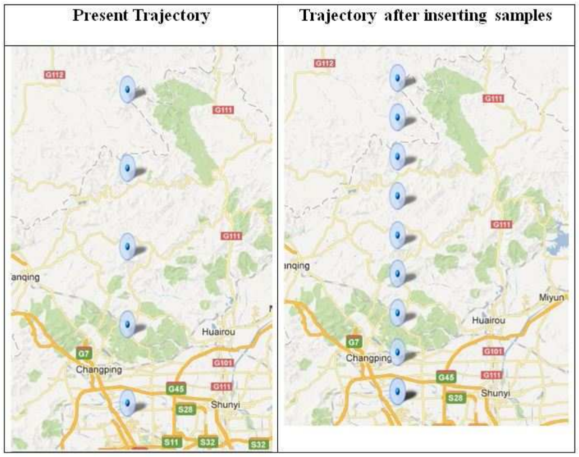

Fig 2: Trajectories on Google Map

Graph Creation

In this module, the trajectory graph is constructed for each equivalent class. Each node represents the each trajectory. There is edge between the two vertexes if both trajectory s- (i.e. x-axis and y-axis projection) overlap each other and weight of the edge is the distance between to trajectories.

Distance calculation: Let Tp and Tq be the two trajectories. Trajectory Tp has location (xp1, yp1, t1), (xp2, yp2, t2) and so on. Trajectory Tq has location samples (xq1, yq1, t1), (xq2, yq2, t2) and so on.

ISSN: 2277-9655

$$Dist(T_p, T_q) = \frac{\sum_{i=1}^{n} \sqrt{(x_{pi} - x_{qi})^2 + (y_{pi} - y_{qi})^2}}{t_n - t_1}$$

Let there are 10 trajectories inside equivalent class, for simplicity, here we assume that x-axis projection means s-overlap. Fig 3 shows trajectory distribution inside equivalent class.

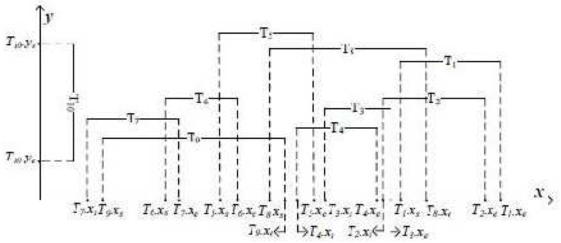

Fig 3: Trajectory Equivalent Class

Trajectory graph based on Fig 3 is created after calculation of trajectory distances as shown in Fig 4. We can see in Fig.3, T10 is not s-overlap to any other trajectory, so there is no edge with T10 in graph shown in Fig 4.

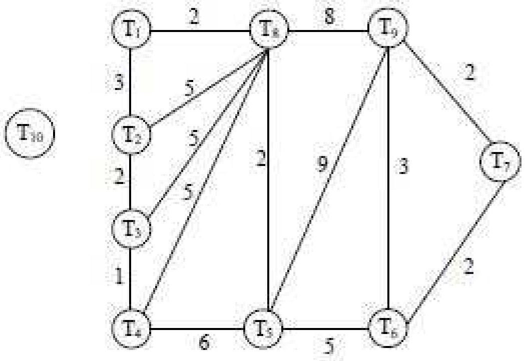

Fig 4: Trajectory Graph

Graph Partition And Best K Calculation

The aim is to partition the graph and determine the best k value for given trajectory. Graph partitioning method is applied on different k and will calculate the loss of trajectory in each partition. Based on the k and loss of trajectory it will declare the Best k. When there is partition on the above graph, it will form the new one having four partitions .i.e. Table 1. The Fig 5 represents the partition of the Fig 4 graph.

Table 1: Partition of the graph of Fig

Partition Number	Vertex
1	T1, T8, T5
2	T2, T3, T4
3	T7, T6, T9
4	T10

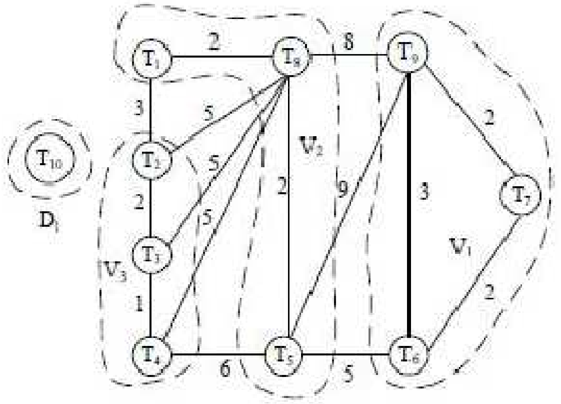

Fig 5: Graph Partition of Fig 4

Since T10 is not having neighbour, it will be removed after k=3 anonymous trajectories are created. So finally Fig 6 will have k-anonymous graph after removing T10.

