IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 9 December 2024, accepted 1 January 2025, date of publication 9 January 2025, date of current version 23 January 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3526934

RESEARCH ARTICLE

A Framework for Privacy-Preserving in IoV Using Federated Learning With Differential Privacy

MUHAMMAD ADNAN[1], MADIHA HAIDER SYED[1], ADEEL ANJUM[1], AND SEMEEN REHMAN[2], (Member, IEEE)

[1]Institute of Information Technology, Quaid-i-Azam University, Islamabad 45320, Pakistan
[2]Institute of Parallel Computing Systems, University of Amsterdam, 1098 XH Amsterdam, The Netherlands

Corresponding author: Semeen Rehman (semeen.rehman@tuwien.ac.at)

This work was supported by Technische Universität Wien (TU Wien) Bibliothek through its Open Access Funding Program.

ABSTRACT Vehicles become more advanced and smarter due to advancements in technology in the modern world. Every person now a days, demand a smart vehicle due to their automobility and smart controls. This is all possible through advancements in VANET (Vehicular Adhoc Network) and the Internet of Vehicles (IoV). Vehicles in the VANET are highly connected to each other and this thing can cause security, safety, and privacy risks for the asset itself and driver also. It can become a reason of major threat. And these threats can occur due to tracing the location of the vehicle. Existing techniques like group-based shadowing schemes, obfuscation, silent periods, and mix-zone have preserved privacy of location somehow, but don't have a good QoS and optimized efficient security. To overcome these issues, we introduced a new privacy framework, which is an improvement of the existing shadowing scheme. We proposed a computationally efficient group leader selection process based on centeredness, rule obeyed, and OBU resources, reducing overhead by 20%, integrating FL with DP to preserve data privacy without sacrificing utility, and achieving a 15% improvement in location accuracy under privacy constraints, validating the scalability and robustness of the framework through extensive simulations involving up to 300 vehicles. Group Leader is used as an optimization of the overall framework including efficiency and implementation of the scheme. This scheme increases privacy if the number of vehicles also increases, and this thing makes our scheme more scalable. This scheme overcomes the many drawbacks of existing techniques like a higher tracing ratio in shadowing schemes, totally depending on the group leader, and reduced utility of all schemes based on distances. The most important thing, the single point of failure in the group leader base shadowing scheme is overcome by using local federated learning with differential privacy. Validation results of our proposed scheme showed that it outperformed the current schemes mainly based on group leader.

INDEX TERMS Privacy, federated learning, differential privacy, LBS, data utility.

I. INTRODUCTION

The Intelligent Transportation Systems (ITS) have ushered in a new era of connectivity and efficiency in vehicular communication, epitomized by Vehicle Ad hoc Networks (VANETs) and the Internet of Vehicles (IoV). These paradigms, built upon a foundation of advanced communication technologies, enable seamless data exchange and collaboration among vehicles, infrastructure, and smart devices. Within VANETs, vehicles communicate using a blend of Dedicated Short-Range Communication (DSRC) and cellular networks, facilitating applications such as traffic management and collision avoidance. Similarly, IoV extends the concept of IoT to the automotive domain, fostering a dynamic ecosystem where vehicles interact with each other V2V, roadside infrastructure V2I, and it can also be a hybrid approach [1] as shown in Figure 1. Vehicles with more advanced technology give more convenience to the people inside the car and outside having increased sales globally expected by 18% in 2023. These advancements also raise privacy and security issues [2].

As the vehicles become smarter, with high mobility and autonomous functionality, VANET becomes more vulnerable

The associate editor coordinating the review of this manuscript and approving it for publication was Chakchai So-In.

to outside attacks [1]. There are three main things which are needed to be addressed while designing a security-proof IoV VANET, these are trust, security, and privacy. Many researchers have proposed many techniques to address this issue in different ways. However, the location privacy in VANET needs to be addressed uniquely.

The networks used in VANET are not similar to other types of networks, like Internet networks. Due to the different nature of networks used in VANET, a different approach is needed to avoid issues like malicious attacks and eavesdropping while sharing sensitive information like location data. Without proper protection or encryption of data, it can be led to a serious issue in the VANET system.

Researchers have done a lot of research on VANET as it's the most demanding and trending in the field of intelligent transport systems (ITS). And this is a need of current and future trends also to modernize and make transport smart. Due to the increase in population, vehicles need and demands also increase and due to this, road accidents are also increasing day by day. 1.3 million people die annually having the age of 15-29 as reported by the World Health Organization (WHO) [3]. We can use the latest and smart technology to avoid road accidents in IoV and secure more lives from accidents of vehicles. The latest technology helps to provide smart controls, road health, weather forecasts, traffic jams, and some emergency services. As the need for technology increases in VANET, privacy concerns mainly location privacy also come into consideration because of having more vehicles communicating with each other and sharing information in the network. With the need for technology, the onboard device was also upgraded with high specifications. There are a lot of privacy solutions introduced by many researchers. Some solutions are somehow good or acceptable to be used for LBS applications. In recent years, there's been a lot of interest from researchers in the field of location privacy in-vehicle networks. With advancements in technology, like GPS in cars and mobile devices, it's become easier to track vehicles. Researchers have come up with various solutions to protect privacy in location-based services (LBS). For example, Levy and Schneier [4] proposed a method to confuse attackers by deviating from the usual route. Ullah and colleagues [5] suggested using ''silent periods'', where vehicles stop sending signals, making it harder to track them.

