International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V72I12P113

Volume 72 Issue 12, 140-150, December 2024

Original Article

Deep Learning Model for Privacy Preservation of Vehicle Trajectories Over Internet of Vehicles

Nikhil Chaurasia[1], Pritaj Yadav[2], Sanjeev Kumar Gupta[3]

[1,2,3]*Department of CSE, Rabindranath Tagore University Bhopal Madhya Pradesh, India.*

[1]*Corresponding Author : nikhilsub97@gmail.com*

Received: 07 June 2024 Revised: 05 October 2024 Accepted: 05 December 2024 Published: 25 December 2024

Abstract - *This research presents a deep learning-based differential privacy Laplace mechanism (DLPM) for the networking trajectories of the Internet of Vehicles. The DLPM is constructed using deep learning and clustering techniques to address consumers' privacy leakage concerns effectively. Segment the trajectory space into separate regions based on timestamps to incorporate the temporal element in trajectories. This will determine the trajectory's distribution points inside each zone. Enhance membership stability in multi-peak clustering for each region and pre-allocate the privacy budget matrix based on the trajectory density of each area. Employ a temporal graph convolutional network model to train and predict the designated privacy budget matrix while extracting the spatiotemporal characteristics of trajectory data. Disseminate the trajectory data just after applying Laplace noise to the predictive outcomes. Both theoretical and empirical research indicate that DLPM exhibits reduced overhead and more precise privacy budget predictions than alternative systems. The Differential Privacy Laplace Mechanism (DLPM) incorporates Laplace noise into trajectory data, enhancing its use.*

Keywords - *Internet of Things, Privacy protection, Connected vehicle, Trajectory, Clustering, Deep learning.*

1. Introduction

In Vehicular Ad-Hoc Networks (VANETs) [1], users generate extensive trajectory data through location-based services [2], posing significant privacy risks due to the sensitive nature of location information [3]. Traditional privacy protection methods, such as k-anonymity [4], l-diversity [5], and t-closeness [6], anonymize trajectory data to prevent privacy breaches. However, these techniques are vulnerable to advanced attacks, including homogeneity, background knowledge, and combination attacks. To mitigate these risks, Dwork's [7] introduction of Differential Privacy (DP) provided a theoretical framework to protect against background knowledge attacks by adding noise to data. Yet, excessive noise often degrades the utility of the protected data, creating a trade-off between privacy and usability [8]. Addressing the challenge of data quality loss in most clustering scenarios, the author [9] introduced a system, DPTD, which uses DBSCAN clustering for trajectory publication, safeguarding most trajectories' privacy. The author [10] introduced the LGAN-DP technique, which utilizes deep learning for trajectory synthesis and applies k-means clustering to analyze trajectory result sets, thereby improving data privacy. K-means and DBSCAN clustering significantly rely on user-defined parameters, resulting in inconsistent performance. The author [11] presented the Stable-Membership Multi-Peak Clustering (SMMP) algorithm, which utilizes stable membership degrees to resolve parameter tuning challenges. However, it is constrained to low-dimensional spaces [12]. This paper presents an enhanced algorithm, the Improved Stable-Membership Multi-Peak Clustering (ISMMPC), which autonomously adjusts clustering thresholds, performs multi-prototype clustering and tackles challenges in multi-dimensional application scenarios associated with the Internet of Things.

Author [13] introduced the DPGeo framework, employing DP-collected perturbed datasets from adversarial autoencoders to train trajectory generators [14]. Nonetheless, trajectory precision is confined to grid-level representations. Recent research by Fed-Inforce-Fusion has presented a federated reinforcement model for safeguarding IoMT networks. [16] examines the differential privacy of deep learning, whereas [17] proposes a framework for the private generation of trajectories utilizing deep learning. The GeoPM-DMEIRL model [18] uses deep inverse reinforcement learning to provide secure IoMT trajectories. These enhancements underscore the ongoing efforts to bolster the privacy and security of the IoMT. The author presented the Spatiotemporal Long Short-Term Memory (LSTM) model, incorporating Laplace noise to forecast position structures; nevertheless, the specified privacy budget is imprecise. The Internet of Things has numerous applications, such as collaboration among urban Internet of

Vehicles, pedestrian safety through IoT and sensors, edge AI for the Internet of Everything, sustainable rail transportation, and future person re-identification via the Internet of Bodies. The author employed a trajectory prediction Hidden Markov Model utilizing spatiotemporal density clustering. Accommodates smooth data but fails to capture concealed nonlinear features in trajectories. This study integrates the T-GCN model with differential privacy technology, effectively forecasting digital privacy budgets and investigating linear and nonlinear characteristics embedded in trajectories.

Research on the security and privacy of the Internet of Things encompasses a comprehensive examination of the challenges related to 5G, a privacy-centric federated learning framework for smart city mobility [24], a federated reinforcement learning-enhanced driving system [25], and an environmentally considerate approach to forest fire classification [23]. A digital privacy and security evaluation of smart city federated learning is also accessible. This study enhances the understanding and management of vulnerabilities and privacy problems associated with the Internet of Things. This analysis examines the influence of data sparsity, caused by timestamp-based trajectory region partitioning, on the digital privacy protection of trajectories, considering the spatiotemporal characteristics of trajectory space.

The ISMMPC algorithm and T-GCN model address significant privacy issues in the Internet of Things, utilizing a Q-Learning model for flow compensation incentives. Various studies employ Multi-Agent Deep Reinforcement Learning to focus on a privacy-preserving offloading mechanism [26]. A survey [27] has thoroughly studied the challenges of trust-based security in the Internet of Things. Moreover, research has examined the potential collaboration between the Internet of Things and blockchain to enhance security [28]. Electric vehicles are associated with the Internet-of-Batteries (IoB) idea, which prioritizes sustainable solutions and is engineered to improve the efficacy of digital privacy protection and the functionality of trajectory data.
Addressing these limitations, this paper introduces a novel privacy preservation method incorporating the Improved Stable-Membership Multi-Peak Clustering (ISMMPC) algorithm and the Temporal Graph Convolutional Network (T-GCN) model. This approach addresses privacy risks in the multi-dimensional trajectory space representing vehicle trajectories by mapping their time-based movement across geographic coordinates and timestamps. The ISMMPC algorithm dynamically adjusts clustering thresholds, supports multi-dimensional spaces, and retains temporal and spatial features during clustering, which is critical for privacy protection in the Internet of Vehicles (IoV). Meanwhile, the T-GCN model improves privacy budget prediction by analyzing spatiotemporal dependencies in trajectory data, allowing for a more adaptive noise allocation that enhances data usability. This dual approach uniquely mitigates privacy risks associated with consecutive location points in trajectory data—a commonly overlooked aspect in prior research, where the focus has primarily been on individual data points. The main contributions of this research include:

1. A novel privacy preservation framework for trajectory data in IoV that integrates ISMMPC with T-GCN to enhance stability and adaptability across complex spatiotemporal structures.
2. A more accurate prediction of privacy budgets by utilizing T-GCN to model spatiotemporal dependencies in trajectory data.
3. A comprehensive experimental validation using real datasets (Divvy Bikes and T-drive) demonstrates the proposed model's superior balance between privacy protection and data usability compared to state-of-the-art methods.

In Section 2, we lay down the groundwork by reviewing the pertinent literature and previous research. Section 3 describes the Dynamic Load Balancing and Power Management technique in depth, whether it is suggested or used. Section 4 details the experimental design and dataset description, whereas Section 5 presents results, discussion and performance assessment. The study concludes with a summary of the main points, consequences, and future prospects in Section 6.