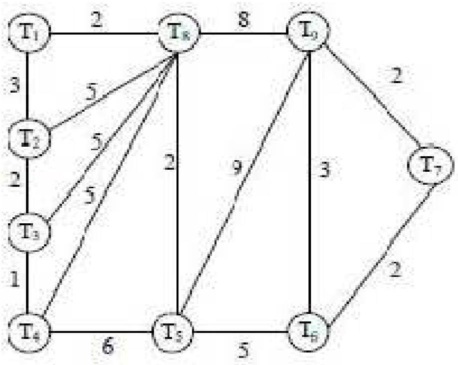

Fig 6: k-Anonymous Trajectory Data Set

First trajectory graph is created and partition of graph are created for all possible values of k. for all values of k trajectory loss is calculated. We assume twenty percentage loss is acceptable while finding best k values. The highest k value for which trajectory loss below twenty percent is assumed as best k.

```
Algorithm: FindBestK (TG)
Input: A trajectory graph TG;
Output: Best K value for which
information loss is under
Acceptable limit(less than 20%);
for each ( 2 <= k < Total
node) do
    Greedy k-node partition (TG,
    k);[12]
    calculate trajectory loss%;
    if (trajectory loss% < 20)
```

```
     bestk = k;
    end if
  end for
```

Space Grouping

In this module each location sample will be anonymized with other nearby location of other samples. For example in Fig 7 there are two trajectories having nearby time for nearest location samples.

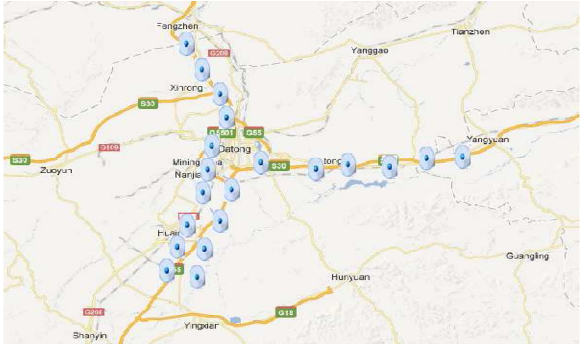

Fig.7: Two Trajectories on Google Map

After space grouping there will merge of two trajectory having location from Nanjiao to Yingxian as shown in Fig 8.

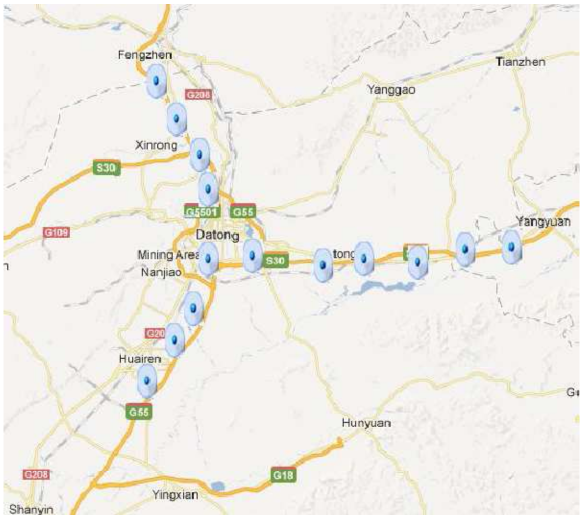

Fig.8: Merging of Trajectories from Fig 7

Space grouping is done for Best K values calculated in Graph partition and best K calculation step. Space grouping algorithm takes k equivalent trajectory as input and gives anomized version of equivalent class. Space grouping is done for each vertices of all trajectories in equivalent class. For each vertex distance is calculated between this vertex and corresponding vertex having same timestamp on all there trajectories. If this distance is less than minimum distance then these two vertices are merged and spacegroupingflag corresponding to these vertices are set so that these vertices further do not participate in anmization.

```
Algorithm : SpaceGrouping (TG,
Bestk)

Input: Trajectory equivalent class
EC=T1 , T2 , . . . , Tk
Output: Anomized version of
equivalent class
spacemergingflag = 0 for all
vertices
for each vertices V (for which
spacemergingflag = 0) on trajectory
Ti (1 < i <= k)
  calculate distance dij between
Vi and Vj;
Vj are vertices on Tj having same
timestamp as Vi
  if dij < Dmin
   merge Vi and vj;
   set spacemergingflag = 1
corresponding to these vertices;
  end if
end for
```

User Interface

In this module the user interacts with the system, user chooses the option for which trajectory dataset is going to be published. Currently it will have following type trajectory data set.