Other methods include group signatures, mix zones, and changing identifiers [6]. However, simply changing IDs might not be enough to stop tracking, as attackers could still figure out a vehicle's home address. There are also safety concerns with some privacy methods, like using fake data. Liang and his team [7] have looked at how capable attackers are of breaching location privacy. This research paper looks at how to protect privacy in the Internet of Vehicles. It explains how using group leaders, federated learning, and differential privacy can keep sensitive information safe while still benefiting from connected vehicles. We have introduced a computationally efficient group leader selection process based on centeredness, rule obeyed, and On-board-unit resources, by reducing overhead by 20%, and by integrating FL with DP to preserve data privacy without sacrificing utility, and achieving a 15% improvement in location accuracy under privacy constraints, validating the scalability and robustness of the framework through extensive simulations involving up to 300 vehicles.

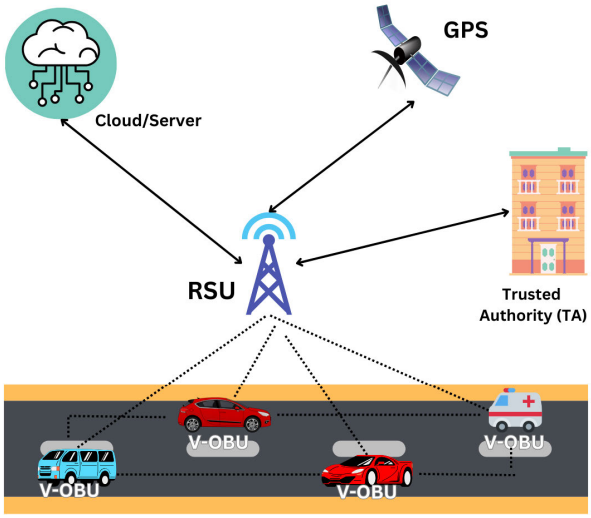


FIGURE 1. A typical VANET system.

The major focus of our study is to maintain privacy while enabling the vehicles to make accurate LBS requests and utilize these requests in real-time scenarios. In Figure 2 we have visualized our proposed system architecture also. Our major contributions to this research are these:

Introducing the group leader with a more optimized selection process. Adding the group leader will optimize the scheme, decrease implementation time, and also utilization of requests in VANET.

Preserving privacy while maintaining the trade-off between utility and location privacy by using Federated Learning with Differential Privacy.

Our proposed model not only maintains privacy, but it also provides scalability of the model of very complex scenarios.

By using our proposed model, vehicles can now ask for LBS (Location Based Services) in the VANET and get accurate responses faster, more timely, and in a more private way.

We have also done a comparison of different privacy-preserving schemes in VANET or IoV and demonstrated the comparison in the form of a table and also shown an analysis of these technique's effects.

We have organized this research paper in the following way: In Section I, we introduce the research problem the objective of our research, and then our methodology. In Section II, we have made related work reviews. In Sections III and IV, we have provided our proposed methods

in more practical detail, And in the next Section, we provide the experiment details of our proposed methodology. Finally, in Section V we concluded our work and also mentioned future works.

II. LITERATURE REVIEW

In the literature review, we have presented different research techniques and directions which are about preserving location privacy in IoV.

A. SILENT PERIODS

Sampigethaya et al. [10] Introduced a scheme known as CARAVAN, which handles pseudonym changes by using group-forming mechanisms and silent-periods intervals. This method suggests using silent periods and forming groups to reduce unnecessary transmissions among vehicles, thus enhancing privacy. However, it assumes scenarios where some vehicles don't send messages for a while, which might not work for safety applications needing frequent broadcasts. Huang et al. [11] propose using silent periods either at specific places or randomly to improve privacy. However, this approach relies on individual decisions made by edge devices (Vehicles), which could lead to synchronization issues. Swing Protocol by Li et al. [12] This protocol focuses on users reducing tracking by changing pseudonyms when speed or direction changes. Attackers find it hard to link movement with positions. The Swap scheme increases location privacy by having vehicles [15] exchange pseudonyms. However, only vehicles that work together and start the change can achieve complete anonymity. CPS Scheme by Wahid et al. [13] This scheme aims to protect vehicle locations. Unlike other plans, CPS says vehicles should only talk when they really need to. They always keep their radios on in case something urgent happens. This plan uses a Roadside Units (RSU) to figure out how far a vehicle can go and how fast it's moving. Then, it sets a timer based on that and the vehicle doesn't send any messages until that time is up.

