Reciprocal Bucketization (RB) - An Efficient Data Anonymization Model for Smart Hospital Data Publishing

Rajesh S M[1], Prabha R[2]
Research Scholar, Department of Information Science and Engineering,
Dr. Ambedkar Institute of Technology, Visvesvaraya Technological University
Department of Computer Science and Engineering,
GITAM Deemed to be University, Bengaluru, India.[1]
Department of Computer Science and Engineering, Dr. Ambedkar Institute of Technology, Bengaluru, India.[2]

***Abstract*—With the lightning growth of the Internet of Things (IoT), enormous applications have been developed to serve industries, the environment, society, etc. Smart Health care is one of the significant applications of the IoT, where intelligent environments enrich safety and ease of surveillance. The database of the Smart Hospital records the patient's sensitive information, which could face various potential privacy breaches through linkage attacks. Publishing such sensitive data to society is challenging in adopting the best privacy preservation model to defend against linkage attacks. In his paper, we propose a novel Reciprocal Bucketization Anonymization model as the privacy preservation method to defend against Identity, Attribute, and Correlated Linkage attacks. The proposed anonymization method creates the Buckets of patient records and then partitions the data into sensor trajectory and Multiple Sensitive attributes (MSA). A local suppression is employed on Sensor Trajectory Data and Slicing on MSA to get the anonymized data to be published gathered by combining anonymized sensor trajectory and MSA. The proposed method is validated on the synthetic and real-time dataset by comparing its data utility loss in both sensor trajectory and the MSA. The experimental results eradicate that the RB – Anonymization exhibits the nature of best privacy preservation against Identity, Attribute, and Correlated linkages attacks with negligible utility loss compared with the existing methods.**

***Keywords*—Anatomization; anonymization; entropy; pearson's contingency coefficient; and KL – Divergence**

I. Introduction

The Internet of Things (IoT) ecosystem facilitates collaboration across various computing devices ranging from sensors to complex processing systems with cloud storage. In the digitally connected world era, IoT adoption has rapidly increased in multiple applications, such as Smart Homes, Smart Healthcare, and so on [1]. The IoT plays a crucial role in building a secure cyber-physical communication system as the essential requirement for deploying Intelligent applications. Smart Healthcare systems primarily focus on patient real-time monitoring through sensors and data management via the cloud for remote access, such as Mobi-Care and MEDiSN [2].

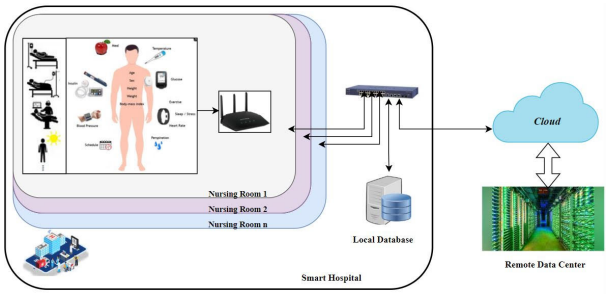


Fig. 1. Structure of smart hospital.

Fig. 1 describes the components deployed in the Smart Hospital Structure with its various benefits to assess the patients and the caretakers to build a pervasive surveillance system. The evolution in the sensor's technology assures innovative and secure patient care in real-time and facilitates the accessible collection of the patient's spatiotemporal sensor data [3]. The sequence of sensor data at the specific time of a patient is known as sensor trajectory data. It helps to predict patient-sensitive information such as symptoms, diseases, etc. However, publishing such trajectory data along with multiple sensitive attributes of a patient for the researchers/data miners may result in a privacy breach [4]. Hence the challenge is preserving the privacy of the patient data by providing equal weightage for trajectory data and multiple sensitive attributes against the potential privacy breach.

Most privacy preservation principles are designed to protect the data against privacy breaches of patient sensitive information. However, the Adversary with prior knowledge of partial sensor trajectory data or a few sensitive attributes could infer the patient data even after removing the identity attributes, such as patient name and unique social identification number from the database [5]. Further, the Adversary can imply various linkage attacks with the prior knowledge to predict the patient's sensitive information with high probability [6].

Example: Consider the SMART Hospital "X," which digitally records and maintains patient data. The records may have the patient's ID, sensor trajectory, and medical data, as represented in Table I. The sensor trajectory data is recorded concerning time from the sensors deployed on the patient's

body, representing the pair of sensor data and time given as (sen,t) [7] [8]. An example of the Patient data is as follows, PId 7 is the patient ID, the data collected from the sensors x, k, n, and p at the timestamps 1, 6, 7, and 8, respectively, with the multiple sensitive attributes Fever, RITD Test, Influenza, and Medicine. The recorded data must be made available for the data miners for research [9]. In parallel, the hospital wants to preserve the privacy of the patient's sensitive data from unauthorized usage by malicious data miners against the following attacks [10]:

Identity linkage attack: In the published dataset, if the trajectory of the sensor data for a patient is unique, then an adversary can quickly identify a patient record along with the patient's sensitive data using his prior knowledge [11].

Attribute linkage attack: In the published dataset, the most frequent occurrence of the sensitive values of a targeted victim could result in an attribute linking attack. The adversary could breach the sensitive information with high confidence even though the unique sensor trajectory information of the victim is not available [11].

Correlated-records linkage attack: In the published dataset, when a patient has multiple records, there could be the possibility of a Correlated-record linkage attack. For example, the patient with PId 1 has the correlated records in row 1 and row 4, as shown in Table I. Having additional knowledge about the correlated records by the adversary can predict both trajectory data and the sensitive value of the victim with high confidence [11].

In literature, various privacy preservation approaches have been proposed for the trajectory data with single sensitive attributes, such as Generalization, Perturbation, Clustering, Differential Privacy, and Suppression, to defend against the various linkage attacks [12] [13]. Similarly, Multi-sensitive Bucketization, (p,k) —Angelization, and Generalization are the approaches to preserve the privacy of the multiple sensitive attributes along with the Quasi Identifiers but not with trajectory data. To the best of our knowledge, we are the first to address the privacy preservation approach for the dynamic trajectory data generated by the sensors and Multiple Sensitive attributes with trustfulness.

In this paper, we implement Reciprocal Bucketization as the overall framework for anonymization. Further Suppression and Slicing are implemented to anonymize the trajectory data with Multiple Sensitive Attributes to ensure privacy from the above three linkage attacks. The Bucketization approach helps in the formation of buckets from the patient data table records, on which the suppression and the slicing methods are parallelly imposed on sensor trajectory data and MSA, respectively, to anonymize the data via K-anonymity threshold by reducing the data loss with efficient anonymization [14] [15].

The major contributions are summarized as follows:

- We present Reciprocal Bucketization (RB) an efficient data anonymization model to preserve privacy in publishing the patient's data by the Smart Hospitals.
- To the best of our knowledge, we are the first to combine the Sensor trajectory data and MSA to ensure the privacy requirements for data publishing to defend against Identity, Attribute, and Correlated Linkage attacks.
- We proposed a suppression method on the sensor trajectory data and slicing on the MSA to achieve an improved anonymization model with a reduced information loss rate compared with earlier approaches.

The rest of the paper organizes as follows: Section II introduces the related works with their benefits, and Section III defines the basic definitions and notations incorporated. Section IV describes the procedure involved in Reciprocal Bucketization as the efficient anonymization model. The experimental results and comparative analysis with the existing approach are given in Section V. Finally, Section VI concludes the proposed approach.

II. Literature Review

In this section presents the advantages of Smart Health Care, the recent research on the privacy preservation of trajectory data, and the multiple sensitive attributes addressing the various potential benefits and shortcomings.

Smart Health Care is a significant application of the IoT, where the various aspects of Health Care are implemented for ease of maintenance under the secured surveillance system. Vedaei *et al.* [16] presented COVID-SAFE, an automated health monitoring and surveillance system integrating IoT devices with Machine Learning algorithms. The primary aim is to improve the efficiency and accuracy of detecting and monitoring COVID-19 patients in urban areas. The depreciation of exposure to the coronavirus is more significant. Still, deployment and maintenance costs are high, and there needs to be a consideration for the security aspects of the data collected in the IoT environment.