2. Background Knowledge

2.1. Trajectory Model

Definition 1: A grid graph H. H = (W;F) , F is the set of edges, $W = \{W_1, W_2, \cdots, W_{n\times n}\}$ is the node at the grid position, $n \times n$ is the number of grid area positions.

Definition 2: Feature Matrix Y. Node attributes contain grid trajectory data, forming the feature matrix Y. $\boldsymbol{Y}_{t_g}$ represents the trajectory information at the time t_g, where t_g is the timestamp. Spatiotemporal trajectories are trajectory data obtained by learning mapping functions in both time t and spatial coordinates Y and can be represented as:

$$\left[Y_{i_g+1}, \cdots, Y_{t_g+T}\right] = G\left(G; \left(Y_{t_g-m}, \cdots, Y_{t_g-1}, Y_{t_g}\right)\right) \quad (1)$$

The forecast time series step is T, while the past time series length is m.

Definition 3: Let B_g be the region partitioned by t_g. Based on B_g ,n×n grid regions are obtained, and matrix T is constructed to record the number of trajectories in B_g . Matrix element $t_{ij}(i, j = 1,2, \cdots, n)$ corresponds to the cumulative trajectory count in the grid region, where i and j are element indices. $\max_T$ represents the maximum t_{ij}, and T can be expressed as:

$$T = \begin{bmatrix} t_{11} & \cdots & t_{1n} \\ \vdots & & \vdots \\ t_{n1} & \cdots & t_{nn} \end{bmatrix} \quad (2)$$

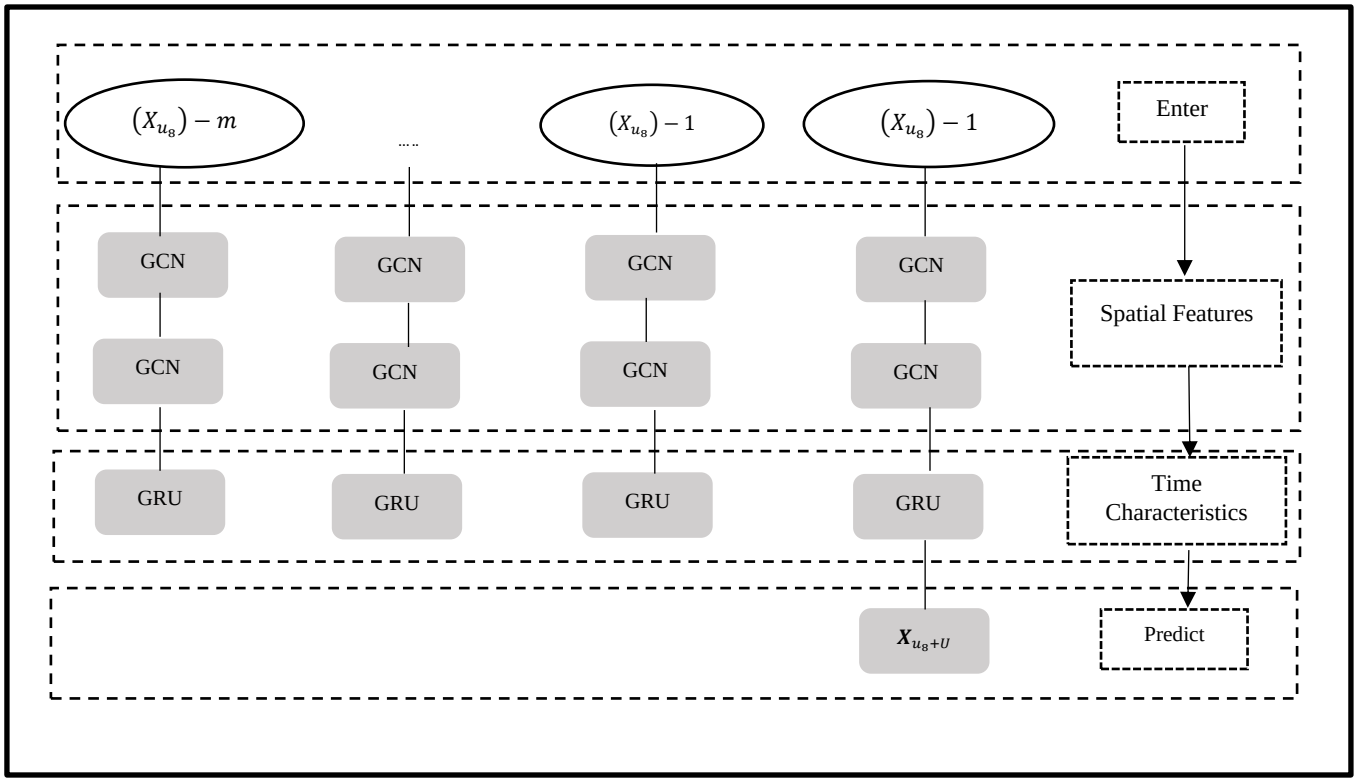


Fig. 1 Temporal graph convolutional network model

Definition 4: Density Matrix Q Construct Matrix Q to represent the trajectory density of B_g . Matrix element $q_{ij}(i,j = 1,2,\cdots,n)$ Indicates the density of trajectories in the corresponding grid region, where $q_{ij} = t_{ij}/a_{B_g}$ is the area of B_g. Q can be expressed as:

$$Q = \begin{bmatrix} q_{11} & \cdots & q_{1n} \\ \vdots & & \vdots \\ q_{n1} & \cdots & q_{nn} \end{bmatrix} \quad (3)$$

Definition 5: Privacy Budget Matrix F. Construct matrix F to represent the privacy budget allocated for B_g. Matrix element $\varepsilon_{ij}(i,j = 1,2,\cdots,n)$ represents the grid region's privacy budget. The initial value of ε_{ij} is 0. F can be expressed as:

$$F = \begin{bmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & & \vdots \\ f_{n1} & \cdots & f_{nn} \end{bmatrix} \quad (4)$$

2.1. Differential Privacy

Definition 6: Differential Privacy. Given a random algorithm N, where the set of all possible outputs is O, and the probability distribution of O is denoted by Q [·], for any two neighbouring datasets E and E′, if the probability distributions of the two neighbouring sets satisfy:

$$Q[K(E) \in O_K] \leqslant \mathrm{e}^{\varepsilon} \times Q[K(E') \in O_K], \quad (5)$$

Then the algorithm N provides ε-Differential Privacy. Here, Q [·] represents the probability of privacy leakage, and ε represents the degree of privacy protection, where ε∈ (0,1].

Definition 7: Global Sensitivity. For a function $f: E \to S^e$ where E is the domain for any neighboring datasets E and E′, the global sensitivity iss given by:

$$\Delta f = \max_{E,E'} \|f(E) - f(E')\|_1, \quad (6)$$

Where e is the query dimension and $\|\cdot\|$ represents the L_1 norm.

