Anonymization and Publication of Trajectories by Sensitive Halting points Generalization

Rajesh N, Sajimon Abraham, Shyni S Das

Abstract*: The profuse use of wireless and GPS enriched mobile devices has left spatio-temporal trajectory traces in an enormous scale. The researchers and mobility management people are keen to extract and make use of the published mobility traces for their own developmental activities. The spatio-temporal traces publication is definitely a major privacy encroachment for the individuals/objects, especially for VVIPs. The publication of trajectory details urges us for a privacy preserved anonymization approach. The observation was the trajectory consists of halts and passes and anonymizing the sensitive halting points is adequate enough for the trajectory anonymization. This paper suggests a different approach, which derives the major halting points from the trajectories and anonymizes them by using the personalized generalization technique. The prototype referred to here safeguards the major sensitive halting points in an area zone, the size of which is specified by the user. This work uses Haversine measure for the spatial distance instead of Euclidean measure, since the former takes into account the spherical shape of the earth and gives an exact distance from sensitive to non-sensitive points. For the evaluation, the model mentioned here utilizes the real world dataset and the outcome proves that, the published trajectory has lesser information loss and greater privacy than the anonymity methods that exists now and it can be used safely for mobility related applications and developments.

***Index Terms*: Location based system, Published trajectory, Sensitive halting points, Spatio-temporal trajectory.**

I. INTRODUCTION

An enormous amount of mobility traces or data are collected by the pervasiveness of location aware devices like GPS embedded vehicles, tablet, RFID, smart mobile phones, PDA's and location based social networks (eg; Facebook and Twitter which use check-ins). This massive movement big data are being collected by the developers of mobile service providers and location- based applications in every moment. All these traces come under the category of spatio-temporal data and represent the user's movement trajectories.

The trajectory is the term given for a moving object's mobility traces collected over a specific period of time, which contains the spatial information such as location coordinates and the temporal coordinate as time. According to the authors in [1], the trajectory data can be classified into explicit and implicit trajectory data based on their spatio-temporal continuity. Since the GPS data is having the strong spatio-temporal continuity, the researchers mostly follow this type of explicit trajectory data. The collection of various user trajectories forms a massive spatio-temporal repository and it is made available for the researchers for further investigations. For knowing the individual's behavior by extracting behavioral patterns from the datasets plays a vital role in the analysis. The trajectory data can be used for the new developments in the applications such as optimization of the traffic congestion on the roads, location-based advertisements, urban planning, managerial decision regarding the starting of a new supermarket and many more. The trajectory data in its original form contains many individual oriented data, and publishing in its raw form could possibly result in the disclosure of sensitive information even if the explicit attributes like personal identification number, name, and phone number are not included in the published data. Suppose an employed person's daily mobility traces reveal their places like home and work, with this the intermediate/future places can be easily predicted. This will certainly harm the privacy of the individual. But most of the individuals were not concerned about this breach of privacy and they voluntarily provide the personal data and location updates to the applications freely. Even if the users are not aware of the privacy harms, the researchers are seriously concerned about this matter and try to find out new methods to protect against privacy attacks. Merely removing the explicit identifiers from the dataset alone does not satisfy the anonymity concerns, since the published data may be linked with publicly available other databases through known identifiers termed as quasi-identifiers for recognizing the objects. In the case of trajectories, they contain spatial and temporal information about the locations of the object or individual visited or passed, which is itself a crucial QID (Quasi-identifier) for the adversaries. Suppose an individual starts his trajectory every day having long stay and proceeds to a place where they spend some time and return to the place where they started, the malevolent will certainly search the address in the known publicly available databases and link this information to get the details of the individuals. These types of attacks by the adversaries are called linkage attacks and it can be prevented to an extent by using the anonymization techniques. The process of anonymization is that the concealing of vital or sensitive information and/or identity of the object from the adversaries. In Fig.1, the first trajectory is a normal trajectory, in which we could see certain clusters being present in some parts. This gives the information that the user has some activity in and around there for a specific period. The second trajectory is the refinement of the first one and such a cluster has been

Revised Manuscript Received on August 03, 2019.

Rajesh N, School of Computer Sciences, Mahatma Gandhi University, Kottayam, India.

Sajimon Abraham, Dept.of Computers& IT, School of Management and Business Studies, Mahatma Gandhi University, Kottayam, India.

Shyni S Das, Department of Computer Science & Applications, S.A.S. S.N.D.P. Yogam College, Konni, Pathanmthitta, India.

identified as some geographical location where the user has stayed for specific purposes. So, our task is to anonymize these sensitive points.