IoT devices are highly vulnerable to attacks, and the medical data collected from those devices are highly significant and need highly secured privacy schemes and policies. Luo *et al.* [17] designed a Privacy Protector framework to defend against linkage attacks and secretly share data by adopting the Slepian-Wolf-coding-based secret sharing (SW-SSS). The distributed nature of the framework stores the patient data collected on the cloud server by assuring an efficient access control scheme. Privacy Protector ensures the security of the data collected, however securing the data in real-time is still challenging.

Komishani *et al.* [18] presents a Preserving personalized privacy in trajectory data publishing (PPTD) model for the trajectory data associated with the sensitive attribute of moving objects. The sensitive attribute generalization and trajectory data local suppression approach balance the data utility and privacy well. The linkages attacks such as identity linkage, attribute linkage, and similarity attacks are demonstrated on the anonymized data to the resistance of the data publishing. The PPTD has been implemented on the City80K and Metro100K datasets, and an extensive comparison is carried out with KCL. With less data loss and high privacy protection, the PPTD outstands in its performance aspects, but multiple sensitive attributes are yet to address with efficient data utility.

Addressing the various linkage attacks, such as attribute, record linkage, and similarity attacks on the trajectory data,

TABLE I. SMART HOSPITAL SENSOR TRAJECTORY DATASET WITH MULTIPLE SENSITIVE ATTRIBUTES (P_{TB}).

PId	Trajectory	Multiple Sensitive Attributes			
		Symptom	Diagnostic Method	Disease	Treatment
1	$x1 \rightarrow d2 \rightarrow z3 \rightarrow p4 \rightarrow k6 \rightarrow p8$	Abdominal pain	X-ray	Abdominal Cancer	Chemotherapy
2	$k6 \rightarrow n7 \rightarrow p8$	Weight loss	Antibody Test	HIV	Medication
3	$z3 \rightarrow k6 \rightarrow n7 \rightarrow p9$	Eating disorders	Body mass index (BMI)	Obesity	Nutrition control
1	$x1 \rightarrow d2 \rightarrow n5 \rightarrow k6 \rightarrow p9$	Fever	Molecular diagnostic methods	Cholera	Antibiotic
4	$x1 \rightarrow d2 \rightarrow k6 \rightarrow n7 \rightarrow p9$	Infection	ELISA Test	HIV	ART
5	$d2 \rightarrow n5 \rightarrow k6 \rightarrow n7$	Diarrhea	RT-PCR Tests	Dengue	Antibiotic
6	$x1 \rightarrow z3 \rightarrow n7 \rightarrow p8$	Shortness of breath	FeNO test	Asthma	Medication
7	$x1 \rightarrow k6 \rightarrow n7 \rightarrow p8$	Fever	RITD Tests	Influenza	Medicine
8	$n5 \rightarrow k6 \rightarrow p9$	Weight loss	MRI Scan	Lung Cancer	Radiation Therapy
9	$d2 \rightarrow n5 \rightarrow n7 \rightarrow p9$	Chest tightness	Methacholine challenge tests	Inflammation	Medication
10	$p4 \rightarrow n7 \rightarrow p8$	Pain or discomfort	Biopsy Test	Skin cancer	Radiation Therapy
11	$z3 \rightarrow p4 \rightarrow k6 \rightarrow p8$	Abdominal pain	Ultrasound	Dyspepsia	Antibiotic

Yao *et al.* [19] have designed an anonymous technique called Enhanced l-diversity Data Privacy Preservation for publishing trajectory data (EDPP). EDPP defends against the background knowledge of the trajectory data to predict the sensitive attributes by identifying critical spatial-temporal sequences that cause privacy leakage. The method adopts perturbation and enhanced l - diversity with well-defined privacy constraints to ensure more excellent data utility. However, the approach must be extended for a greater trajectory length with indexing and multiple sensitive attributes.

Adding Noise to the trajectory data to ensure privacy through a vector-based grid environment is a new effort by Tojiboev *et al.* [20] Adding Noise before data publishing results in low complexity and greater privacy, but data handling and rebuilding the original data is challenging. Differential privacy is a modern, robust privacy preservation approach that implements a query mechanism to minimize privacy loss. Added Noise on the trajectory data and implementing differential privacy as the privacy protection model poses a challenge in data utility. They are considering the sensitive attribute label and adopting Generative Adversarial Network (GAN), an effective privacy preservation model implemented by Yao [21] The GAN ensures balanced data privacy and data utility. However, the computational speed and addressing the multi-source trajectory data and sensitive attributes are challenges for Differential Privacy.

Wen *et al.* [22] proposed a dynamic privacy level model by defining the relationship between privacy requirements and location features. The optimal differential privacy model is designed on the trajectory data by filtering the template trajectory with the semantic similarities on the trajectory as the constraint. The model publishes the data on randomization of the locations on the user trajectory data to balance privacy and data utility. However, adopting differential privacy on the multiple sensitive attributes and the dynamic trajectories is challenging.

Kanwal *et al.* [23] present (p, l)- Angelization for publishing 1:M, an individual with multiple records resisting a correlation attack. The Angelization method eliminates explicit Identifier by splitting the table into quasi-identifiers and multiple sensitive attributes. Quasi-attribute generalization and multiple sensitive attribute weight and dependency are computed for anonymization through (p, l)- Angelization. The algorithm performs well under static datasets without republications, but republication is an issue with data utility for the dynamic dataset.

The privacy preservation methods like generalization and bucketization pose a challenge to data utility. To overcome these issues slicing method is proposed by Li *et al.* [24] The slicing method can be applicable horizontally and vertically on the given data records; the membership disclosure attack is primarily defended. The significant advantage of slicing is to handle the high dimensional data with minimized data loss on the complete records. Overlapping slicing ensures high privacy by duplicating an attribute into more than one column. However, the utilization of the anonymized data still needs to be improved.

Heap Bucketization-anonymity (HBA) model is proposed by *et al.* [25], where the method develops an anonymization approach for quasi-identifiers and the sensitive attribute. HBA anatomizes the complete records to anonymize the sensitive attributes using slicing and Heap Bucketization of the quasi-identifier using k-anonymity and slicing. The KL - Divergence is used for validation in terms of utility and privacy by defending against background knowledge attacks, quasi-identifier attacks, membership attacks, and fingerprint correlation attacks. HBA results in less utility loss with greater privacy; HBA has yet to address the dynamic, unstructured data and semi-sensitive attributes.

Sensitive Label Privacy Preservation with Anatomization (SLPPA), a scheme for privacy preservation, is designed by Yao *et al.* [26] to address the various background knowledge attacks. The SLPPA adopts two phases in implementation, i.e., Table Division and Group Division. The entropy and mean-square contingency coefficient is computed for anonymizing by adding uncertainty during table division. The group division is performed by adopting the privacy constraints and ensuring no overlapping groups in the published data. SLPPA enhances data utility by defending against background knowledge attacks. However, dynamic data anonymization is yet challenging.

III. PROBLEM DESCRIPTION AND BASIC NOTATIONS

A. Problem Definition

The patient's dataset (P_{TB}), combines trajectory data (P_{TB}^{T}) and multiple sensitive attributes (P_{TB}^{SA}). Anonymize the dataset so that the Adversary with the prior knowledge fails to decode the individual identity through the trajectory data or the MSA. The anonymization approach must ensure the defense

mechanism against linkages attacks with optimal equilibrium between privacy and data utility.

B. Notations

A patient's dataset consists of patient data records, where each record allows a unique patient identifier, sensory trajectory data, and a set of sensitive values. The patient sensor trajectory data with the MSA can be represented as:

$$P_{TB} = \{(PId_1, T_1, SA_1), ...(PId_i, T_i, SA_i)\} \quad (1)$$

Where, PId is unique identifier for a patient in P_{TB}. T_i is an Sensor Trajectory data for a patient possessing the MSA as SA_i. T_i is a sequence of data collected for the sensor's at a particular timestamp for a user i and it given as follows:

$$T_i = \{(sen_1, t_1)^i, (sen_2, t_2)^i, ...(sen_n, t_n)^i\} \quad (2)$$

We compute $|T_i|$ as the number of sensors moving points for a patient 'i'. For an Example $|T_4|$ = 5, i.e., five sensors are recording the data in sequence for patient with ID 4, concerning Table I.