Definition 8: Laplace Mechanism. For a given dataset E, assuming there is a function $f: E \to S^e$ with sensitivity Δf, the Laplace Mechanism is defined as K(E)=f(E)+Y, where the noise Y follows the Laplace distribution.

$$Y \sim \mathrm{Lap}\left(\frac{\Delta f}{\varepsilon}\right). \quad (7)$$

Y Δf" is directly proportional to and inversely proportional to

2.2. Temporal Graph Convolutional Network (T-GCN) Model

The GCN-GRU Temporal Graph Convolutional Network (T-GCN) model is shown in Figure 1. T-GCN graphs convolution on m time series data using a 2-layer GCN model to understand spatial features from the road network region's intricate structure. Using spatial time series, the GRU model captures temporal data through dynamic information transmission between units. T-GCN effectively learns spatiotemporal dependencies, enabling trajectory prediction. This can be expressed as:

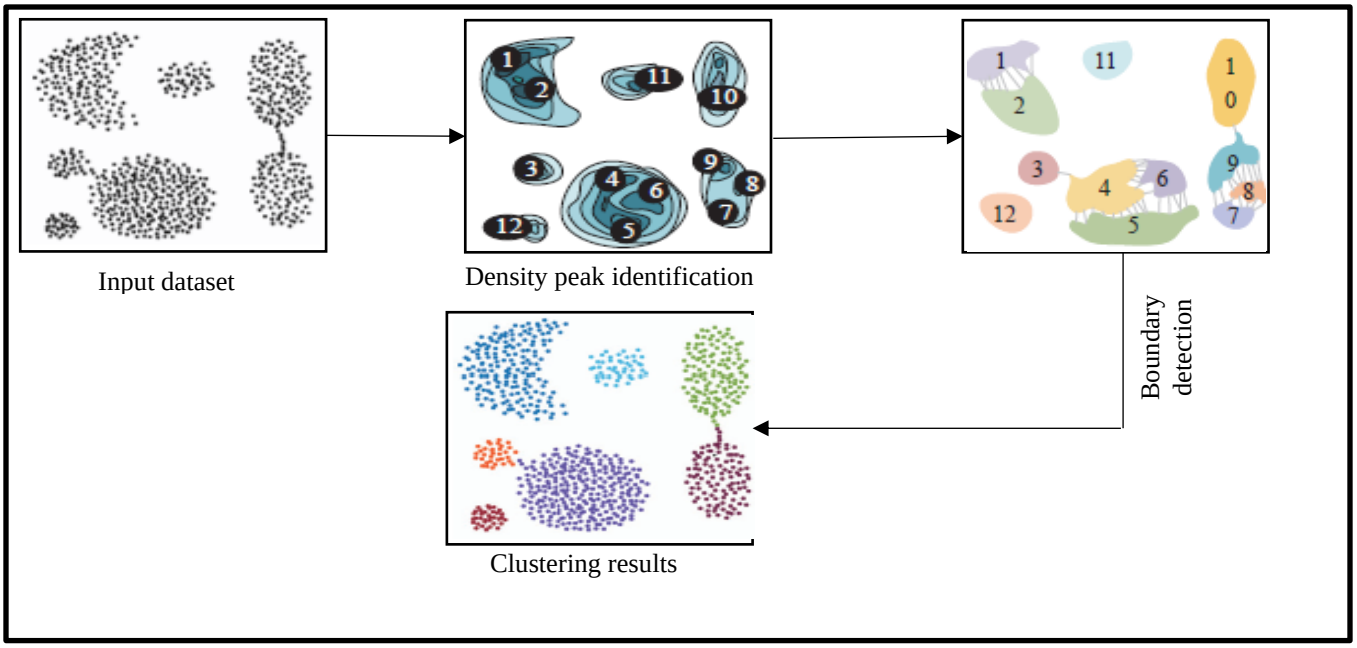


Fig. 2 Improved stable-membership multi-peak clustering process

$$
\begin{aligned}
&G(N,X) = \sigma\left(\widehat{N}\mathrm{Relu}\left(\widehat{N}XZ_0\right)Z_0\right),\\
&v_{t_g} = \sigma\left(Z_v\left[G\left(N,X_{t_g}\right),g_{t_g-1}\right]+c_v\right),\\
&r_{t_g} = \sigma\left(Z_r\left[G\left(N,X_{t_g}\right),g_{t_g-1}\right]+c_r\right),\\
&c_{t_g} = \tanh\left(Z_c\left[G\left(N,X_{t_g}\right),\left(r_{t_g}g_{t_g-1}\right)\right]+c_c\right),\\
&g_{t_g} = v_{t_g}g_{t_g-1}+\left(1-v_{t_g}\right)d_{t_g}.
\end{aligned}
\tag{8}
$$

Where: G (·) represents the graph convolution process. $\widehat{N}$ is the normalized Laplacian matrix, where $\widehat{M} = \widehat{N} = \widetilde{H}^{-\frac{1}{2}}\widetilde{N}\widetilde{H}^{-\frac{1}{2}},\}$ and $\widehat{N}$ is the matrix indicating the number of edges connected to the node in the graph.

X_0 and X_1 are weight matrices. g_{t_g} and $d_{t_g}$are the output and memory information stored at the time t_g.

r_{t_g} and $v_{t_g}$are the reset and update gate results at the time t_g. σ, ReLU and tanh are activation functions.

3. DLPM Mechanism

3.1. Problem Description

Many methods for protecting trajectory data in the connected vehicle often overlook the spatiotemporal features of trajectories. Geographic constraints and the correlation of locations over time make it highly likely for attackers to infer users' real sensitive locations and trajectory information. In fact, most trajectory privacy protection mechanisms only consider the privacy protection of individual location points, neglecting the impact of consecutive location points on trajectory privacy protection. This oversight makes it easy for attackers to infer the geographical relationship between two location points, deducing the locations where users have passed through or stayed, leading to potential privacy leaks in user locations or trajectories. Positions in trajectories are time-dependent, and introducing timestamps is essential for obtaining the distribution of trajectory positions at different times, exploring the correlation between positions, and understanding user behaviour patterns. However, introducing timestamps leads to sparse trajectory data, making it challenging to withstand injected noise, thereby reducing data value. The ISMMPC clustering is introduced to alleviate the data sparsity caused by timestamps. While clustering trajectory data, ISMMPC retains both temporal and spatial features. Privacy budget pre-allocation is based on the density of clustered trajectory regions, forming a privacy budget matrix. The T-GCN model predicts the privacy budget matrix. During the training process of the T-GCN model, continuous optimization of privacy budget allocation is performed. This optimization protects the privacy of trajectory data while minimizing noise injection into the data.

3.2. Implementation of DLPM

3.2.1. Trajectory Clustering

The ISMMPC algorithm [34] performs clustering in regions after timestamp division, capturing the position distribution of trajectories. Based on the distribution of trajectories [35], it forms arbitrarily shaped and quantified sub-clusters without the need to pre-determine the number of sub-clusters [36]. Figure 2 illustrates the ISMMPC clustering process, where n represents the number of sub-clusters. The membership in the algorithm is a clustering function R(δ;τ) [37], which takes a dataset δ and a similarity function τ as inputs, returning a logical matrix $\chi \in S^{n\times n}$describing membership. Here, the element z_{ij} =1 indicates that points x_i and x_j belong to the same cluster [38]. Assuming the similarity between x_i and x_j is $w_\xi = \xi\left(x_i, x_j\right)$ a clustering

threshold θ= [θ min, θ max] is set, where $\theta = I_\theta \subseteq \left[minw_\xi, maxw_\xi\right]$ and I_θ represents the range of the clustering threshold θ. If θ exceeds the range of R such that $w_\xi \geq \theta$, then z_{ij} =1. This clustering function can have a property of consistency, measuring the stability of the clusters. A hypothesis based on the range of θ suggests that, based on the similarity of δ and τ, when changing θ, reasonable clustering should have relatively stable membership. Therefore, θ is set as:

$$I_\theta = \text{mean}(I_\theta^*), I_\theta^* \to \left(\underset{z_{ij}=0}{max} w_\xi, \underset{z_{ij}=1}{min} w_\xi\right)$$

Where I_θ represents the optimal threshold subinterval.