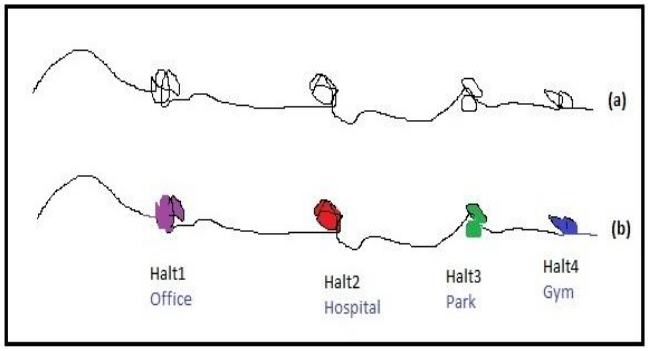


Fig. 1: Normal trajectory Vs trajectory with halting point information

The process of anonymization starts with the collection of user trajectories into databases and removal or hiding explicit identifiers and sensitive identifiers from the real trajectory data. After that extract semantic points from the QIDs and finally mask these points using the appropriate anonymization techniques such as generalization, suppression or perturbation. Many works use the model k-anonymity [2] for the anonymization of location traces as well as the whole trajectories. But from our point view it may be noted that, protecting sensitive halting samples is enough rather than anonymizing the whole location points in a trajectory. So, for guarding the trajectory is needed stay points [3] that are to be extracted from trajectory. Then check the sensitive stay points and apply anonymization in them, which are sufficient to prevent the privacy threats from the malevolent. This strategy is well suited because most of the vital information lies around the sensitive halting samples and protecting this information surely prevents the whole trajectory exposure. So, our proposal mainly aims to protect the revealing of entire trajectory information by generalizing the sensitive halting points along with the non-sensitive halting samples in a Minimum-Bounding-Rectangular (MBR) area zone. For the creation of area zone, the selection of non-sensitive halting points from the sensitive halting points were taken according to spatial distance calculated using the Haversine distance measure. It provides greater accuracy in its measure than Euclidean measure, while considering the spherical shape of the earth. Thus, the attackers would not be able to distinguish the real sensitive points. This work studies the function of personalized trajectory anonymization during the phase of data publication. The primary task of this work is extracting the sensitive as well as non-sensitive points among the user trajectories and anonymizing them in an MBR-zone and its size has to be defined by the user with a greater number of non-sensitive halting points for better privacy.

II. RELATED STUDY

Privacy preserved approach in trajectory anonymization is rather a new research area in which a number of work has been carried out in the recent years. Only few works has been done with semantic trajectories rather than raw trajectories. This is because the semantic trajectories contain the geographical details of the location that a user has visited, which is more important from the side of user perspective and some of the locations contain private information about the user and this has to be protected from the attackers before publishing. During the trajectory publishing phase, we have the approaches like generalization and cloaking methods, position sharing approaches, false path/location, perturbation and clustering, suppression and generalization etc. to keep away from adversaries from disclosing the sensitive data under the LBS (Location Based System). But only the generalization methods are discussed here. With the use of cloaking and generalization method, the authors in [4] proposed effective algorithms to anonymize moving object data for retaining the data utility with the help of apriori principle. The work adopted by k^m-anonymity model that can restrict the possibility of the disclosure of identity at the time data publishing. But this work uses Euclidean distance in their approach, not suited to the distance-based generalization in spatio-temporal trajectories. At first, in [2] was specified the concept of *k*-anonymity for the protection of shared clinical data from malevolent threat so that it cannot be separated from other *k - 1* record. Further this has been extended with *l*-diversity [5], which contains *l* different places. Next the authors in [6] went on to develop another privacy idea called *t*-closeness, it is an improved version of *l*-diversity. The authors in [7] tried to protect data privacy through their approach and it is done for the protection of sensitive attributes while providing *k*-anonymity. By the use of suppression and generalization method, some items are being deleted from the ordered sequences of places of interest visited by a particular user by taking the adversaries background knowledge about the user. By adding generalization which considers tree taxonomy to this concept, replaces the lower nodes with the parent node. Use of this may result in the distortion of data, if the full tree is not supplied. This method can be applied in two ways.

(i) Global Suppression: Here same type of transformation is applied to all trajectories in the dataset.

(ii) Local suppression: Here the transformation is taking place only for the items of specific trajectories in the data-set. Among these two suppression methods, local suppression incurs less information loss than the global suppression.

III. PROBLEM STATEMENTS

The spatio-temporal points left by the moving objects are called trajectories and they are accumulated in the moving object databases. Suppose a user trajectory UTrj, represented as UTrj={Ut_{id}, ($lati_1$, $longi_1$, t_1), ……., ($lati_n$, $longi_n$, t_n), where t_{id} be the trajectory-id; ($lati_i$, $longi_i$, t_i) be the i^{th} location's spatio-temporal coordinate at a sampling time t_i.