Joinable trajectories are formed if there exists a sub-trajectory for the given trajectory. Let's consider $T_K = \{t_1^K, t_2^K, t_3^K, ..., t_n^K\}$ as the sensor trajectory data and $T_S = \{t_1^s, t_2^s, t_3^s, \ldots, t_m^s\}$ as the sub trajectory of user K then $|T_S|$is the sub trajectory satisfying the condition n¡=m and $|T_S|$ must be subset of $|T_K|$. The trajectories can be merged using union operation.

C. Adversary Model

The Adversary could breach the patient's privacy in two significant ways (1) the Adversary's Prior Knowledge of Trajectory Data and (2) the Adversary's goal to hit the victim by knowing a sensitive values from MSA. Let's consider that Avok is one of the patients in the smart hospital, and his details are recorded in Table I. Consider 'A' as the Adversary, which could be a data collector itself and aims to find the sensitive attributes of the targeted victim. With the prior knowledge about Avok, the adversary 'A' can perform the following attacks[27]:

Identity linkage attack: If adversary 'A' knows about Avok's sensor trajectory, i.e., sensor 'd' measures the oxygen level at timestamp 2 and sensor 'p' measure the blood pressure at timestamp 4, respectively, then A can claim that the record T1 belongs to Avok. The adversary can declare with 100% confidence that the record belongs to Avok and access the sensitive attributes because d2, p4 is the only sub-trajectory that belongs to the T1 record.

Attribute linkage attack: If adversary 'A' knows about Avok's partial sensor trajectory, i.e., sensor 'k' measures the heart rate at timestamp 6 and sensor 'n' measures the body temperature at timestamp 7, respectively. The table contains three records, T2, T4, and T5, with the sub-trajectory k6, n7 can be identified by A. Adversary 'A' can forecast that Avok has HIV disease with 67% confidence because out of three records under multiple sensitive attributes, two records have the disease as HIV.

Correlated-records linkage attack: From the given Table, Avok has multiple records. If adversary 'A' knows the number of records Avok has in the dataset and also knowledge about Avok's sensor trajectory x1, k6. This makes the adversary predict with 100% confidence with his prior correlated knowledge that Avok has Asthma plus dengue could be specifically on the Sensor trajectory data. Similarly, if the Adversary 'A' knows the two sensitive attributes out of four, like the Symptom: Abdominal pain and Treatment: Chemotherapy, the adversary could correlate them to say confidently that the patient has Abdominal Cancer.

D. Privacy Requirement

The goal of anonymizing the patient dataset from adversaries requires adopting the following privacy requirements in our proposed approach [28].

1) Bucketization: The patient's dataset (P_{TB}) , which is a combination of sensor trajectory data (P_{TB}^T) and multiple sensitive attributes (P_{TB}^{SA}), is partitioned into 'n' buckets with a constraint of a maximum bucket size of three patient records. On partitioned, each bucket is named with Bucket ID. Further bucketization helps in carrying the suppression and slicing parallelly.

2) Suppression: Suppression is the method of eliminating the critical sensor trajectory point from the patient trajectory data. A sensor trajectory point is critical if and only if the trajectory point fails to satisfy the K-1 threshold defined and the suppression metric. The critical point is eliminated only with the corresponding trajectory to enhance the data utility by minimizing the data loss, which could express as Local Suppression [29].

3) Slicing: It's a method to anonymize the data using a partition. Partition could be carried vertically or horizontally to get the randomly permutated sliced table. The attribute partition is carried out by slicing Table III vertically into two slices, say (Disease, Symptom) and (Diagnostic Method, Treatment) [30].

E. Utility Metrics

To maintain the trade-off between the patient's privacy and data utility in the published anonymized dataset. It's required to define the Reciprocal Bucketization to ensure less data loss through cumulated sensitive attribute representation under the same bucket. The suppression metric measures the suppression score of every sensor trajectory point and helps find the critical trajectory point to remove. The computation of the suppression score of one sensor point from the critical trajectory is as follows.

$$SupressionScore = \frac{NT\_B_i\_CP}{NT\_B_i} \quad (3)$$

Where,NT_B_i_CP : Number of Trajectories in Bucket with the Critical Point and NT_B_i : Number of Trajectories in the bucket

On the computation of the suppression score for both the sensor points from the trajectory, one point is selected as critical by finding the maximum among them. If both the points have the same suppression score, the leftmost sensor trajectory point is declared the critical point to remove [31].

IV. Proposed System

The reciprocal Bucketization anonymity model proposes the design workflow as shown in the Fig. 2 and algorithm to prevent the privacy leakage of patients' data from the three linkage attacks. The proposed model consists of the following four phases,

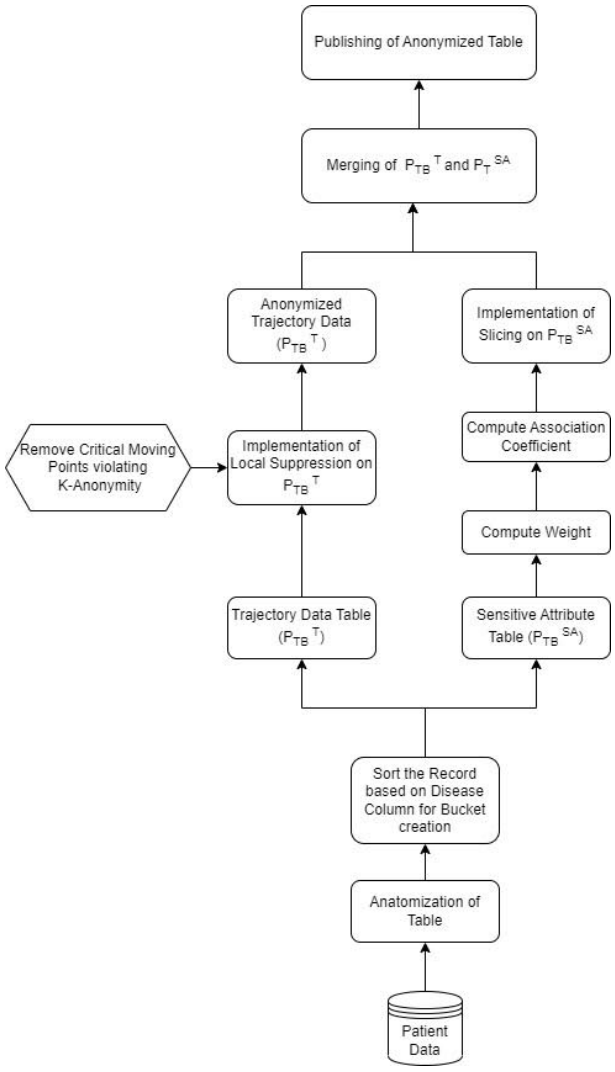


Fig. 2. Workflow of reciprocal bucketization model.

1) Bucketization and Partition.
2) Local Suppression on Sensor Trajectory Data (P_{TB}^{T}).
3) Slicing of MSA (P_{TB}^{SA}).
4) Publishing Anonymized Dataset on Merging P_{TB}^{T} and P_{TB}^{SA}

The primary goal of the proposed model is to ensure less data loss with high privacy preservation. The detailed procedure for the above four phases is in the following sub-sections.