The ISMMPC algorithm utilizes density peak clustering technology [39] to obtain the density peak set of the region B_g, selecting the highest density peak as the centroid and allocating the surrounding unallocated data points to the same cluster, forming sub-clusters $\widehat{D}$. After allocating all data points, the connectivity of boundary-linked sub-clusters $\widehat{D}$ is evaluated. Among the boundary points across clusters, those not linked and their closest unlinked boundary points are associated, forming boundary links $c_{li} = \{\tau_i, \tau_j\}$ to assess intracohesion within the clusters. Each boundary point is assigned a specific value $\gamma(\gamma \in [0,1.0])$to quantitatively assess the adjacency between sub-clusters. Well-linked sub-clusters with high similarity values will exhibit multiple high-value boundary links. The Mahalanobis distance assesses the similarity between sub-clusters, where a smaller distance indicates higher similarity and can be expressed as:

$$E_{\text{ma}}(\widehat{D}_i, \widehat{D}_j) = \sqrt{(\widehat{D}_i - \widehat{D}_j)^{\text{T}} Z_\gamma{}^{-1} (\widehat{D}_i - \widehat{D}_j)} \quad (9)$$

Where Z represents the weight of eigenvalues.

Using n to represent the total number of boundary links, $\Gamma(E_{\text{ma}})$ returns all E_{ma} values relative to the maximum E_{ma} value, which can be represented as:

$$\Gamma(E_{\text{ma}}) = 1 - \frac{n_\gamma^{-1} \sum_{i=1}^{n_\gamma} \left|E_{\text{ma}_i} - max\left(E_{\text{ma}}(\widehat{\boldsymbol{D}}_i, \widehat{\boldsymbol{D}}_j)\right)\right|}{max\left(E_{\text{ma}}(\widehat{\boldsymbol{D}}_i, \widehat{\boldsymbol{D}}_j)\right)} \quad (10)$$

If the specific values of all boundary link samples are nearly equal for two intersecting sub-clusters, they are highly similar. According to Equation (11), the elements in the similarity matrix (x_i, xj) are similarity values $\widehat{\xi(D_i, D_j)}$. After obtaining the similarity matrix $\xi_{\text{m}} \in S^{n \times n}$ between sub-clusters, sub-clusters are merged into the final cluster D based on the similarity between the sub-clusters.

$$\xi(\widehat{D}_i, \widehat{D}_j) = max\left(D_{\text{ma}}(\widehat{D}_i, \widehat{D}_j)\right) \times \Gamma\left(D_{\text{ma}}(\widehat{D}_i, \widehat{D}_j)\right) \quad (11)$$

Algorithm 1: ISMMPC Clustering Algorithm Implementation

Input: Trajectory data set $\{X_{u_g+1}, \cdots, X_{u_g+U}\}$

Output: Clustered result set D

1. Obtain set $\left\{X_{u_g+1}, \cdots, X_{u_g+U}\right\}$and set $k = \left\lceil\sqrt{u_g + U}\right\rceil$
2. $I_k = \{1,2, \cdots, k\}$, gap $= \left\lceil\frac{\text{range}(I_k)}{20}\right\rceil$
3. for $k = minI_k$: gap: $maxI_k$with step gap
4. Compute the density of Z
5. end for
6. for x_i in $\{X_{u_g+1}, \cdots, X_{u_g+U}\}$
7. $Q_k = max\rho, \gamma_i = 1 \parallel Q_k$is the density peak
8. end for
9. if $Z_j > Z_i$, then measure the coherence between sub-clusters
10. $\gamma_i \leftarrow \gamma_j \rho_i / \rho_j$
11. end if
12. Form sub-clusters $\widehat{\boldsymbol{D}}$ with points having the same Z
13. for each pair of density peaks $\widehat{\boldsymbol{D}}_i, \widehat{\boldsymbol{D}}_j$in $\widehat{\boldsymbol{D}}$
14. Compute D_{ma} according to Equation (9)
15. Calculate similarity $(\widehat{\boldsymbol{D}}_i, \widehat{\boldsymbol{D}}_j)$
16. end for
17. $I_\theta \leftarrow min\left(\xi(\widehat{\boldsymbol{D}}_i, \widehat{\boldsymbol{D}}_j)\right), max\left(\xi(\widehat{\boldsymbol{D}}_i, \widehat{\boldsymbol{D}}_j)\right)\Big] \subseteq [0,1.0]$
18. for $\theta = minI_\theta$: gap: $maxI_\theta \parallel$ gap is the iteration step size
19. Count the number of clusters: count
20. end for
21. Automatically adjust $\theta = \text{mean}\ (I_\theta^*)$
22. Adaptive merging of subclusters $D \leftarrow \widehat{D}$
23. Return clustering result $D = \{D_1, D_2, \cdots, D_{\text{count}}\}$

The first line is for obtaining the dataset. Lines 2-8 involve selecting k neighbouring points for each data point, with a time complexity of $O(n\tilde{k})$, where ˜k represents the average k nearest high-density points. The density peak and sub-cluster count for each region are obtained with a time complexity of $O(n)$. Lines 9-11 calculate boundary linking values to quantitatively assess the correlation between sub-clusters, with a complexity of $O(nk_{\text{b}})$, where $k_{\text{b}} = min\{k/2, 2\ln n\}$. Lines 12-16 compute Mahalanobis distance and estimate the similarity between sub-clusters with an O(n2) time complexity. Lines 17-23 involve adaptive merging of sub-clusters to obtain the clustering result set C, with a time complexity of $O(n^2)$. The total time complexity of ISMMPC clustering is $O\left(n\left(\tilde{k} + k_{\text{b}}\right) + n^2\right)$.

3.2.2. Prediction of Temporal Graph Convolution Privacy Budget Matrix

Figure 3 shows the structure of the T-GCN model. The T-GCN model can learn spatial features from traffic data. The GRU takes h_{u_g-1} and current traffic information as input to obtain the traffic information at the time. The model captures current traffic information while preserving the trend of historical traffic information.