A. Definition 1 (Location)

A location Lc, consists of a pair of points (lati, longi), where lati and longi being the latitude and longitude of the location. Each point in the moving object database points to a location at the time of sampling. These location points may be either halting points or pass-by points.

The halting points are the points where the moving object is halted for a fixed time and the pass-by points being the location points that the moving object has moved without stopping.

B. Definition 2 (Halting points)

The point of halt Sp comprises of coordinates (Sp_{id}, Sp_t, lati, longi, γt), in which Sp_{id} being the halting point identifier; Sp_t is the halting point type, (lati, longi) are halting point's location coordinates and γtbeing the halting time duration. The halting points may be sensitive or non-sensitive in nature.

Sensitive halting point: It is the halting point that the user has halted for more than the threshold time Δt. The user is reluctant to disclose the information about this sensitive point because of privacy fear factor.

Non-sensitive halting point: It is the point where the user has halted for a marginal time (γt) which is always less than a time threshold Δt. Non-sensitive points can be disclosed to others without privacy fear.

C. Definition 3 (Area Zone)

An area zone Z_A consists of minimum one sensitive halting point and "n" non-sensitive halting points. The Z_A denoted as ($Z_{Aid,}$ Ul_c, Br_c, SP_n, NSP_n), where Z_{Aid} be the identifier of area zone, Ul_c and Br_c represent the upper left corner and bottom right corner coordinates of the Minimum-Bounding-Rectangle and SP_n and NSP_n represents the count of sensitive and non-sensitive halting points contained in the area zone.

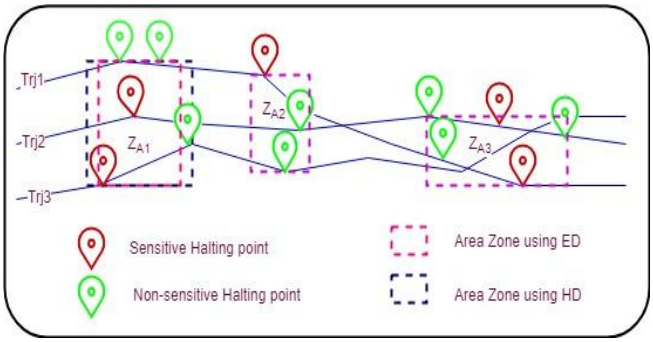


Fig. 2: Area zone using Euclidean and Haversine distance

An example of the trajectory anonymization and area zone creation is as shown in Fig. 2. Here we can see that the number of non-sensitive halting points were more in area zones with Haversine based distance strategy than the Euclidean based strategy. Haversine distance measure gives more accurate distance than the Euclidean measure since it considers sphericalshape of the earth for the calculation.

IV. METHODOLOGY

A. Outline of the Work

Anonymization of real trajectories of a user contained in the user database (Trj_Db) and publishing the anonymized database for the aid of researchers (PuTDb) is the primary goal of this work. The procedure is shown in Fig. 3: and described as follows.

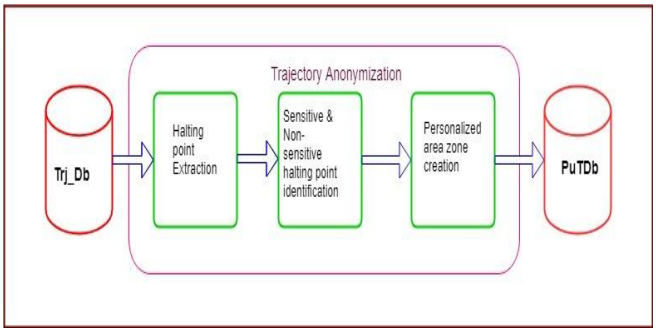


Fig. 3: Procedure of proposed anonymization method

(i) *Halting point extraction*. First, we extract the halting points from trajectories that are stored in Trj_Db, which is the location were the user has stopped for a fixed time interval.

(ii) *Sensitive & non-sensitive halting point identification*. In this step, we extracted the locations from halting points that the individual has stayed in a location, greater than the threshold time is treated as sensitive and other halting points as non-sensitive.

(iii) *Creation of personalized area zone*. Here, we created a Minimum Bounding Rectangle in the user trajectories, which contains minimum one sensitive halting point and a specified count of non-sensitive halting points depending on the Haversine distance measure from the sensitive to non-sensitive halting point.