Algorithm 1 Reciprocal Bucketization

INPUT: Patient Data P_{TB}, Bucket_Size
OUTPUT: Anonymised Patient Data P_{TB}^{A}.

anatomize (Patient Data P_{TB})
Sort_Disease_Value(Patient Data P_{TB})
$Bucket\_ID = 0, count = 0, i = 1$
for $k \leftarrow 1$ to $len(P_{TB})$ **do**
  if $count \leq Bucket\_Size$ **then**
    $Bucket\_ID = i$
    $count = count + 1$
  else
    $count = 0$
    $i = i + 1$
  end if
end for
$SplitP_{TB}$ =$P_{TB}^{T} withBucket\_ID$,$P_{TB}^{SA} withBucket\_ID$
P_{TB}^{TA} = Anonymise_Trajectory(P_{TB}^{T})
P_{TB}^{SAA} = Anonymise_Sensitive_Attribute(P_{TB}^{SA})
P_{TB}^{A} = merge(P_{TB}^{TA},P_{TB}^{SAA})
return P_{TB}^{A}

A. Bucketization and Partition

The primary goal of this step is to create the Bucket based on the user input bucket size, then partition Table II into two table's Table III and Table V, having sensor trajectory and MSA, respectively. Before the bucket creation, the entire Table I is sorted according to the Disease column from the MSA to get Table II and a new column, Bucket ID, where each patient's record is added with the bucket number sequentially corresponding to the bucket size. I.e., if the user input bucket size is 3, then the first three records from Table II are allocated with bucket number 1, and so on for the rest of the buckets. Further, Table II is divided into two Tables to carry Suppression on the Trajectory data and Slicing on MSA to achieve the anonymization effectively.

B. Local Suppression on Sensor Trajectory Data (P_{TB}^{T})

On the sensory trajectory data P_{TB}^{T}, the Local Suppression method is implemented to remove the critical trajectory points from the patient's PTB dataset and generate the anonymized trajectory dataset. The algorithms depict the steps involved in the trajectory suppression, and the procedure is as follows:

TABLE III. Sensor Trajectory Data (P_{TB}^{T})

PId	Trajectory	Bucket ID
1	$x1 \rightarrow d2 \rightarrow z3 \rightarrow p4 \rightarrow k6 \rightarrow p8$	1
6	$x1 \rightarrow z3 \rightarrow n7 \rightarrow p8$	1
1	$x1 \rightarrow d2 \rightarrow n5 \rightarrow k6 \rightarrow p9$	1
5	$d2 \rightarrow n5 \rightarrow k6 \rightarrow n7$	2
11	$z3 \rightarrow p4 \rightarrow k6 \rightarrow p8$	2
2	$k6 \rightarrow n7 \rightarrow p8$	2
4	$x1 \rightarrow d2 \rightarrow k6 \rightarrow n7 \rightarrow p9$	3
9	$d2 \rightarrow n5 \rightarrow n7 \rightarrow p9$	3
7	$x1 \rightarrow k6 \rightarrow n7 \rightarrow p8$	3
8	$n5 \rightarrow k6 \rightarrow p9$	4
3	$z3 \rightarrow k6 \rightarrow n7 \rightarrow p9$	4
10	$p4 \rightarrow n7 \rightarrow p8$	4

TABLE II. SORTING DISEASE COLUMN AND BUCKET ID ASSIGNMENT

PId	Trajectory	Multiple Sensitive Attributes				
		Symptom	Diagnostic Method	Disease	Treatment	Bucket ID
1	$x1 \rightarrow d2 \rightarrow z3 \rightarrow p4 \rightarrow k6 \rightarrow p8$	Abdominal pain	X-ray	Abdominal Cancer	Chemotherapy	1
6	$x1 \rightarrow z3 \rightarrow n7 \rightarrow p8$	Shortness of breath	FeNO test	Asthma	Medication	1
1	$x1 \rightarrow d2 \rightarrow n5 \rightarrow k6 \rightarrow p9$	Fever	Molecular diagnostic methods	Cholera	Antibiotic	1
5	$d2 \rightarrow n5 \rightarrow k6 \rightarrow n7$	Diarrhea	RT-PCR Tests	Dengue	Antibiotic	2
11	$z3 \rightarrow p4 \rightarrow k6 \rightarrow p8$	Abdominal pain	Ultrasound	Dyspepsia	Antibiotic	2
2	$k6 \rightarrow n7 \rightarrow p8$	Weight loss	Antibody Test	HIV	Medication	2
4	$x1 \rightarrow d2 \rightarrow k6 \rightarrow n7 \rightarrow p9$	Infection	ELISA Test	HIV	ART	3
9	$d2 \rightarrow n5 \rightarrow n7 \rightarrow p9$	Chest tightness	Methacholine challenge tests	Inflammation	Medication	3
7	$x1 \rightarrow k6 \rightarrow n7 \rightarrow p8$	Fever	RITD Tests	Influenza	Medicine	3
8	$n5 \rightarrow k6 \rightarrow p9$	Weight loss	MRI Scan	Lung Cancer	Radiation Therapy	4
3	$z3 \rightarrow k6 \rightarrow n7 \rightarrow p9$	Eating disorders	Body mass index (BMI)	Obesity	Nutrition control	4
10	$p4 \rightarrow n7 \rightarrow p8$	Pain or discomfort	Biopsy Test	Skin cancer	Radiation Therapy	4

The algorithm accepts P_{TB}^{T} sensor trajectory data with Bucket IDs, adversary prior knowledge delta, and threshold K to produce anonymized sensor trajectory data as the output. In the suppression procedure, the primary step is to group all the trajectory data into two groups: one with a similar Bucket ID for which the anonymization is to be carried and the other group with all remaining Bucket IDs. Split the trajectory data into two sets, say set X consists of trajectory data with '1' as the Bucket ID, and set Y consists of the records with remaining Bucket ids. Now from set X, find all the sub-trajectories of length 1 and verify that those trajectories appeared with a minimum of K-1 times in set Y. If any of the sub-trajectories fails to leave traces of K-1 times, those trajectories are eliminated; else, trajectories are preserved as same.

TABLE IV. ANONYMIZED TRAJECTORY DATA (P_{TB}^{TA})

PId	Trajectory	Bucket ID
1	$z3 \rightarrow p4 \rightarrow k6 \rightarrow p8$	1
6	$z3 \rightarrow n7 \rightarrow p8$	1
1	$d2 \rightarrow n5 \rightarrow k6 \rightarrow p9$	1
5	$d2 \rightarrow n5 \rightarrow k6 \rightarrow n7$	2
11	$z3 \rightarrow p4 \rightarrow k6 \rightarrow p8$	2
2	$k6 \rightarrow n7 \rightarrow p8$	2
4	$d2 \rightarrow k6 \rightarrow p9$	3
9	$d2 \rightarrow n5 \rightarrow p9$	3
7	$k6 \rightarrow n7 \rightarrow p8$	3
8	$n5 \rightarrow k6 \rightarrow p9$	4
3	$z3 \rightarrow k6 \rightarrow n7 \rightarrow p9$	4
10	$p4 \rightarrow p8$	4

Similarly, find all the sub-trajectories of length 2 from set X and validate with set Y for the minimum traces to K-1 times, failing to eliminate a trajectory point that does not satisfy the condition by computing the suppression score. Else keep the trajectory the same in the record. Repeat the procedure to calculate the critical trajectory point till the length of the sub-trajectory equals the adversary prior knowledge length delta. The complete process has to iterate for all the Bucket ID's for the PTBT to generate an anonymized sensor trajectory record PTBTA, as shown in Table IV.

Consider an example of the sensor trajectory $x1 \rightarrow d2 \rightarrow z3 \rightarrow p4 \rightarrow k6 \rightarrow p8$, which belongs to Bucket ID 1. The sensor trajectory is validated by computing the suppression score for critical trajectory points of the length 1 and 2, respectively, and found that the trajectory points x1 and d2 are critical to eliminating. On the complete trajectory iteration, the anonymized trajectory is $z3 \rightarrow p4 \rightarrow k6 \rightarrow p8$.

Algorithm 2 Anonymise_Trajectory(P_{TB}^{T})

INPUT: Sensor-Trajectory data P_{TB}^{T} with Bucket_ID, $A's$ prior knowledge ∂ with maximum length ρ, K threshold
OUTPUT: Anonymized Sensor Trajectory data P_{TB}^{TA}.