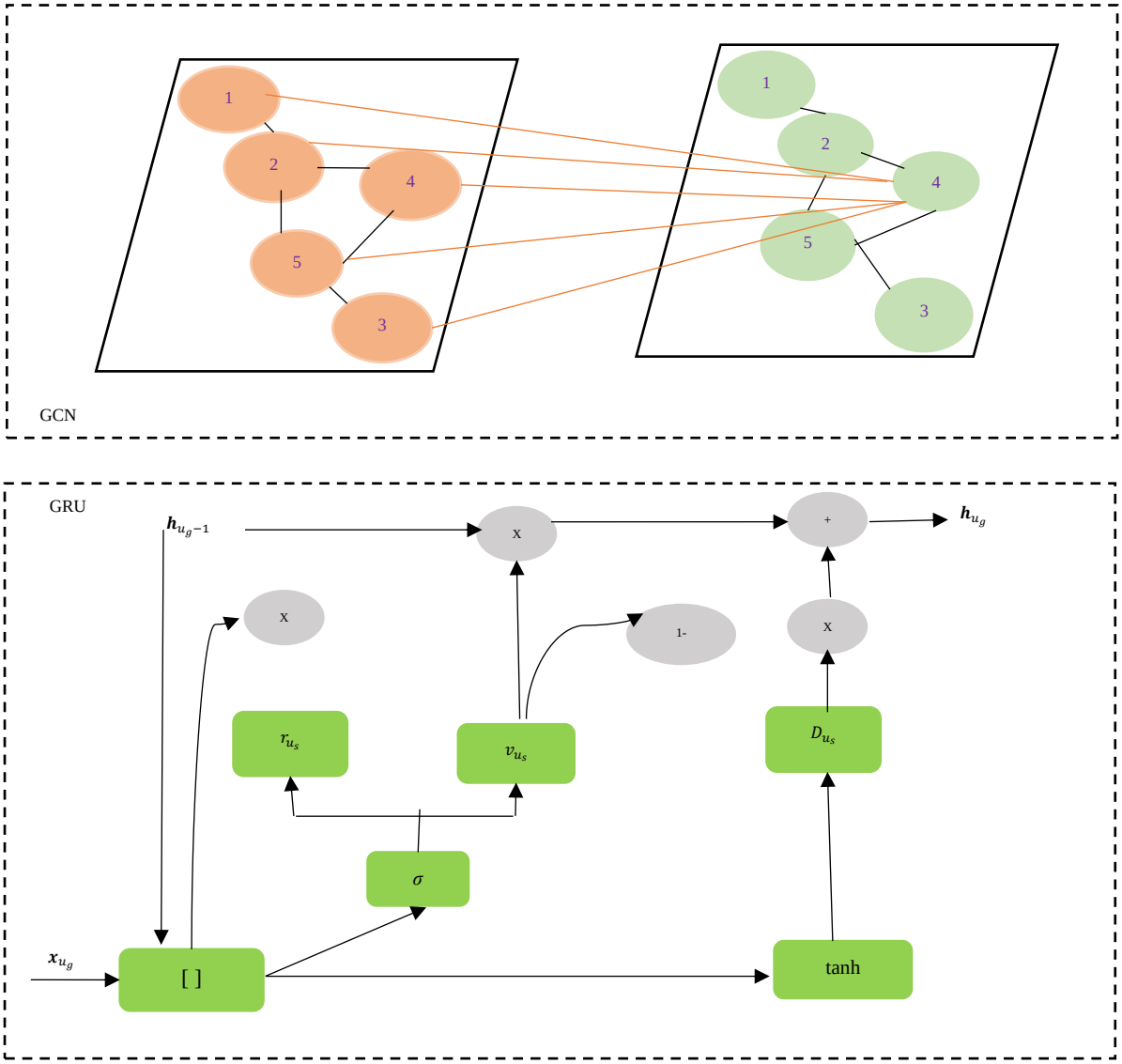


Fig. 3 The compositional structure of the temporal graph convolution network model

The GCN model encapsulates the topological interactions between node 5 and surrounding trajectory points, representing the road network's spatial dependencies and trajectory properties. After graph convolution, GRU extracts temporal characteristics from the node feature matrix. The trajectory data for each region, obtained by timestamp division, is input into the T-GCN model.

Under the influence of spatial dependencies and temporal features, the spatiotemporal information [X_{ug+1}, ..., X_{ug+U}] of the regional trajectory is obtained from Definitions 1 and 2. This information is used as input for ISMMPC clustering, forming the trajectory dataset D. Through D, the total matrix set $\{T_{ij} \mid i = 1, \cdots, x; j = 1, \cdots, y\}$ for each region is obtained, and according to Definition 4, the local density of each region is calculated. The regions are then sorted from top to bottom based on density, resulting in the corresponding density matrix set $\{Q_{ij} \mid i = 1, \cdots, x, j = 1, .., y\}$. Privacy budget pre-allocation is performed according to Equation (12), yielding the privacy budget matrix set $\{F_{ij} \mid i = 1, \cdots, x, j = 1, .., y\}$.

The notion of differential privacy suggests crowded places have lower values and larger values for sparse regions. The privacy budget allocation formula is given by

$$F_{ij} = \frac{\alpha_{ij}}{\sum_{i=1}^{x} \sum_{j=1}^{y} \alpha_{ij}} F \tag{12}$$

In the Equation: $\alpha_{ij} = ij(ij + 1)/2$ it represents the total privacy budget. By integrating the T-GCN model to extract temporal and spatial features from the data, the Model estimates privacy budget matrix F_{ij} .

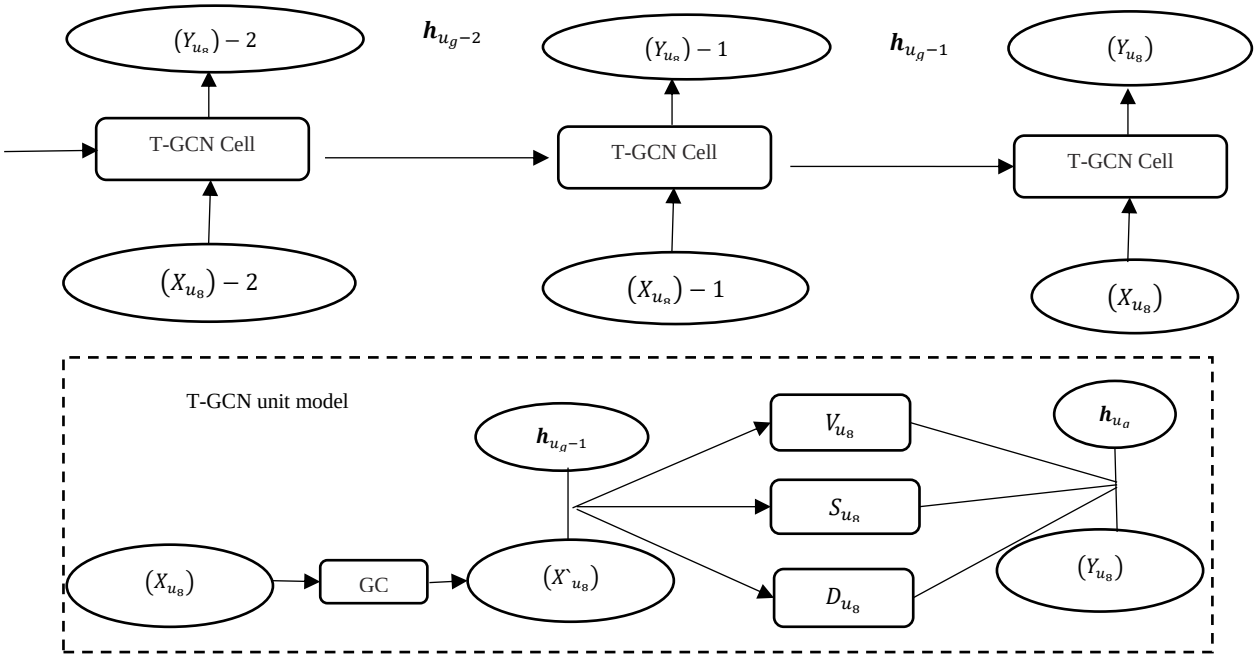


Fig. 4 Model for temporal graph convolutional neural networks: a spatiotemporal prediction procedure

Figure 4 illustrates the spatiotemporal prediction process of Model T-GCN. Starter privacy budget matrix F_{ij} obtained is organized into a spatiotemporal sequence matrix set J_{ij}chronological order. The spatiotemporal data J_{ij} is input into a deep learning model, continuously undergoing the learning and T-GCN unit training to forecast privacy budget matrix $F`_{ij}$ .Based on $F`_{ij}$, Laplace noiseM^*_{ij} is calculated for each total matrix R_{ij} according to Equation (13). The Laplace noise $M^*_{ij}$is then added to the trajectory information of each region, and the perturbed trajectory data is released.

$$M^*_{ij} = \text{Lap}\left(\frac{max}{R}\right), l \in \{1, \cdots, i\}. \quad (13)$$

4. Experimental Setup

Using real datasets, Divvy Bikes and T-drive, for simulation, the study validates the effectiveness and time complexity of DLPM and evaluates the protective effects of differential privacy.