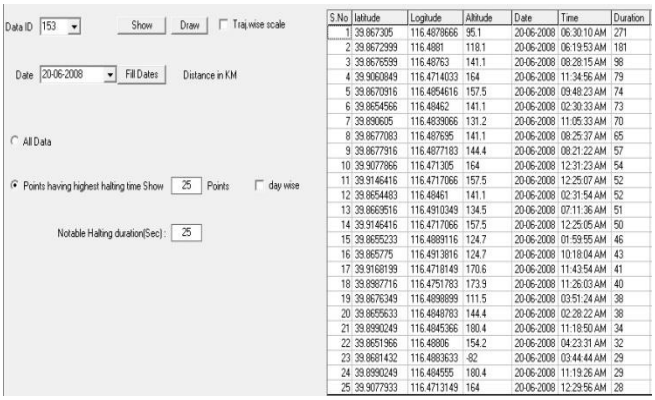
And finally is published the trajectory details with the ordinary location points and the coordinates of the anonymized area zone.

B. Halting points extraction

Each trajectory contains various halts and moves. These halts occurred while the user has halted in a location for their purposes. Many works have treated these halts as some important points for their work. Moves are the path between two consecutive halts. We followed the halting point's extraction strategy adhered in the paper [8] with slight modifications. The extraction of halting points was calculated by taking the time differences between a location point's time and the consecutive location's time.

For a given trajectory UTrj = {(L_{c1}, t_1), ….., (L_{cn}, t_n)}, where L_{ci} being the point's location coordinate and t_i is the sampling time, if $|t_{i+1} - t_i| > \gamma t$, in which γt is the user fixed halting time threshold, then L_{ci} is regarded as a halting point.

We store these halting point details in another table with halting location coordinates and halting time duration. The Fig.4: shows the output of the program that specifies the extraction of various halting point coordinates of a user from a particular day's trajectory and their halted duration in a sorted order.



S.No	latitude	Logitude	Altitude	Date	Time	Duration
1	39.867305	116.4878666	95.1	20-06-2008	06:30:10 AM	271
2	39.8672999	116.4881	118.1	20-06-2008	06:19:53 AM	181
3	39.8676599	116.48763	141.1	20-06-2008	08:28:15 AM	98
4	39.9060849	116.4714033	164	20-06-2008	11:34:56 AM	79
5	39.8670916	116.4854616	157.5	20-06-2008	09:48:23 AM	74
6	39.8654566	116.48462	141.1	20-06-2008	02:30:33 AM	73
7	39.890605	116.4839066	131.2	20-06-2008	11:05:33 AM	70
8	39.8677083	116.487695	141.1	20-06-2008	08:25:37 AM	65
9	39.8677916	116.4877183	144.4	20-06-2008	08:21:22 AM	57
10	39.9077866	116.471305	164	20-06-2008	12:31:23 AM	54
11	39.9146416	116.4717066	157.5	20-06-2008	12:25:07 AM	52
12	39.8654483	116.48461	141.1	20-06-2008	02:31:54 AM	52
13	39.8669516	116.4910349	134.5	20-06-2008	07:11:36 AM	51
14	39.9146416	116.4717066	157.5	20-06-2008	12:25:05 AM	50
15	39.8655233	116.4889116	124.7	20-06-2008	01:59:55 AM	46
16	39.865775	116.4913816	124.7	20-06-2008	10:18:04 AM	43
17	39.9168199	116.4718149	170.6	20-06-2008	11:43:54 AM	41
18	39.8987716	116.4751783	173.9	20-06-2008	11:26:03 AM	40
19	39.8676349	116.4898899	111.5	20-06-2008	03:51:24 AM	38
20	39.8655633	116.4848783	144.4	20-06-2008	02:28:22 AM	38
21	39.8990249	116.4845366	180.4	20-06-2008	11:18:50 AM	34
22	39.8651966	116.48806	154.2	20-06-2008	04:23:31 AM	32
23	39.8681432	116.4883633	-82	20-06-2008	03:44:44 AM	29
24	39.8990249	116.494555	180.4	20-06-2008	11:19:26 AM	29
25	39.9077933	116.4713149	164	20-06-2008	12:29:56 AM	28

Fig. 4: Extracted Halting points for a specific day

C. Identification of sensitive halting points

The sensitive halting point is the location where the user has halted for more than or equal to the time threshold Δt. Mostly these sensitive halting points may be of the types like religious place, hospital or a place where the user has some personal needs to satisfy. For this purpose, we adopt the strategy as in [4] with slight improvements. For the extraction, we take the halting points from halting point table and find out the sensitive halting points by extracting the points which halted more or equal to the time threshold Δt. Then store these sensitive points' details in another table and delete the sensitive points from the halting table. The non-sensitive halting points are the halting points which are not sensitive points or the halting points less than the time threshold Δt, extracted from the halting table. The Algorithm:1 (Fig.5) describes the halting points extraction within the user trajectories which identifies and keeps these sensitive halting points to St_tab and non-sensitive halting points values to NSE_tab and NSH_tab based on Euclidean and Haversine distance measures respectively.