Scan Sensor Trajectory Table P_{TB}^{T}
let S_A = {set of all distinct sensitive values under the same Bucket_ID}
for each $sa \in S_A$ **do**
  $i = 1, \mathtt{Cr}_p = \emptyset, \mathtt{Dr}_i = \emptyset$
  $\mathtt{P} = \{P_{TBr}^{T} \mid P_{TBr}^{T} \in P_{TB}^{T} \wedge P_{TBr}^{T}(sa) = sa\}$
  $\mathtt{Q} = P_{TBr}^{T} - \{\mathtt{P}\}$
  for each $P_{TBr}^{T} \in \mathtt{P}$ **do**
    $\mathtt{Cr}_p = \{\tau_r \mid \tau_r \subseteq P_{TBr}^{T} \wedge \mid \tau_r \mid = 1\}$
    for each $\tau_r \in \mathtt{Cr}_p$ **do**
      if $(\mid \tau_r \in P_{TBr}^{T} \mid_{\forall P_{TBr}^{T} \in \mathtt{Q}} \geq \mathtt{K})$ **then**
        $\mathtt{Dr}_i = \mathtt{Dr}_i \cup \tau_r$
      else
        remove τ_r from $P_{TBr}^{T} \in \mathtt{P}$
      end if
    end for
  end for
  while $(i + 1 \leq \rho)$ **do**
    for each $\tau_r \in \mathtt{Dr}_i$ join with successive τ_{r+i} in $\mathtt{Dr}_i$ **do**
      if $(\mid \tau_r \cup \tau_{r+i} \in P_{TBr}^{T} \mid_{\forall P_{TBr}^{T} \in \mathtt{Q}} \geq \mathtt{k})$ **then**
        $\mathtt{Dr}_{i+1} = \mathtt{Dr}_{i+1} \cup \{\tau_r \cup \tau_{r+i}\}$
      else
        $\mathtt{tr} = \tau_r \cup \tau_{r+i}$
        remove $\chi(t)$ from $P_{TBr}^{T} \in \mathtt{P}$
      end if
    end for
    $i = i + 1$
  end while
end for

C. Slicing of Multiple Sensitive Attributes (P_{TB}^{SA})

Anonymizing the MSA P_{TB}^{SA} obtained after Bucketization and Partition implements the following slicing steps. In the slicing procedure, we first measure the weight of each sensitive attribute through the concept of entropy. Entropy refers to the average value of the information in each message gained. The sensitive attribute with higher entropy measures results as the qualitative information container. The entropy is measured with the following formula:

$$W_{SA} = -\sum_{j=1}^{D_{SA}} p(S_{vi}) \log(p(S_{vi})) \tag{4}$$

TABLE V. MULTIPLE SENSITIVE ATTRIBUTE (P_{TB}^{SA})

PId	Symptom	Diagnostic Method	Disease	Treatment	Bucket ID
1	Abdominal pain	X-ray	Abdominal Cancer	Chemotherapy	1
6	Shortness of breath	FeNO test	Asthma	Medication	1
1	Fever	Molecular diagnostic methods	Cholera	Antibiotic	1
5	Diarrhea	RT-PCR Tests	Dengue	Antibiotic	2
11	Abdominal pain	Ultrasound	Dyspepsia	Antibiotic	2
2	Weight loss	Antibody Test	HIV	Medication	2
4	Infection	ELISA Test	HIV	ART	3
9	Chest tightness	Methacholine challenge tests	Inflammation	Medication	3
7	Fever	RITD Tests	Influenza	Medicine	3
8	Weight loss	MRI Scan	Lung Cancer	Radiation Therapy	4
3	Eating disorders	Body mass index (BMI)	Obesity	Nutrition control	4
10	Pain or discomfort	Biopsy Test	Skin cancer	Radiation Therapy	4

Where SA is Sensitive Attribute, $\{S_{v1}, S_{v2}, S_{v3}, ..., S_{vi}\}$ set of possible values under SA. $p(S_{vi})$ possibility that S_{vi} is considered and D_{SA} number of distinct sensitve attributes.

The slicing procedure continues to find the association after computing the weight of each sensitive attribute using entropy. We adopt Pearson's Contingency Coefficient in the association computation to measure the association between two sensitive attributes. The analysis obeys the following formula:

$$\phi^2(SA_1, SA_2) = \frac{\sum_{i=1}^{n1}\sum_{j=1}^{n2}\frac{(p(SA_{ij})-p(SA_i)p(SA_j))^2}{p(SA_i)p(SA_j)}}{min\{n1, n2\} - 1}) \quad (5)$$

where, n1 and n2 are the total number of distinct values of SA_1 and SA_2 respectively. $p(SA_{ij})$ represents the chance from SA_{ij}. $p(SA_i)$ and $p(SA_j)$ are the boundary totals, where $p(SA_i) = \sum_{j=1}^{n2}$ and $p(SA_j) = \sum_{i=1}^{n1}$

On the computation of the weights of each sensitive attribute and the association among them, we combine the four columns of the MSA to form two columns for generating the sliced table. The sensitive attribute with the higher weights is selected as the slice's first column to avoid the adversaries' ease of data access. Then we choose the second column of the slice, with the maximum average association coefficient, with the other column. Doing so maximizes the association between the sensitive attributes in the same silce, and the coefficient for attributes in the different slices is minimized. After completing the above procedure, Table V with sensitive attributes like Symptoms, Diagnostic Method, Disease, and Treatment generates Table VI. Table VI has two slices on the attributes (Disease, Symptom) and (Diagnostic Method, and Treatment) to efficiently anonymize the data.

D. Publishing Anonymized Dataset on Merging P_{TB}^{TA} and P_{TB}^{SAA}

The anonymized sensor trajectory data and MSA data are merged to get the Table to give the input to the Reciprocal Bucketization. In this procedure, the sensor trajectory data remains untouched. In return to the Bucket ID, the MSA is combined to form three tuples having the three patients' data records under the same Bucket ID, which can reduce the adversary's confidence in identifying the individual through sensor trajectory data or the multiple sensitive attributes. In merging the corresponding records, the Patient ID is the referral point. Finally, the anonymized data is published by sorting the records according to the patient ID as represented in the Table VII.

Algorithm 3 Anonymise_Sensitive_Attribute(P_{TB}^{SA})

INPUT: Multiple Sensitive Attributes P_{TB}^{SA} with Bucket_ID
OUTPUT: Anonymized Multiple Sensitive Attributes P_{TB}^{SAA}.

1: **Scan** Multiple Sensitive Attributes Table P_{TB}^{SA}
2: **for** each attribute in P_{TB}^{SA} **do**
3: compute W_{SA} store in $W_{SA}Array$
4: **end for**
5: $SA\_Max_1 = First\_Max(W_{SA}Array)$
6: $SA\_Max_2 = Second\_Max(W_{SA}Array)$
7: $Slice\_01 = (SA\_Max_1)$
8: $Slice\_02 = (SA\_Max_2)$
9: Compute Pearson Contingency Coefficient of MSA excluding $(SA\_Max_1)and(SA\_Max_2)$
10: compute $\phi^2(SA_1, SA_2)$ store in $PCC_{SA}Array$
11: $PCC\_Max_1 = First_Max(PCC_{SA}Array)$
12: $PCC\_Max_2 = Second_Max(PCC_{SA}Array)$
13: $Slice\_01 = (SA\_Max_1, PCC\_Max_1)$
14: $Slice\_02 = (SA\_Max_2, PCC\_Max_2)$
15: return($Slice\_01$,$Slice\_02$)

The RB – Anonymization model generates the anonymized dataset to be published, as represented in Table VII, and it is resistant to Identity, Attribute, and Correlated linkage attacks. To our knowledge, RB -Anonymization model is the first approach to anonymize the trajectory data with the MSA. If adversary prior expertise with the sensor trajectory length delta = 2 as d2,p4, then Adversary 'A' can perform all three linkage attacks on Table I as discussed in Section III. However, the Adversary fails to identify Avok's record from Table VII because the not even one record with d2,p4. Only by finding the sensory trajectory point d2, the Adversary could declare the identification with less than 37% confidence. The MSA couldn't be reached with the same Adversary's prior knowledge.