4.1. Dataset

The T-GCN training model utilizes the Adam optimizer and employs the ELU activation function. For the input layer, 80% of the dataset is used for training, while the remaining 20% is reserved for testing. DLPM is compared with PTD [9], LGAN-DP [10], and DPGeo [12]. The Divvy Bikes dataset comprises shared bike usage data in Chicago from 2015 to 2020, including start points, timestamps, start times, and start coordinates. The T-drive dataset covers the trajectories of taxis in Beijing, with a total distance of approximately 9 million km and over 15 million location points. Trajectory data includes each taxi's ID, timestamp, longitude, and latitude. In experimental pre-processing, regions with relatively high trajectory density are selected.

4.2. Evaluation Metrics

Privacy protection aims to release useful information while concealing sensitive data. Three metrics are employed to quantify differences between original and released data. Root Mean Square Error (RMSE): RMSE evaluates the data effectiveness of DLPM, measuring the differences between original and released data. Assuming the true and predicted values of the privacy budget matrix are represented as E ij and E ij′ respectively, with a sample size N, the RMSE formula is given by Equation 14.

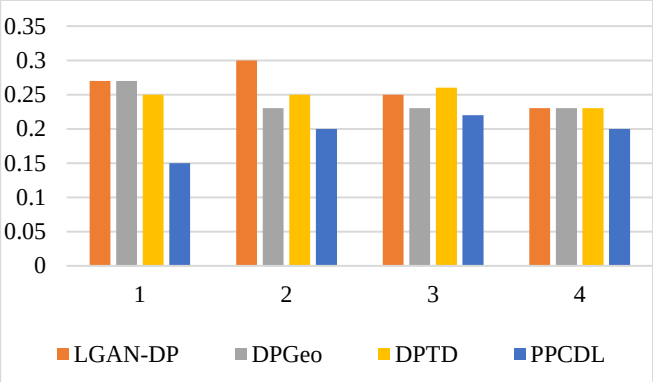
$$\text{RMSE} = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(F_m - \hat{F}_m\right)^2} \quad (14)$$

Query Error (QE): QE is employed to assess the protective effects of differential privacy. Given a query function f(B) for a query area B, where |B| is its size and $f(\widetilde{B})$ represents the noisy query result. The query error is described by Equation 15

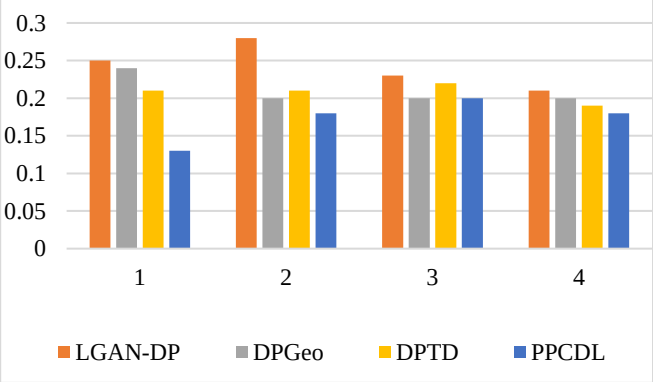
$$\text{QE} = \frac{|f(B) - f(\tilde{B})|}{max\{f(B), 0.01|B|\}} \quad (15)$$

Using Jensen-Shannon Divergence (JS), the similarity between real trajectories and trajectories with added noise is evaluated. Given the probability distribution functions ϕ and ω for the published original data and the noisy data, respectively, JS Divergence is defined as:

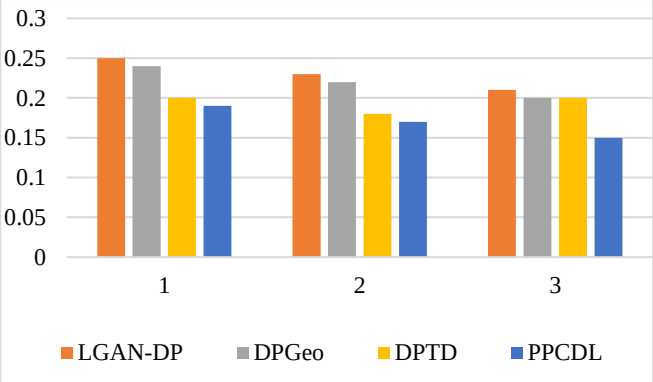
$$\begin{aligned} \text{JS}(\phi \parallel \varpi) = {} & \frac{1}{2}\sum_{i=1}^{m} \varphi_i \ln \frac{\varphi_i}{\varphi_i + \omega_i} + \\ & \frac{1}{2}\sum_{i=1}^{m} \omega_i \ln \frac{\omega_i}{\varphi_i + \omega_i} + \ln 2 \end{aligned} \quad (16)$$



(a) Root means square error



(b) Query error



(c) JS divergence

Fig. 5 Various indicators on the divvy bikes data set

5. Results Analysis and Discussion

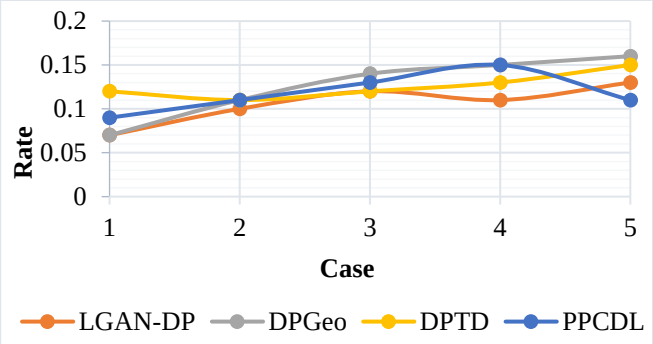
To validate the privacy protection effectiveness of the trajectory dataset, T-GCN forecasts various total privacy budgets' privacy budget matrices ε= {0.1,0.3,0.5,0.7,0.9}. Queries are conducted on both the original and noise-added trajectory data, obtaining RMSE error, QE error, and JS Divergence for testing data. By adjusting ε, the protective level of the dataset under different total privacy budgets is evaluated, and the experimental results are shown in Figures 5 and 6.

Figure 5 (a): DLPM's RMSE is observed to be smaller than the other three mechanisms. As the privacy budget increases, RMSE gradually decreases. Using spatiotemporal features through the TGCN algorithm allows DLPM to continuously predict the privacy budget matrix, providing a more reasonable allocation in trajectory regions that balance noise errors and data availability.