Algorithm: 1 - Extract_sens: Extraction and identification of sensitive and non-sensitive halting Points

Input: Unprocessed individual's trajectories from Trj_Db
Output: Tables of sensitive halting points and non-sensitive halting points

```
1   Read and pre-process the entire trajectories of a user from Trj_Db -->Ptrj.
2   Initialize Hlt_tab<-- {φ}
3   Input hlt_time_thrsh as γt
4   Initialize j=0
5   while (! eof (Ptrj)) )                    /* Halting point Extraction*/
6         for i=0 to n-1
7               δt = | t_{i+1} – t_i |
8   if (δt> = γt ) then Hpt_i -->Hlt_tab
9         calculate j=j+1
10  Input sen_ pt_ thrsh_time as Δt and initialize St_tab, NSE_tab and NSH_tab <-- {φ}
11  while (! eof (Hlt_tab)) )                 /* Sensitive and non-sensitive point identification */
12        Initialize rep_visit as cnt=0, i=0
13         for j=0 to n-1
14               if (Hpt_i = = Hpt_{i+1}) then cnt=cnt+1
15                     Remove Hpt_{i+1} from Hlt_tab
16         for k=0 to n-1
17               if ((δt(Hpt_k) >= Δt) and (cnt>1)) then store Spt_i , δt -->St_tab, i=i+1
18                     Remove Hpt_k from Hlt_tab
19   Initialize p =0, q=0      /*Storing of points based on Euclidean and Haversine distances*/
20        for i=0 to n-1
21               for j=0 to m-1
22                     while (!eof(St_tab && Hlt_tab)
23                           Calculate Edt_p(Hpt_i, Spt_j), p=p+1
24                                 Store Spt_i, Hpt_i, δt_i, Edt_p -->NSE_tab
25                           Calculate Hdt_q(Hpt_i, Spt_j), q=q+1
26                                 Store Spt_i, Hpt_i, δt_i, Hdt_q -->NSH_tab
27  return.
```

Fig. 5: Algorithm: 1 - Extract_sens

The Table: 1 shows the count of sensitive and non-sensitive halting points for a particular user's trajectory on a specific day. The data-set used here is the Microsoft Geolife data-set [10].

Table 1: No. of sensitive and non-sensitive halting points on each day's trajectory

Date	Sensitive halting points	Non-sensitive halting points
01-05-2008	2	22
30-05-2008	2	10
20-06-2008	2	68

Also, you can notice that the sensitive points are very few in a single day as compared to non-sensitive halting points. Here we took the threshold time(Δt) for sensitive halting point as 120 sec.

D. Creation of Area Zone

Privacy of the user relies on the non-exposure of the sensitive halting points. In order to protect these sensitive points, we applied here the generalization strategy on the personalized trajectory. For the purpose of anonymization we created area zones. Zone is a rectangular shaped area over the trajectory in the form of Minimum-Bounding-Rectangle, it contains a minimum of one sensitive halting point and a non-sensitive halting point as per the MBR area size.

The authors in [4] used Euclidean distance for the calculation of distance between the points. But here uses the *Haversine-distance* formula for the distance calculation from sensitive to non-sensitive halting points. This is because, Euclidean distance is suitable for the geometric distance in a 2D plane between the points, but we used here the spatial points. The points were not located on plane surface. Some of the points may lay on the curved surface of the earth. So *Haversine* distance gives more accurate measure than the Euclidean distance since it recognizes the spherical shape of the earth. It was noted that privacy increases as the size of the zone increases. But along with, it increases the information loss also.

$$ED = \sqrt{(ulat2 - ulat1)^2 + (ulong2 - ulong1)^2} \qquad (1)$$

In the formula (1), ED be the Euclidean distance among two points, *ulat* and *ulong* are the corresponding UTM (Universal Transverse Mercator) coordinate values for latitude and longitude respectively.

$$HD = 2R \arcsin\left(\sqrt{sin^2\left(\frac{\sigma 2-\sigma 1}{2}\right) + \cos(\sigma 1)\cos(\sigma 2)\, sin^2\left(\frac{\alpha 2-\alpha 1}{2}\right)}\right) \qquad (2)$$

In the formula (2), HD is the Haversine distance [11] between two points, R is the radius of the earth(6371000m), σ and α are latitude and longitude respectively.

The Table 2: shows the spatial distance, Euclidean (ED) and Haversine (HD) in metres(m) from each of the non-sensitive (NS) halting points to sensitive(S) halting point with its corresponding halting duration (Hlt_tm) at the specified location.