Similarly, knowing the partial sensor trajectory k6, n7 with the significant MSA as the disease HIV could be predicted in Table I. But from Table VII even though the k6, n7 trajectory is repeated more than one time, the adversary failed to identify its

TABLE VI. SLICED MULTIPLE SENSITIVE ATTRIBUTE (P_{TB}^{SAA})

PId	(Disease, Symptom)	(Diagnostic Method , Treatment)	Bucket ID
1	(Abdominal Cancer , Abdominal pain)	(X-ray , Chemotherapy)	1
6	(Cholera , Fever)	(Molecular diagnostic methods , Antibiotic)	1
1	(HIV , Weight loss)	(Antibody Test , Medication)	1
5	(Obesity , Eating disorders)	(Body mass index (BMI) , Nutrition control)	2
11	(HIV , Infection)	(ELISA Test , ART)	2
2	(Dengue , Diarrhea)	(RT-PCR Tests , Antibiotic)	2
4	(Asthma , Shortness of breath)	(FeNO test , Medication)	3
9	(Influenza , Fever)	(RITD Tests , Medicine)	3
7	(Lung Cancer , Weight loss)	(MRI Scan , Radiation Therapy)	3
8	(Inflammation , Chest Tightness)	(Methacholine challenge tests , Medication)	4
3	(Skin cancer , Pain or discomfort)	(Biopsy Test , Radiation Therapy)	4
10	(Dyspepsia , Abdominal pain)	(Ultrasound , Antibiotic)	4

TABLE VII. MERGING AND RECIPROCAL BUCKETIZATION P_{TB}^{A}

PId	Trajectory	(Disease, Symptom)	(Diagnostic Method , Treatment)
1	$z3 \rightarrow p4 \rightarrow k6 \rightarrow p8$	(Abdominal Cancer , Abdominal pain) (Cholera , Fever) (HIV , Weight loss)	(X-ray , Chemotherapy) (Molecular diagnostic methods , Antibiotic) (Antibody Test , Medication)
2	$k6 \rightarrow n7 \rightarrow p8$	(Obesity , Eating disorders) (HIV , Infection) (Dengue , Diarrhea)	(Body mass index (BMI) , Nutrition control) (ELISA Test , ART) (RT-PCR Tests , Antibiotic)
3	$z3 \rightarrow k6 \rightarrow n7 \rightarrow p9$	(Inflammation , Chest Tightness) (Skin cancer , Pain or discomfort) (Dyspepsia , Abdominal pain)	(Methacholine challenge tests , Medication) (Biopsy Test , Radiation Therapy) (Ultrasound , Antibiotic)
1	$d2 \rightarrow n5 \rightarrow k6 \rightarrow p9$	(Abdominal Cancer , Abdominal pain) (Cholera , Fever) (HIV , Weight loss)	(X-ray , Chemotherapy) (Molecular diagnostic methods , Antibiotic) (Antibody Test , Medication)
4	$d2 \rightarrow k6 \rightarrow p9$	(Asthma , Shortness of breath) (Influenza , Fever) (Lung Cancer , Weight loss)	(FeNO test , Medication) (RITD Tests , Medicine) (MRI Scan , Radiation Therapy)
5	$d2 \rightarrow n5 \rightarrow k6 \rightarrow n7$	(Obesity , Eating disorders) (HIV , Infection) (Dengue , Diarrhea)	(Body mass index (BMI) , Nutrition control) (ELISA Test , ART) (RT-PCR Tests , Antibiotic)
6	$z3 \rightarrow n7 \rightarrow p8$	(Abdominal Cancer , Abdominal pain) (Cholera , Fever) (HIV , Weight loss)	(X-ray , Chemotherapy) (Molecular diagnostic methods , Antibiotic) (Antibody Test , Medication)
7	$k6 \rightarrow n7 \rightarrow p8$	(Asthma , Shortness of breath) (Influenza , Fever) (Lung Cancer , Weight loss)	(FeNO test , Medication) (RITD Tests , Medicine) (MRI Scan , Radiation Therapy)
8	$n5 \rightarrow k6 \rightarrow p9$	(Inflammation , Chest Tightness) (Skin cancer , Pain or discomfort) (Dyspepsia , Abdominal pain)	(Methacholine challenge tests , Medication) (Biopsy Test , Radiation Therapy) (Ultrasound , Antibiotic)
9	$d2 \rightarrow n5 \rightarrow p9$	(Asthma , Shortness of breath) (Influenza , Fever) (Lung Cancer , Weight loss)	(FeNO test , Medication) (RITD Tests , Medicine) (MRI Scan , Radiation Therapy)
10	$p4 \rightarrow p8$	(Inflammation , Chest Tightness) (Skin cancer , Pain or discomfort) (Dyspepsia , Abdominal pain)	(Methacholine challenge tests , Medication) (Biopsy Test , Radiation Therapy) (Ultrasound , Antibiotic)
11	$z3 \rightarrow p4 \rightarrow k6 \rightarrow p8$	(Obesity , Eating disorders) (HIV , Infection) (Dengue , Diarrhea)	(Body mass index (BMI) , Nutrition control) (ELISA Test , ART) (RT-PCR Tests , Antibiotic)

MSA due to the bucketization where three tuples are confusing to decide, and the probability of finding is less than 0.2, which is slightly negligible.

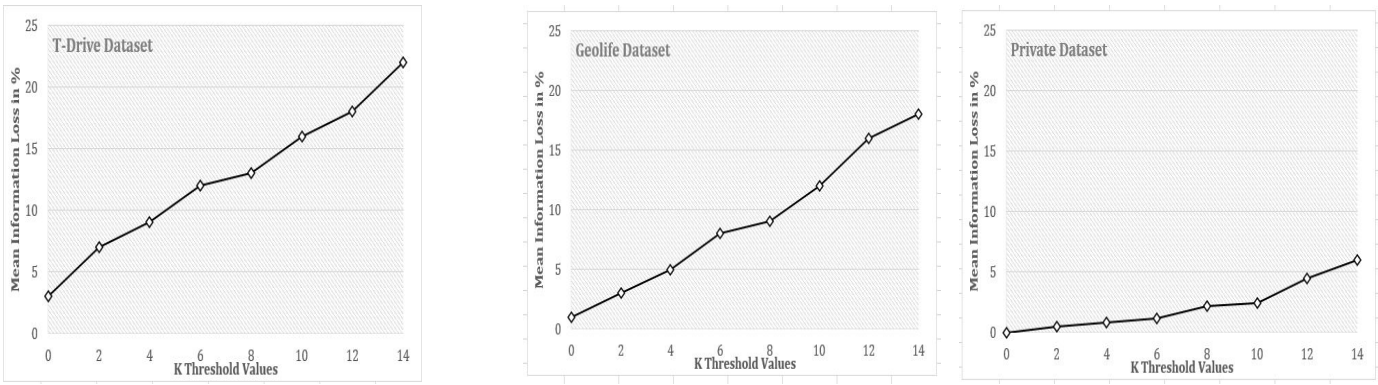
Correlated linkage attacks concerning multiple sensory trajectories with the same patient ID and establishing the relations among the sensitive attributes are well addressed by RB – Anonymization. The significant bucketization process brings down the confidence level of the adversary in the correlated linkage attack to less than 30% as the outcome of Bucketization and combining the records under the same bucket ID.

V. RESULTS AND DISCUSSION

The proposed approach uses the Windows 10 operating system, i5 processor with a minimum of 4GB RAM and 256 GB SSD. The algorithm implementation uses the Python 3 version of the Synthetic dataset. The number of instances considered is more excellent than 10,000. The dataset's attributes are PId, Sensor Trajectory, Disease, Symptom, Diagnostic Method, and Treatment. Disease, Symptom, Diagnostic Method, and Treatment categorizes as MSA. The proposed approach's performance evaluates in terms of information loss in anonymized sensor trajectory data and the utility loss in MSA (P_{TB}^{A}).