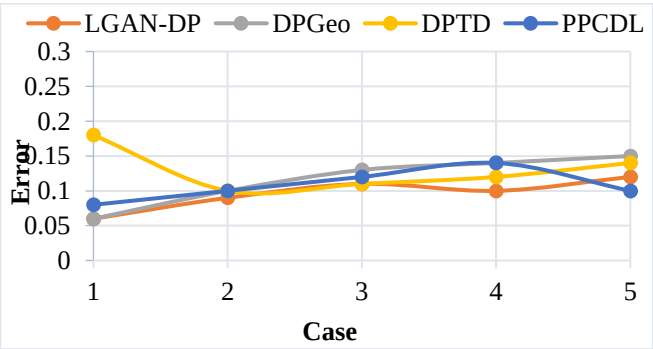
Figure 5 (b): As ε grows, the supplementary Laplace noise diminishes, decreasing QE. The iterative procedure facilitates a more rational budget allocation in trajectory regions, reaching an equilibrium between noise error and data availability.

Figure 5(c): illustrates that the JS divergence between the two datasets diminishes as ε grows, indicating a greater similarity between the probability distributions of the original and noisy data. This occurs because as ε increases, the supplementary noise progressively diminishes, enhancing the similarity between trajectories.

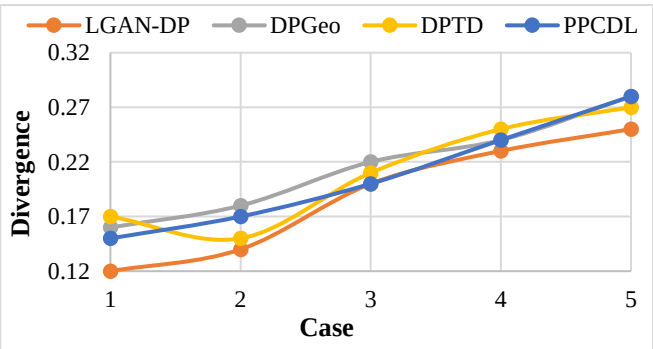
A reduced JS divergence guarantees enhanced privacy by implementing greater perturbations to the actual locations. Conversely, an increased JS divergence results in diminished noise at the actual sites, hence compromising privacy. The JS divergence results demonstrate that the data utility of PPCDL surpasses that of the comparison technique. Fine-grained experiments are conducted on the T-drive dataset. Five timestamps with varying intervals are established on the T-drive dataset, maintaining an identical privacy budget. By incrementally extending the timestamp duration and retrieving trajectory data for the day, the average performance of the metrics is derived. Figure 6 illustrates that altering the timestamp to evaluate the effect of partitioning on privacy protection reveals that PPCDL surpasses the comparable technique in RMSE error, QE error, and JS divergence. This is due to PPCDL's judicious allocation of the privacy budget. Laplace noise is more pronounced in densely populated areas than in sparse ones. Incorporating noise at each point effectively provides varying degrees of privacy protection while improving the accessibility of trajectory data. As the timestamp progresses, the trajectory sequence lengthens, and the RMSE error exhibits an upward trend, resulting in a continual increase in noise that impacts data availability. As the timestamp progresses, RMSE diminishes because the duration of the covered trajectory sequence becomes significantly greater than during other time intervals. The data is currently compromised by noise from non-trajectory regions. As the average metric is computed, the RMSE diminishes. Querying locations in crowded areas for QE error is comparatively straightforward but introduces a greater degree of noise. In the computation of QE error, an increase in noise correlates with a rise in QE. The disparity between the original data and the published data increases JS divergence as noise levels rise. A reduced JS divergence signifies that the area encompasses many places that do not conform to the trajectory sequence.



(a) Root means square error



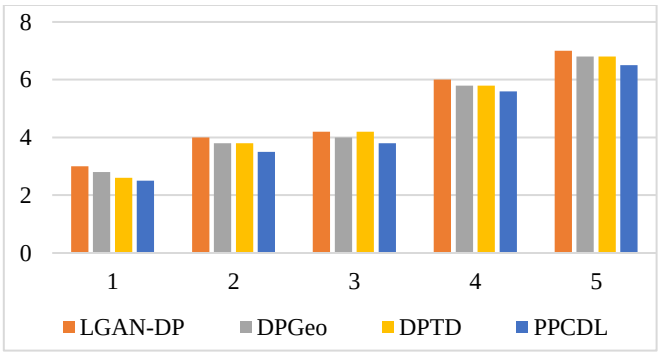
(b) Query error



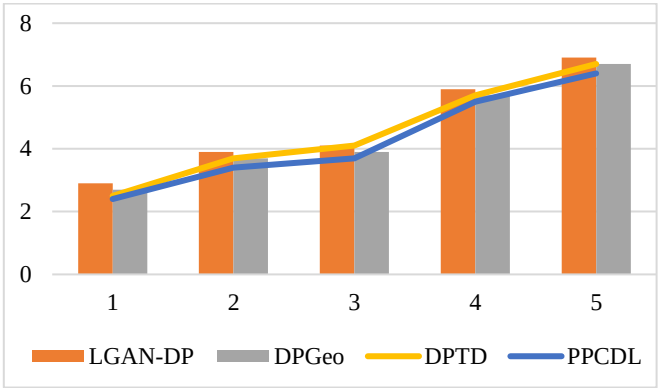
(c) JS divergence

Fig. 6 Fine-grained analysis of time division

To validate DLPM's efficiency, the trajectories in the dataset are divided into different numbers of groups. In the case of ε=0.1, DLPM is compared with other mechanisms for time complexity. The results in Figure 7 show that as the number of participant trajectory data increases, the runtime also increases. Larger group numbers imply a more complex clustering process, requiring more time. Therefore, the average trajectory generation time increases with the number of groups. DLPM's runtime is shorter than the compared mechanisms, indicating faster result generation.



(a) Divvy bikes



(b) T-drive

Fig. 7 Runtime complexity of different schemes

6. Conclusion

The efficiency of the disclosed trajectory dataset and the extent of the utilized privacy budget render DLPM superior to alternative methods. The constancy of DLPM's privacy budget does not alter this reality. This enhancement can be accomplished without altering the allocated budget for privacy. To thwart attackers from acquiring authentic trajectories, anticipated privacy budgets might use temporal intervals in data dissemination, rendering them a potentially effective security measure, particularly vital in nascent technologies like the Internet of Things.

They can use the versatility of the data releases, which explains this situation. All these measures are implemented to avert analogous acquisitions from occurring. Our forthcoming efforts will focus on enhancing the initialization of privacy budgets to augment the training efficiency of deep learning models while adequately safeguarding the privacy of trajectory data in vehicular networks. This will enable us to achieve both objectives. Our capacity to attain both objectives is directly contingent upon this. The outcome is that we will have achieved both of our goals.