Table 2: Euclidean and Haversine distance measures from a sensitive to non-sensitive hating points

Date	Lat	Long	ED (m)	HD(m)	Type	Hlt_Tm
	39.867305	116.4878666			S	271
	39.8676599	116.48763	44.28814966	44.32922363	NS	98
	39.9060849	116.4714033	4528.548159	4535.146442	NS	79
	39.8670916	116.4854616	207.0461842	206.6234531	NS	74
20-06-2008	39.8654566	116.48462	345.238939	344.993755	NS	73
	39.890605	116.4839066	2608.160238	2612.78492	NS	70
	39.8677083	116.487695	47.1111335	47.17573316	NS	65
	39.8677916	116.4877183	55.4766915	55.56805182	NS	57
	39.9077866	116.471305	4710.947316	4717.926016	NS	54

The Algorithm: 2 (Fig.6:) describes the formation of area zone (Z_A) using MBR and area zone's Ul_c and Lr_c are given for publication instead of sensitive and non-sensitive halting points coordinates. Thus, the trajectory anonymization is achieved.

Algorithm: 2 – Trajectory anonymization and publication

Input: Ptrj, St_tab, NSE_Tab and NSH_tab

Output: Publishable form of Anonymized Trajectory to PuTDb

1 Call Extract_sens.
2 Sort NSE_tab on Spt_i, Edt_p and NSH_tab on Spt_i, Hdt_q
3 Create Z_A using MBR with Spt_i's from St_tab and non-senstive halting points from NSE_tab.
4 Create Z_A using MBR with Spt_i's from St_tab and non-senstive halting points from NSH_tab.
5 Enlarge Z_A with "n" number of non-senstive halting points for more privacy.
6 Store the coordinates of Ul_c and Lr_c of Z_A.
7 Publish the trajectory points with Ul_c and Lr_c of Z_A, instead of sensitive halting points.

Fig. 6: Algorithm: 2 - Trajectory anonymization and publication

The selection of the non-sensitive halting points in an MBR was on the Haversine distance-based strategy. The count of non-sensitive points included in the zone gives more privacy to the user, since location exposure probability is larger than 1/*l*, where *l* represents the total count of location points in the zone.

For finding the average information loss [7], the formula needed is (3)

$$Inf_L = (\sum_{i=1}^{n}\sum_{j=1}^{n}(1 - 1/area\_size\_of\_Zone(Z_{Ai}, t_j)) + \sum_{x=1}^{k} L_x)/(n \times m) \quad (3)$$

Where Inf_L denotes average information loss happenedat the specific zone of Z_{Ai} at time t_j when Z_{Ai} halted. L_x is the probable sensitive halting point eliminator. $n \times m$ being the count of total location samples.

V. EXPERIMENTS AND ANALYSIS

A. Experimental Environment and Dataset

The trials are mostly done with on trajectory data-set, which were available from Geolife project of Microsoft [10] and spatio-temporal points of 182 users covers a total of 17,621 trajectories during 5 years (from Apr. 2007 to Aug. 2012) were contained in it. These trajectories contain the spatio-temporal details recorded by the various GPS-equipment and GPS-phone with a time interval of 1 to 5 seconds duration and in a distance of 5 to 10 meters per point. Within the data-set 73 users were labeled their trajectories with transportation mode like bus and bike and also includes driving and walking parameters.

The tests were performed on a Windows 10 computer with 4 GB RAM and Intel i5-3337U CPU @ 1.80GHz processor.

B. Analysis

According to our algorithm, various halting points were extracted from individualized trajectories of a single user by processing them. Since original data-set is prepared with utmost accuracy and it contains plenty of trajectories for a single user and the user's halting points we got in seconds. For the purpose of analysis, we took three day's trajectory and given it halting point threshold γt as 5 seconds and sensitive halting point threshold Δt as 120 sec. We also counted the Haversine and Euclidean distance for the extraction and identification of halting and sensitive point's identification. Also, the remaining halting points apart from sensitive halting points were taken into account as non-sensitive halting points for this work. Later we created the rectangular area zone using MBR with multiple numbers of non-sensitive halting points against sensitive halting points.

C. Evaluations

On evaluation, four major analyses were considered. The first analysis as shown in Fig.7 is the time for processing during the area zone creation, and we noted that as the count of non-sensitive halting points increases, the area zone creation time also increases.

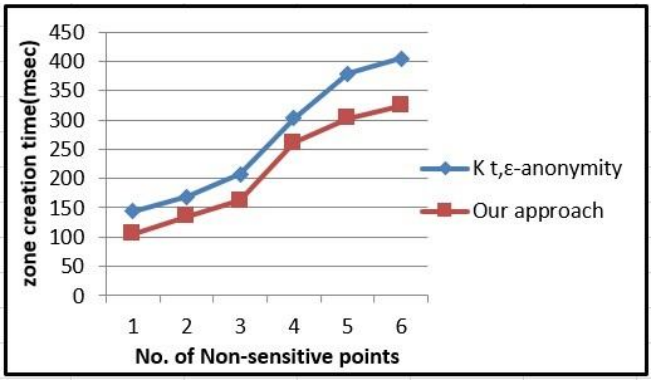


Fig. 7: Evaluation for zone creation time

But as compared our approach with the $k^{t,\epsilon}$-anonymity approach [9], we could clearly say that our approach has an edge over the existing approach.