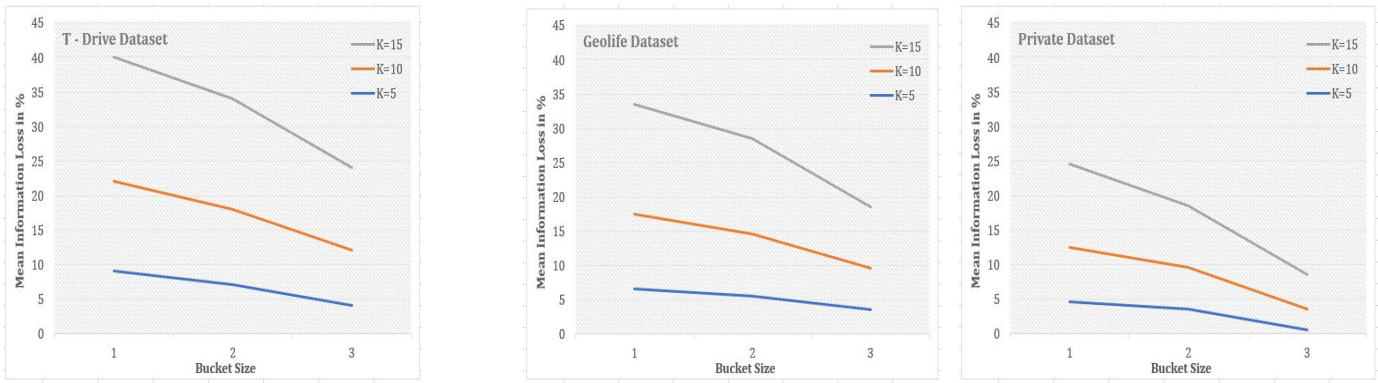
A. Sensor Trajectory Information Loss

Information loss occurs during the anonymization process because of the methods used, like generalization or suppression. Analysing the number of sensor trajectories information loss in the resultant anonymized dataset is quite significant. Eliminating the critical trajectory moving point results in



(a) (b) (c)

Fig. 3. Mean sensor trajectory information loss in P_{TB}^{TA} with K threshold values. (a) T- Drive dataset. (b) Geolife dataset. (c) Private dataset.



(a) (b) (c)

Fig. 4. Mean sensor trajectory information loss in P_{TB}^{TA} with bucket size. (a) T- Drive dataset. (b) Geolife dataset. (c) Private dataset.

information loss during the procedure to satisfy the privacy requirements.

The sensor trajectory information loss is computed as follows between the P_{TB}^{T} (Original Sensor Trajectory) and P_{TB}^{TA} (Anonymized Sensor Trajectory).

Information Loss(I_L) of each record is computed as:

$$I_L(P_{TBi}^{TA}) = \frac{|P_{TBi}^{T}| - |P_{TBi}^{TA}|}{|P_{TBi}^{T}|} \tag{6}$$

Total Trajectory Information Loss is computed as:

$$I_L(P_{TB}^{TA}) = \sum_{i=1}^{|P_{TB}^{TA}|} I_L(P_{TBi}^{TA}) \tag{7}$$

where, P_{TBi}^{T} represents total number of trajectory points in i^{th} record of P_{TB}^{T} and P_{TBi}^{TA} represents total number of trajectory points in i^{th} of P_{TB}^{TA}.

Fig. 3 shows the mean sensor trajectory information loss in the seven sensor trajectories corresponding to the various K threshold values. The graph shows that the increase in K values is directly proportional to the information loss on the anonymized data due to the random increase in the critical trajectory points, which failed to reach the privacy requirements. Hence the data publisher has to adopt the K value carefully to hold the moderate data loss and to preserve privacy.

Effect of Bucket Size: The RB anonymity model splits the given records in the user input bucket size, where we have taken as 3. The number of sensor trajectories that fall under the bucket is directly proportional to the bucket size. Anonymizing the trajectories while keeping the constraints of K threshold values and adversary prior knowledge varies on the bucket size. Fig. 4 represents the mean Information loss of the patient's sensor trajectory by keeping the adversary prior knowledge sigma to 2 and changing the K threshold values. The significant observation from the graph is that as the bucket's size increases, the information loss reduces due to the more trajectories falling into the bucket and the less elimination of trajectory critical point. It leads to quick access to the sensitive attributes hence the bucket size has to be chosen appropriately by providing equal significance to the sensor trajectory and the multiple sensitive attributes.

B. Utility Loss in Multiple Sensitive Attributes

The utility loss in MSA measures probability distribution across the actual values and the RB – Anonymize data. We adopt the Kullback – Leibler divergence metric to estimate the probability distribution differences. The MSA P_{TB}^{SA} table is considered as actual values for implementing the KL divergence as x1. x1(r) are the elements of the records for R, i.e., (r belongs to R). x2 is an estimated probability distribution considered after anonymization (P_{TB}^{A}). The KL divergence is given as follows:

$$KL_{div}(x1, x2) = \sum_{r \in R} x1(r) \log(\frac{x1(r)}{x2(r)}) \tag{8}$$

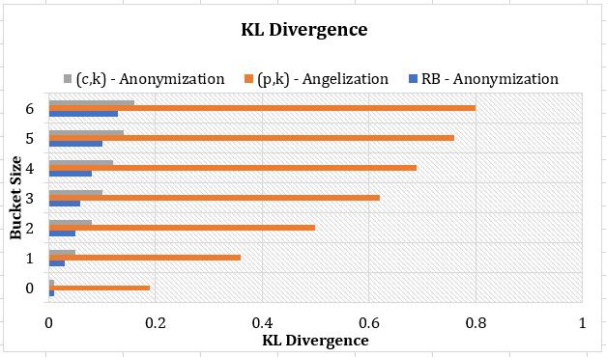


Fig. 5. KL Divergence of MSA.

Fig. 5 represents the utility loss comparison between the existing methods discussed in Section II with the proposed RB – Anonymization model. We can notify the significant differences, such as the (p,k) – angelization may assure high privacy. Still, the estimated probability distribution increases, resulting in a moderate utility loss. (c,k) – Anonymization and RB – Anonymization approach has a slight difference in the probability distribution with negligible utility loss; RB – Anonymization model outperforms consistently even if bucket size increases.

The analysis of the RB – Anonymization is validated by measuring the execution time and the privacy loss. Since no exact dataset has the sensor trajectory data and the multiple sensitive attributes, we have built our own synthetic dataset to evaluate the approach. The proposed model ensures its high adaptability towards real-time applications through its little execution time for anonymization with greater privacy. A moderate number of records and sensitive attributes are preferable since the execution time is proportional to both attributes. As the RB – Anonymization approach outperforms in terms of privacy, it also requires assuring the privacy loss on publishing the anonymized data. The privacy loss is negligible if there is less patient data exposure in terms of sensor trajectory and the MSA. The proposed approach brings new challenges to the adversaries through its stringent privacy preservation approach.

VI. Conclusions

In this paper, we present a novel privacy preservation method considering the Smart Hospital data, consisting of sensor trajectory and the MSA. The proposed method outperforms in defending Identity, Attribute, and Correlated linkage attacks on data publishing. Our approach adopts a local suppression to anonymize the sensor trajectory and the slicing for MSA with a constraint on the bucket size. The proposed method outperforms on comparing the information loss of the sensor trajectory and MSA with (p,k) – angelization and (c,k) – anonymization approaches implemented on the real-time and synthetic dataset. As a future study, we are interested in addressing all the various linkage attacks and, as the primary concentration on the MSA, enabling our proposed method as the best practice.

References

[1] G. Xu, "IoT-Assisted ECG Monitoring Framework with Secure Data Transmission for Health Care Applications," *IEEE Access*, vol. 8, pp. 74 586–74 594, 2020.

[2] Y. Shi, Z. Zhang, H. C. Chao, and B. Shen, "Data Privacy Protection Based on Micro Aggregation with Dynamic Sensitive Attribute Updating," *Sensors (Switzerland)*, vol. 18, no. 7, pp. 1–16, 2018.