References

[1] Olusogo Popoola et al., "A Critical Literature Review of Security and Privacy in Smart Home Healthcare Schemes Adopting IoT & Blockchain: Problems, Challenges and Solutions," *Blockchain: Research and Applications*, vol. 5, no. 2, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] Yichen Wan et al., "Privacy-Preserving Blockchain-Enabled Federated Learning for B5G-Driven Edge Computing," *Computer Networks*, vol. 204, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[3] Ikram Ud Din et al., "Machine Learning in the Internet of Things: Designed Techniques for Smart Cities," *Future Generation Computer Systems*, vol. 100, pp. 826-843, 2019. [CrossRef] [Google Scholar] [Publisher Link]
[4] Praveen Kumar Reddy Maddikunta et al., "Incentive Techniques for the Internet of Things: A survey," *Journal of Network and Computer Applications*, vol. 206, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[5] Aqeel Thamer Jawad, Rihab Maaloul, and Lamia Chaari, "A Comprehensive Survey on 6G and Beyond: Enabling Technologies, Opportunities of Machine Learning and Challenges," *Computer Networks*, vol. 237, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[6] Mubarak Umar et al., "Physical Layer Authentication in the Internet of Vehicles through Multiple Vehicle-Based Physical Attributes Prediction," *Ad Hoc Networks*, vol. 152, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[7] Jie Liu et al., "Internet of Things Challenges and Future Scope for Enhanced Living Environments," *Advances in Computers*, vol. 133, pp. 201-246, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[8] Atefeh Hemmati, Mani Zarei, and Alireza Souri, "UAV-Based Internet of Vehicles: A Systematic Literature Review," *Intelligent Systems with Applications*, vol. 18, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[9] Hamed Alqahtani, and Gulshan Kumar, "Machine Learning for Enhancing Transportation Security: A Comprehensive Analysis of Electric and Flying Vehicle Systems," *Engineering Applications of Artificial Intelligence*, vol. 129, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[10] Jean-Paul A. Yaacoub, Hassan N. Noura, and Benoit Piranda, "The Internet of Modular Robotic Things: Issues, Limitations, Challenges, & Solutions, *Internet of Things*, vol. 23, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[11] Homayun Kabir, Mau-Luen Tham, and Yoong Choon Chang, "Internet of Robotic Things for Mobile Robots: Concepts, Technologies, Challenges, Applications, and Future Directions," *Digital Communications and Networks*, vol. 9, no. 6, pp. 1265-1290, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[12] Nisha, and Urvashi, "A Systematic Literature Review of Internet of Video Things: Trends, Techniques, Datasets, and Framework," *Internet of Things*, vol. 24, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[13] Hongwei Zhang, Wei Fan, and Jinsong Wang, "Bidirectional Utilization of Blockchain and Privacy Computing: Issues, Progress, and Challenges," *Journal of Network and Computer Applications*, vol. 222, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[14] Chenxi Huang et al., "Internet of Medical Things: A Systematic Review," *Neurocomputing*, vol. 557, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[15] Lu Chen et al., "An Approach of Flow Compensation Incentive Based on Q-Learning Strategy for IoT User Privacy Protection," *AEU - International Journal of Electronics and Communications*, vol. 148, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[16] Guowen Wu et al., "Privacy-Preserving Offloading Scheme in Multi-Access Mobile Edge Computing Based on MADRL," *Journal of Parallel and Distributed Computing*, vol. 183, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[17] Mirsaeid Hosseini Shirvani, and Mohammad Masdari, "A Survey Study on Trust-Based Security in Internet of Things: Challenges and Issues," *Internet of Things*, vol. 21, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[18] Abderahman Rejeb et al., "Unleashing the Power of Internet of Things and Blockchain: A Comprehensive Analysis and Future Directions," *Internet of Things and Cyber-Physical Systems*, vol. 4, pp. 1-18, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[19] Heng Li et al., "IoB: Internet-of-Batteries for Electric Vehicles-Architectures, Opportunities, and Challenges, *Green Energy and Intelligent Transportation*, vol. 2, no. 6, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[20] Sabrina Sicari, Alessandra Rizzardi, and Alberto Coen-Porisini, "5G in the Internet of Things Era: An Overview on Security and Privacy Challenges," *Computer Networks*, vol. 179, 2020. [CrossRef] [Google Scholar] [Publisher Link]
[21] Godwin Badu-Marfo et al., "An Ensemble Federated Learning Framework for Privacy-by-Design Mobility Behaviour Inference in Smart Cities," *Sustainable Cities and Society*, vol. 97, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[22] Xiaolan Tang et al., "Assisted Driving System Based on Federated Reinforcement Learning," *Displays*, vol. 80, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[23] Ali Akbar Siddique et al., "Sustainable collaboration: Federated Learning for Environmentally Conscious Forest Fire Classification in Green Internet of Things (IoT)," *Internet of Things*, vol. 25, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[24] Rasha Al-Huthaifi et al., "Federated Learning in Smart Cities: Privacy and Security Survey," *Information Sciences*, vol. 632, pp. 833-857, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[25] Izhar Ahmed Khan et al., "Fed-Inforce-Fusion: A Federated Reinforcement-Based Fusion Model for Security and Privacy Protection of IoMT Networks Against Cyber-Attacks," *Information Fusion*, vol. 101, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[26] Yanling Wang et al., "Differential Privacy in Deep Learning: Privacy and Beyond," *Future Generation Computer Systems*, vol. 148, pp. 408-424, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[27] Jong Wook Kim, and Beakcheol Jang, "Deep Learning-Based Privacy-Preserving Framework for Synthetic Trajectory Generation," *Journal of Network and Computer Applications*, vol. 206, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[28] Yi-rui Huang et al., "GeoPM-DMEIRL: A Deep Inverse Reinforcement Learning Security Trajectory Generation Framework with Serverless Computing," *Future Generation Computer Systems*, vol. 154, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[29] Sujie Shao et al., "Multi Task Dynamic Edge-End Computing Collaboration for Urban Internet of Vehicles," *Computer Networks*, vol. 227, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[30] Raiful Hasan, and Ragib Hasan, "Pedestrian Safety Using the Internet of Things and Sensors: Issues, Challenges, and Open Problems, *Future Generation Computer Systems*, vol. 134, pp. 187-203, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[31] Yassine Himeur et al., "Edge AI for Internet of Energy: Challenges and perspectives," *Internet of Things*, vol. 25, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[32] Prashant Singh et al., "Internet of Things for Sustainable Railway Transportation: Past, Present, and Future," *Cleaner Logistics and Supply Chain*, vol. 4, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[33] Nayan Kumar Subhashis Behera et al., "Futuristic Person Re-Identification over Internet of Biometrics Things (IoBT): Technical Potential versus Practical Reality," *Pattern Recognition Letters*, vol. 151, pp. 163-171, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[34] Shahnawaz Ahmad et al., "Deep Learning Models for Cloud, Edge, Fog, and IoT Computing Paradigms: Survey, Recent Advances, and Future Directions," *Computer Science Review*, vol. 49, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[35] Chen Fang et al., "A Privacy-Preserving and Verifiable Federated Learning Method Based on Blockchain," *Computer Communications*, vol. 186, pp. 1-11, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[36] Francine Berman et al., "The Impact Universe-A Framework for Prioritizing the Public Interest in the Internet of Things," *Patterns*, vol. 3, no. 1, 2022. [Google Scholar] [Publisher Link]

[37] Sofiane Zaidi, Mohammed Atiquzzaman, and Carlos T. Calafate, "Internet of Flying Things (IoFT): A Survey," *Computer Communications*, vol. 165, pp. 53-74, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[38] Yalin Liu et al., "Unmanned Aerial Vehicle for Internet of Everything: Opportunities and Challenges," *Computer Communications*, vol. 155, pp. 66-83, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[39] Tao Chang et al., "PAGroup: Privacy-Aware Grouping Framework for High-Performance Federated Learning," *Journal of Parallel and Distributed Computing*, vol. 175, pp. 37-50, 2023. [CrossRef] [Google Scholar] [Publisher Link]