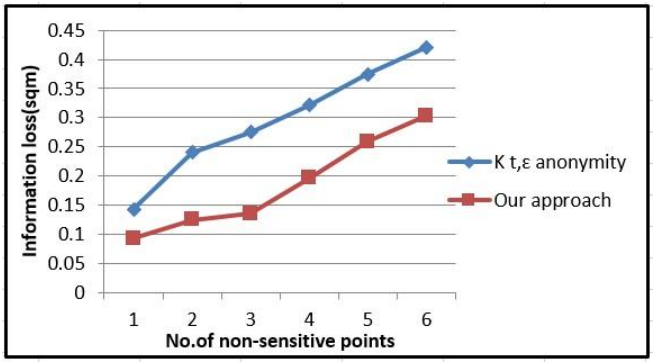


Fig. 8: Evaluation for information loss

In the second analysis as shown in Fig.8, we reckoned the information loss against the number of non-sensitive halting points. Here also we found that as the count of non-sensitive halting point's increases, the information loss also increases. As the count of non-sensitive points increases within the area zone, the malevolent is very hard to identify the sensitive traits of an individual, which results in the increase of privacy. Also, we can find that the information loss is less in our approach compared to the existing approach [9].

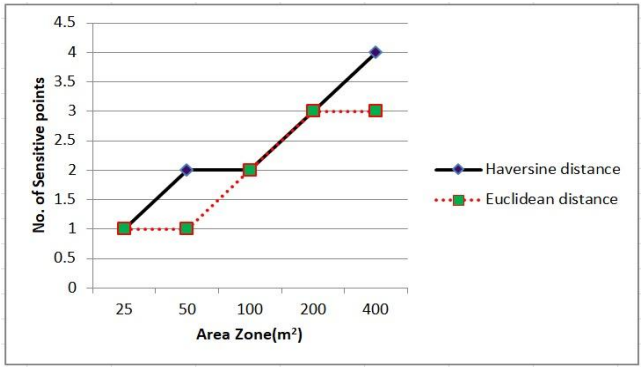


Fig. 9: Evaluation for sensitive points within the area zone

The third analysis as shown in Fig.9: deals with the inclusion of sensitive halting points within the area zone using Euclidean and Haversine distance measures. The result indicates that as the size of the area zone increases, the inclusion of the number of sensitive halting points in both the distance measures may vary according to its current position in GPS system. However Haversine distance measure considers the spherical shape of the earth, it gives better accuracy for the distance measures.

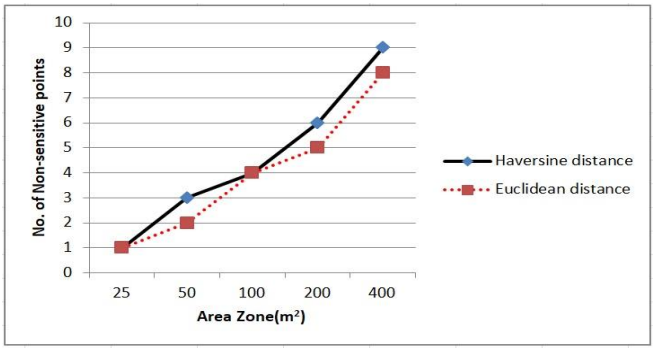


Fig. 10: Evaluation for non-sensitive points within the area zone

The analysis shown in Fig.10: gives the evaluation that, the count of non-sensitive halting points contained in the area zone increases as the size of the area zone increases and there is slight variation for the number of points from Euclidean to Haversine distance. The smaller number of halting points in the area zone results in less generalization of location points, and thereby information loss is less and privacy gain is high. When compared to the approach in [4, 9] and ours is better than the current models, since it mainly it uses Haversine distance measure for the work.

VI. CONCLUSION

Privacy preserved data publishing has become an important research area in the recent years. The thrust area of every user is the personal privacy, especially for the people who need most privacy in their day to day activities,and the modern society cannot do away with the amenities of LBS. Therefore, gathering and publishing the trajectories need extreme care. Almost all works have been carried out in the trajectory anonymization by using Euclidean distance measure for the distance-based calculations. Our preposition in this paper is that, the anonymization of personalized trajectories by concealing the sensitive halting points within the area zone which also contains a user specified number of non-sensitive halting points based on Haversine distance measure rather than the unnecessary anonymization of all halting points.The said approach is also proved that the Haversine distance measure is the more suitable distance measure in spatio-temporal trajectories, since it considers the spherical shape of the earth. The area zone accommodates the non-sensitive and sensitive halting points with a correct distance measure. The evaluation result and analysis also reveals that our approach is more suitable than other current approaches in terms of privacy gain and less information loss. So the suggested approach considerably reduced the privacy leakage fear from the mind of publishing people and they could give more sensible anonymized trajectory details for the research cum developmental activities. Further, we intend to extend the scope of our approach to defy multiple adversary threats like velocity-based linkage attacks.