[3] A. Majeed and S. Lee, "Anonymization Techniques for Privacy Preserving Data Publishing: A Comprehensive Survey," *IEEE Access*, vol. 9, pp. 8512–8545, 2021.

[4] N. Y. Philip, M. Razaak, J. Chang, M. S. Suchetha, M. Okane, and B. K. Pierscionek, "A Data Analytics Suite for Exploratory Predictive, and Visual Analysis of Type 2 Diabetes," *IEEE Access*, vol. 10, pp. 13 460–13 471, 2022.

[5] R. Khan, X. Tao, A. Anjum, H. Sajjad, S. U. R. Malik, A. Khan, and F. Amiri, "Privacy Preserving for Multiple Sensitive Attributes against Fingerprint Correlation Attack Satisfying c -Diversity," *Wireless Communications and Mobile Computing*, vol. 2020, pp. 1–18, 2020.

[6] N. L. and Raju, M. Seetaramanath, and P. S. Rao, "A Novel Dynamic KCi - Slice Publishing Prototype for Retaining Privacy and Utility of Multiple Sensitive Attributes," *International Journal of Information Technology and Computer Science*, vol. 11, no. 4, pp. 18–32, 2019.

[7] E. G. Komishani and M. Abadi, "A Generalization-Based Approach for Personalized Privacy Preservation in Trajectory Data Publishing," *In the Proceedings of 6th International Symposium on Telecommunications, IST 2012*, pp. 1129–1135, 2012.

[8] Y. Alotaibi, "A New Secured E-Government Efficiency Model for Sustainable Services Provision," *Journal of Information Security and Cybercrimes Research*, vol. 3, no. 1, pp. 75–96, 2020.

[9] A. Ye, Q. Zhang, Y. DIao, J. Zhang, H. Deng, and B. Cheng, "A Semantic-Based Approach for Privacy- Preserving in Trajectory Publishing," *IEEE Access*, vol. 8, pp. 184 965–184 975, 2020.

[10] F. Jin, W. Hua, M. Francia, P. Chao, M. Orowska, and X. Zhou, "A Survey and Experimental Study on Privacy-Preserving Trajectory Data Publishing," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 6, pp. 5577–5596, 2022.

[11] D. Hemkumar, S. Ravichandra, and D. V. Somayajulu, "Impact of Prior Knowledge on Privacy Leakage in Trajectory Data Publishing," *Engineering Science and Technology, an International Journal*, vol. 23, no. 6, pp. 1291–1300, 2020. [Online]. Available: https://doi.org/10.1016/j.jestch.2020.06.002

[12] J. N. Vanasiwala and N. R. Nanavati, "Multiple sensitive attributes based privacy preserving data publishing," *In the Proceedings of the 2nd International Conferenceon Computing Methodologies and Communication, ICCMC*, pp. 394–400, 2018.

[13] Y. Xiao and H. Li, "Privacy Preserving Data Publishing for Multiple Sensitive Attributes Based on Security Level," *Information (Switzerland)*, vol. 11, no. 3, p. 166, 2020.

[14] X. Liu and Y. Zhu, "Privacy and Utility Preserving Trajectory Data Publishing for Intelligent Transportation Systems," *IEEE Access*, vol. 8, pp. 176 454–176 466, 2020.

[15] J. Zhao, J. Mei, S. Matwin, Y. Su, and Y. Yang, "Risk-Aware Individual Trajectory Data Publishing with Differential Privacy," *IEEE Access*, vol. 9, pp. 7421–7438, 2021.

[16] S. S. Vedaei, A. Fotovvat, M. R. Mohebbian, G. M. Rahman, K. A. Wahid, P. Babyn, H. R. Marateb, M. Mansourian, and R. Sami, "COVID-SAFE: An IoT-based System for Automated Health Monitoring and Surveillance in Post-Pandemic Life," *IEEE Access*, vol. 8, pp. 188 538–188 551, 2020.

[17] E. Luo, M. Z. A. Bhuiyan, G. Wang, M. A. Rahman, J. Wu, and M. Atiquzzaman, "PrivacyProtector: Privacy-Protected Patient Data Collection in IoT-Based Healthcare Systems," *IEEE Communications Magazine*, vol. 56, no. 2, pp. 163–168, 2018.

[18] E. Ghasemi Komishani, M. Abadi, and F. Deldar, "PPTD: Preserving Personalized Privacy in Trajectory Data Publishing by Sensitive Attribute Generalization and Trajectory Local Suppression," *Knowledge-Based Systems*, vol. 94, pp. 43–59, 2016.

[19] L. Yao, Z. Chen, H. Hu, G. Wu, and B. Wu, "Sensitive Attribute Privacy Preservation of Trajectory Data Publishing Based on l-diversity," *Distributed and Parallel Databases*, vol. 39, no. 3, pp. 785–811, 2021. [Online]. Available: https://doi.org/10.1007/s10619-020-07318-7

[20] R. Tojiboev, W. Lee, and C. C. Lee, "Adding Noise Trajectory for Providing Privacy in Data Publishing by Vectorization," *In the Proceedings of - IEEE International Conference on Big Data and Smart Computing, BigComp 2020*, pp. 432–434, 2020.

[21] L. Yao, Y. Zhang, Z. Zheng, and G. Wu, "GAN-Based Differential Privacy Trajectory Data Publishing with Sensitive Label," *In the Proceedings of - 8th International Conference on Big Data Computing and Communications, BigCom 2022*, pp. 112–119, 2022.

[22] R. Wen, W. Cheng, H. Huang, W. Miao, and C. Wang, "Privacy Preserving Trajectory Data Publishing with Personalized Differential Privacy," *In the Proceedings of - ISPA-BDCloud-SocialCom-SustainCom 2020*, pp. 313–320, 2020.

[23] T. Kanwal, S. A. A. Shaukat, A. Anjum, S. u. R. Malik, K. K. R. Choo, A. Khan, N. Ahmad, M. Ahmad, and S. U. Khan, "Privacy-Preserving Model and Generalization Correlation Attacks for 1:M Data with Multiple Sensitive Attributes," *Information Sciences*, vol. 488, pp. 238–256, 2019. [Online]. Available: https://doi.org/10.1016/j.ins.2019.03.004

[24] T. Li, N. Li, J. Zhang, and I. Molloy, "Slicing: A New Approach for Privacy Preserving Data Publishing," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 3, pp. 561–574, 2012.

[25] J. Jayapradha, M. Prakash, Y. Alotaibi, O. I. Khalaf, and S. A. Alghamdi, "Heap Bucketization Anonymity - An Efficient Privacy-Preserving Data Publishing Model for Multiple Sensitive Attributes," *IEEE Access*, vol. 10, pp. 28 773–28 791, 2022.

[26] L. Yao, Z. Chen, X. Wang, D. Liu, and G. Wu, "Sensitive Label Privacy Preservation with Anatomization for Data Publishing," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 2, pp. 904–917, 2021.

[27] Z. S. H. Abad, D. M. Maslove, and J. Lee, "Predicting Discharge Destination of Critically Ill Patients Using Machine Learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 3, pp. 827–837, 2021.

[28] F. Song, T. Ma, Y. Tian, and M. Al-Rodhaan, "A New Method of Privacy Protection: Random k-Anonymous," *IEEE Access*, vol. 7, pp. 75 434–75 445, 2019.

[29] L. Zhang, J. Xuan, R. Si, and R. Wang, "An Improved Algorithm of Individuation K-Anonymity for Multiple Sensitive Attributes," *Wireless Personal Communications*, vol. 95, no. 3, pp. 2003–2020, 2017.

[30] Z. Li and X. Ye, "Privacy Protection on Multiple Sensitive Attributes," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4861 LNCS, pp. 141–152, 2007.

[31] V. S. Susan and T. Christopher, "Anatomisation with Slicing: A New Privacy Preservation Approach for Multiple Sensitive Attributes," *SpringerPlus*, vol. 5, no. 1, pp. 1–21, 2016.