- Business Analysis
- City Planning
- Intelligent Transportation
- Point of Interest
- To Understand the Infrastructure

Time Grouping or Shifting

After selecting the type of trajectory data set, time grouping or shifting is done using time grouping module. Time sifting is done for “Intelligent Transportation” user option and for other Time Grouping is done.

Time Grouping

Time grouping is done for Best K values calculated in Graph partition and best K calculation step. Time grouping algorithm takes k equivalent trajectory as input and gives anomized version of equivalent class. Time grouping is done for each vertices of all trajectories in equivalent class. For each vertex time difference is calculated between this vertex and corresponding vertex having same location on all other trajectories. If this time difference is less than minimum time difference then these two vertices are merged and time grouping flag corresponding to these vertices are set so that these vertices further do not participate in anomization.

```
Algorithm : TimeGrouping (TG, Bestk)

Input: Trajectory equivalent class EC=T1 , T2 , . . . , Tk
Output: Anomized version of equivalent class
timemergingflag = 0 for all vertices
for each vertices V (for which timemergingflag = 0) on trajectory Ti (1 < i <= k)
    calculate time difference TIMEij between Vi and Vj;
    where Vj are vertices on Tj having same location as Vi
    if TIMEij < TIMEmin
        merge Vi and vj;
        modify TIMEi = TIMEj = (TIMEi + TIMEj) / 2;
        set timemergingflag = 1 corresponding to these vertices;
    end if
end for
```

Example

Assume the trajectory data is to be created for point of Interest. There are two trajectories TA and TB having following sample locations information on same day. First Trajectory data is sampled from 10:00 am and other is taken from 10.05 am. Table 2 is the trajectory of TA and Table 3 is the Trajectory samples of TB. After time grouping both TA and TB will have following location information.

Table 2: Location samples of Trajectory TA with Time

Time	10: 00 am	10: 10 am	10: 20 am	10: 30 am	10: 40 am	10: 50 am	11: 00 am
Location	L1	L2	L3	L4	L5	L6	L7

Table 3: Location samples of Trajectory TB with Time

Time	10:05 am	10: 15 am	10: 25 am	10: 35 am	10: 45 am	10: 55 am	11: 05 am
Location	L1	L2	L3	L4	L5	L6	L7

Table 4: Location samples of Trajectories TA and TB with time after TIME GROUPING

Time	10: 03 am	10: 13 am	10: 23 am	10: 33 am	10: 43 am	10: 53 am	11: 03 am
Location	L1	L2	L3	L4	L5	L6	L7

Time Shifting Example

Let us see another example. The user inputs the "Intelligent Transportation" option in
user interface module and there is Trajectory TX as shown in Table 5.

Table 5: Trajectory TX Location with Date and Time

Time	10: 00 am	10: 10 am	10: 20 am	10: 30 am	10: 40 am	10: 50 am	11: 00 am
Location	L1	L2	L3	L4	L5	L6	L7
Date	10/11/12	10/11/12	10/11/12	10/11/12	10/11/12	10/11/12	10/11/12

When time shifting is done the date will be changed to 11/11/12 which is one day shifting
as shown in Table 6.

Table 6: Trajectory TX Location after Time Shifting

Time	10: 00 am	10: 10 am	10: 20 am	10: 30 am	10: 40 am	10: 50 am	11: 00 am
Location	L1	L2	L3	L4	L5	L6	L7
Date	11/11/12	11/11/12	11/11/12	11/11/12	11/11/12	11/11/12	11/11/12

Description of Data set and Results

The Dataset which is collected from Microsoft Research, anonymized dataset with possible results.