B. OBFUSCATION

The goal of this technique is to mess with tracking of location. They do this by messing with how accurate the location data is and by making the time between signals from On-Board-Units (OBUs) longer. Takbiri et al. [16] propose using Markov chains, a type of math model, for an information-theory method. They came up with a smart idea to handle mistakes in locations. They use Markov chains to make errors that were already figured out before. This helps when testing Location-Based Services (LBS) because it makes the position less exact and gives a different name.

Mutual Offscale Path (MOP) is a method described by [17] that gets real-time location [14] info without needing users' paths to cross. It uses DSRC radios to mix things up when 2 vehicles chat with the LBS nearby server. MOP needs help from nearby vehicles to mess up the routes for everyone involved. Zhou et al. [18] came up with a way to change how likely it is for each user to be somewhere, without using past user info. Other studies, like the one mentioned by [19], suggest many ways to keep someone's location private.

C. GROUP-BASED AUTHENTICATION

In a study by Lu et al. [22], they suggest a way to keep your location private using a system called the NTRU cryptosystem. They also use something called a post-quantum safe transfer forget protocol. While Location-Based Services (LBS) are really useful, they also make people worried about their privacy. That framework is not hard to put into action [23], but it makes the leader of the group do more work compared to others in the network.

Wahid et al. [13] came up with another way to keep your location private called the Synchronized Pseudonym-Changing Protocol (SPCP). This plan works with groups and makes sure they all change their fake names at the same time. The plan has six parts, like signing up your car and giving it the right settings at the start, joining a group, and swapping fake names at the end. The researchers tested their plan with computers, and they say it works better at keeping your location private compared to other plans like Silent Period, AMOEBA, and Random Encryption Period (REP).

D. MIX-ZONES

Mix-context strategy proposed by Asuquo et al. [20] In which Vehicles change pseudonyms synchronously at certain triggering points. This means cars change their fake names at the same time when they reach certain points. PCS scheme proposed by Ni et al. [21] Short-life keys are generated for vehicles, and pseudonyms are changed at specific locations.

Heuristic pseudonym change method proposed by Guo et al. [9] which maximizes anonymity. They suggested a way to change fake names that make sure cars stay anonymous. Different ideas were put forward to keep where cars are private, like the silent period trick [9], the Mix group plan [3], [24], the Path confusion idea [8], the obfuscation trick [25], and the Shadowing method [26]. It's really important to make a plan that keeps where cars are private without messing up Location-Based Services, makes sure cars stay anonymous, and makes it hard to track them. Also, it's important to keep the records about cars safe and secret.

A study introduced by Li et al. [27] the Fog computing-based Pseudonym Management Program (FPMP). It's like a special system for managing fake names for vehicles. Instead of having one place in charge of giving out these fake names, it moves the job to a layer called the fog. This helps make things easier for the people who give out certificates and makes it quicker for vehicles to get their fake names. To keep things safe, they create an algorithm called the dynamic pseudonym swap program (DPSP).

E. CRITICAL REVIEW

We have made Table 1 to compare different Techniques and Services for Location Privacy. The research gap that we have identified in our proposed paper is that the Group-Leader

TABLE 1. Summary of Location-Based Services (LBS) techniques.

Ref	LBS Technique	Pseudonyms used	Cryptography	Certificate Authority	Limitations
[8]	Path confusion	X	X	X	Disclosure control; Segmentation; Perturbation. Intersection challenges in achieving accurate routes.
[9] [10] [11] [12] [13]	Silent-Period	X	X	✓	CSMA/CD unsuitable for real-time location due to collision delays.
[14]	Tracking-Approach	✓	X	✓	Multi-Hypothesis Tracking ineffective against GPA attacks.
[6]	Endpoint Protection	✓	X	✓	Increased resource consumption and crash risks due to Parrots.
[15] [16] [17] [18] [19]	Obfuscation Scheme	X	X	X	MOP. Location entropy.
[3] [8] [9] [20] [21]	Mix-Zone	✓	✓	✓	Pseudonyms Exchange Algorithm relies on excessive assumptions.
[3] [13] [22] [23]	Group-Based Auth	✓	✓	✓	Heavy work load on group leader.

Based Scheme totally depends upon the leader. If the leader fails to work and compromises then the scheme won't be functional. And data sent toward group leaders and servers can be vulnerable to attacks. Different proposed solutions are still not satisfactory enough to compare privacy methods related to safety, privacy, and security levels against an adversary.

TABLE 2. Notations and symbols.