REFERENCES

1. Kong, X., Li, M., Ma, K., Tian, K., Wang, M., Ning, Z., & Xia, F., Big Trajectory Data: A Survey of Applications and Services. *IEEE Access*, *6*, 2018, pp. 58295-58306.
2. Sweeney, L., Achieving *k*-anonymity privacy protection using generalization and suppression. International Journal of uncertainty, fuzziness and knowledge-based systems, 10(5), 2002, pp. 571-588.
3. Huo, Z., Meng, X., Hu, H., Huang, Y.: You can walk alone: Trajectory privacy preserving through significant stays protection. *DASFAA 2012*, Part 1, LNCS 7238, Springer-Verlag Berlin Heidelberg, 2012, pp. 351-366.
4. Poulis, G., Skiadopoulos, S., Loukides, G., Gkoulalas-Divanis, A., Apriori-based algorithms for k^m- anonymizing trajectory data. Transactions on data privacy, 7:2, 2014, pp. 165-194.
5. Machanavajjhala, A., Kifer, D., Gehrke, J., Venkitasubramaniam, M.: *l*-diversity: Privacy beyond *k*- anonymity. TKDD, 2007, 1(1).
6. Li, N., Li, T., Venkatasubramanian, S., *t*-closeness: Privacy beyond *k*-anonymity and *l*-diversity. In ICDE, 2007, pp. 106-115.
7. Xiao, X., Tao, Y.: Personalized Privacy Preservation. In *ACM SIGMOD*, 2006, pp. 229-240.

8. Zheng, Y., Zhang, L., Xie, X., Ma, W.: Mining interesting locations and travel sequences from GPS trajectories. In: 18th International conference on World Wide Web, ACM press, New York, 2009, pp.791-800.
9. Gramaglia, M., Fiore, M., Tarable, A., Banchs, A. : k $^{t,\epsilon}$ anonymity: Towards Privacy-Preserving Publishing of Spatiotemporal Trajectory Data, 2017, https://arxiv.org/abs/:1701.02243 [cs.CY].
10. Microsoft Research Geolife,Date-of-access:21/05/2018 Available: *http://research.microsoft.com/en-us/projects/geolife/*.
11. Haversine distance, Available:http://www.ryanduell.com/2012/12/determining-the-distance-between-two-geographic-points/

AUTHORS PROFILE

Rajesh N (MCA). He is a Part-time Research Scholar in the School of Computer Sciences, Mahatma Gandhi University, Kottayam, Kerala, India and also an Assistant Professor in the Department of Computer Science & Applications, S.A.S. S.N.D.P. Yogam College, Konni, Pathanamthitta, Kerala, India. He has 20 years of undergraduate and 10 years of post-graduate teaching experience. His main area of research interests are Privacy preservation and computing, trajectory data publishing and anonymization, Big data Analysis and Data Mining,Spatio-temporal mobility databases. He is also a member of academic bodies in the University and College. He has published over10 articles in International and National journals and Conference proceedings.

Sajimon Abraham. (MCA, MSc. (Mathematics), MBA, PhD (Computer Science). He has been working as Faculty Member in Computer Applications & IT, School of Management and Business Studies, Mahatma Gandhi University, Kottayam, Kerala, India. He currently holds the additional charge of Director (Hon), University Center for International Co-operation. He was previously working as Systems Analyst in Institute of Human Resource Development, Faculty member of Computer Applications in Marian College, Kuttikkanam and Database Architect in Royal University of Bhutan under Colombo Plan on deputation through Ministry of External Affairs, Govt. of India. His research area includes Data Science, Spatio-Temporal Databases, Mobility Mining, Sentiment Analysis, Big Data Analytics and E-learning and has published 70 articles in National, International Journals and Conference Proceedings.

Shyni S. Das, (M.C.A.) is an Assistant Professor in the Department of Computer Science & Applications, S.A.S. S.N.D.P. Yogam College, Konni, Pathanamthitta, Kerala, India. She has 20 years of under graduate and 10 years of post-graduate teaching experience. She is also a member of the academic bodies in University and College. She is having the experience of guiding the project works of the students from both under graduate and post-graduates. Her main area of research interests are Privacy computing, Internet of Things and Spatio-temporal data mining. She has published and presented over 5 articles in International and National Journals and Conference proceedings.