Dataset

GPS trajectory dataset was collected in (Microsoft Research Asia) Geolife project. We measured different parameters with different 10 samples of trajectory dataset. Different samples of dataset consist of number of trajectories as shown in Table 7 and Fig 9 graphically. For example sample-1 has seventy one users' trajectories.

Table 7: Ten Trajectories Data Set with Number of Trajectories in each set

Sample Number	1	2	3	4	5	6	7	8	9	10
Number of Trajectories	71	175	322	395	86	28	54	34	49	171

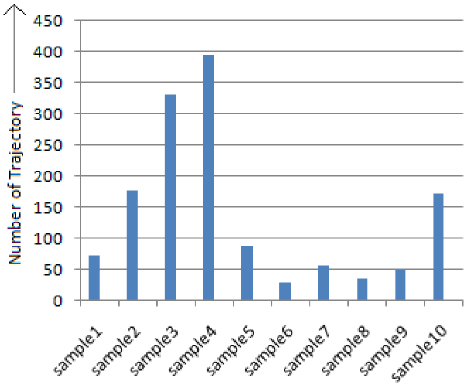

Fig 9: Graphical Representation of Trajectory Count (Table 7)

Fig 10 and Fig 11 show the best k value calculated and the percentage loss of the trajectories. Finally Fig 12 is the real trajectory on Google Map and the Fig 13 is anonymized Data Set on Google Map.

Best k Calculated for Different Samples

The application system was run on the sample trajectories, and it was found that the trajectory loss was very high as we increased the value of k. Finally the k values were selected having less number of trajectory losses and higher value of k to make dataset more anonymous. We found that value of k varies from 3 to 7 in most of the samples trajectories. Best k value plotted in Fig 10 X-axis represents the different samples and Y- axis represents the best k value for the sample.

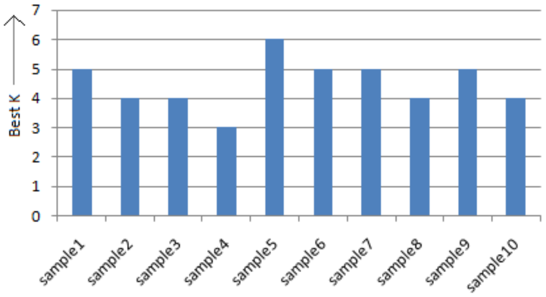

Fig 10: Best k value Calculated for different Samples

Trajectory Loss while Applying Best k Value

The trajectory loss was more than 50% as we increased the value of k. Finally best value of k selected as in earlier section. The corresponding loss for given trajectories are in Fig11. The X-axis represents the different samples and Y-axis represents the percentage loss of the trajectories that couldn't be anonymized.

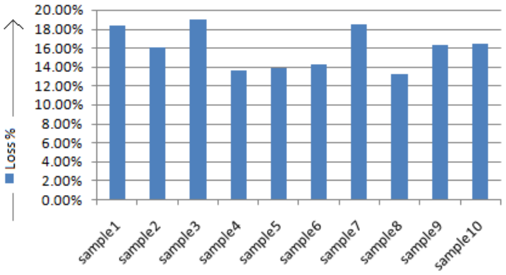

Fig 11: Loss % value calculated for different samples

Results

The Trajectory data set as shown on Google Map Fig 12 is without anonymization. This is one of the sample Data from Microsoft Research. The Trajectory which is not part of any partition is removed after processing. After applying the utility, the trajectory data set is given below in Fig. 13. We can clearly see from the map that the lonely Trajectory is removed after anonymization. Space grouping and time grouping added extra privacy so that location samples of the trajectory can't be differentiated from other's trajectory.

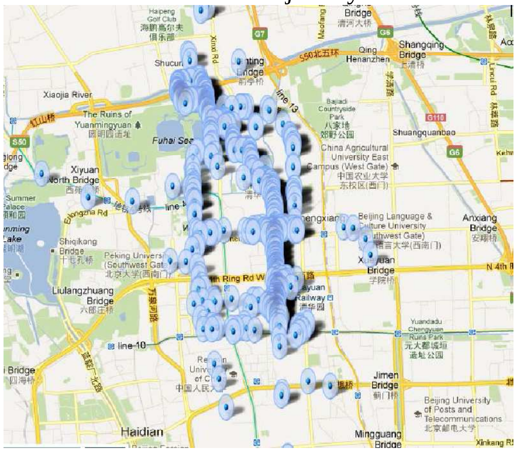
Xiaojia River
Shucun
The Ruins of Yuanmingyuan
Fuhai Sea
Xiyuan
North Bridge
Qing Henanzhen
Shangqing Bridge
Shuangquanbao
China Agricultural University East Campus (West Gate)
Beijing Language & Culture University
Anxiang Bridge
Peking University (Southwest Gate)
Liulangzhuang Bridge
line-10
Haidian
Jimen Bridge
Mingguang Bridge
Beijing University of Posts and Telecommunications

Fig 12: Trajectory Dataset without Anonymization

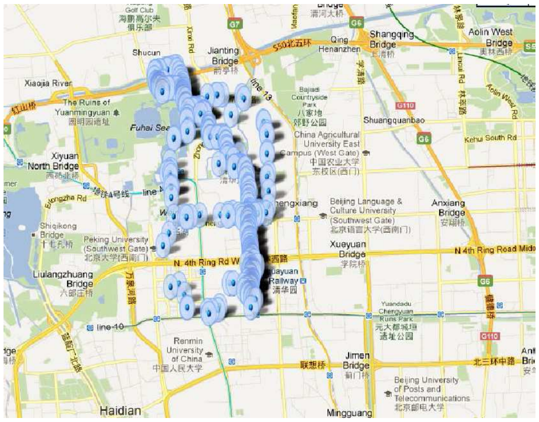

Fig 13: Trajectory Dataset with Anonymization

Conclusion And Future Work

In this paper, we have implemented the system that take one trajectory dataset as input and output the anonymized version of trajectory data set in space and time domain depending on the input parameter for generating the dataset. If data set is required for traffic analysis purpose, then there is time translation of 24 hrs. If dataset in independent of time, it will be translated randomly. With the concept of space merging, the probability of disclosing whole trajectory is very low.

These dataset can be used in following analysis

- business analysis
- city planning
- intelligent transportation
- Point of interest
- To understand the infrastructure

Future Work

One of attack which is possible is that if adversary knows the speed of the target which is unique among the trajectory, then he can generate the trajectory for the same if he has constant speed. We can have speed merging also if the output dataset is independent of the speed.

References

[1] B. Gedik and L. Liu, “A Customizable k-Anonymity Model for Protecting Location Privacy,” Proc. IEEE Int'l Conf. Distributed Computing Systems (ICDCS '05), pp. 620-629, 2005.

[2] B. Gedik and L. Liu, “Location Privacy in Mobile Systems: A Personalized Anonymization Model.” in Proc. of the 25th Int. Conf. on Distributed Computing Systems (ICDCS’05). Pages 620-629

[3] Chi-Yin Chow "Trajectory Privacy in Locationbased Services and Data Publication" ACM, Pages 19-29, 2011

[4] Femi Olumofin Piotr K.Tysowski, Ian Goldberq, Urs Henqartner "Achieving Efficient Query Privacy for Location Based Services" ACM, Pages 93-110, 2010,

[5] H. Kido, Y. Yanagisawa, and T. Satoh, “Protection of Location Privacy using Dummies for Location-based Services.” in Proc. of the 21st IEEE Int. Conf. on Data Engineering (ICDE’05). Page 1248

[6] M. Terrovitis and N. Mamoulis. “Privacy preserving in the publication of trajectories”. In roc. MDM 2008, Pages 65-72, 2008

[7] M. Gruteser and D. Grunwald, “Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking.” in Proc. of the First Int. Conf. on Mobile Systems, Applications, and Services (MobiSys 2003), Pages 31-42

[8] O. Abul, F. Bonchi, and M. Nanni. Never walk alone: uncertainty for anonymity in moving objects databases. In Proc. ICDE 2008, pages 376– 385, 2008.

[9] R. Cheng, Y. Zhang, E. Bertino, and S. Prabhakar. “Preserving user location privacy in mobile data management infrastructures”. In Proceedings of International, Pages 393-412, 2006

[10] Sharad Jaiswal and Animesh Nand "Trust No One: A Decentralized Matching Service for Privacy in Location Based Services" ACM, Pages 51-56, 2010

[11] T. H. You,W. C. Peng, andW. C. Lee. “Protecting moving trajectories with dummies”. In Proc. MDM 2007, Pages 278-282, 2007.

[12] Zheng Huo,Yi Huang, Xiaofeng Meng, "History trajectory privacy-preserving through graph partition", ACM, 2011