Symbol	Description
D	Dataset
D_i	Local dataset on device i
θ	Model parameters
θ_t	Model parameters at iteration t
θ_i	Model parameters from device i
θ_{global}	Aggregated global model parameters
ϵ	Privacy parameter in Differential Privacy (DP)
Δf	Sensitivity of function f
$\text{Lap}(\mu, b)$	Laplace distribution with mean μ and scale b

III. PRELIMINARIES

The advent of the Internet of Vehicles (IoV) has revolutionized transportation systems, enabling seamless connectivity and communication among vehicles and infrastructure. IoV facilitates real-time data exchange, leading to enhanced safety, efficiency, and convenience for commuters. However, alongside these advancements come challenges related to privacy and security, particularly concerning the sensitive nature of location-based data.

Background: As the need for technology increases in VANET, privacy concerns mainly location privacy also come into consideration because of having more vehicles communicating with each other and sharing information in the network. With the need for technology, the onboard device was also upgraded with high specifications. Problem Statement: Many researchers have introduced a lot of privacy solutions. Some solutions are somehow good or acceptable for use in LBS applications. With advancements in technology, like GPS in cars and mobile devices, it's become easier to track vehicles.

Research Objective: Our research's main objective is to propose a comprehensive methodology for preserving location privacy in IoV networks. By leveraging optimized group leader selection techniques and federated learning with differential privacy, our proposed scheme aims to maintain the balance between privacy preservation, and data utility while also enabling the vehicles to make accurate LBS requests and utilize these requests in real-time scenarios. We have proposed an efficient location privacy preservation scheme using FL with DP in IoV VANET.

TABLE 3. Abbreviations.

Abbreviation	Description
FL	Federated Learning
DP	Differential Privacy
IoV	Internet of Vehicles
LBS	Location-Based Services
VANET	Vehicular Ad hoc Network

Significance of the Study: By using our proposed model, vehicles can now ask for LBS (Location Based Services) in the VANET and get accurate responses faster, more timely, and in a more private way.

We have also done a comparison of different privacy-preserving schemes in VANET or IoV and demonstrated the comparison in the form of a table and also shown an analysis of these technique's effects.

A. FEDERATED LEARNING

Federated Learning (FL) is a method where data stays on local devices (like phones or vehicles), and only the trained model updates are transferred to the central server. This way, privacy is maintained as the data never leaves the device. In our vehicular network, the global model θ is updated by aggregating the contributions from multiple vehicles while maintaining privacy. The objective is to train a global model without sharing raw data.

Global Model Update:

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{1}{n} \sum_{i=1}^{n} \nabla J_i(\theta_t)$$

where:

- θ_t is the global model at iteration t.
- η is the learning rate.
- $J_i(\theta_t)$ is the loss function of the i-th vehicle's model.
- $\nabla J_i(\theta_t)$ is the gradient of the loss function for vehicle i.

Local Model Update (Per Vehicle):

$$\theta_i^{t+1} = \theta_i^t - \eta \cdot \nabla J_i(\theta_i^t)$$

where:

- θ_i^t is the local model for vehicle i at iteration t.
- Each vehicle updates its local model independently using its own data.

B. DIFFERENTIAL PRIVACY

DP is a technique that adds random noise to data to protect individual privacy. By using the Laplace mechanism, we ensure that the output does not reveal specific details about any individual in the dataset. To preserve privacy during model updates and location-based service (LBS) requests, differential privacy is applied using the Laplace mechanism.

Differential Privacy Applied to Model Updates:

$$\tilde{\theta}_i^{t+1} = \theta_i^{t+1} + \text{Lap}\left(\frac{\Delta\theta_i}{\epsilon}\right)$$

where:

- $\tilde{\theta}_i^{t+1}$ is the differentially private model update for vehicle i.
- $\Delta\theta_i$ is the sensitivity of the model update.
- ϵ is the privacy parameter controlling the amount of added noise.

Differential Privacy Applied to LBS Requests:

$$\tilde{R}_i = R_i + \text{Lap}\left(\frac{\Delta R_i}{\epsilon}\right)$$

where:

- $\tilde{R}_i$ is the differentially private LBS request for vehicle i.
- R_i is the original request.
- Adding Laplace noise ensures the privacy of sensitive location data.

IV. PROPOSED METHODOLOGY

In the Proposed Methodology section, we will discuss the two major parts of our proposed scheme: An optimized group leader selection process, and the implementation of federated learning with a differential privacy process. To ensure both privacy and scalability, we integrate Federated Learning (FL) with Differential Privacy (DP) in the IoV context. Our methodology comprises two major components: decentralized training and noise injection for privacy preservation.

First, we describe the completely optimized process of selection of the group leader and then after that introduce the Federated Learning with the addition of differential privacy in our proposed scheme. A good understanding of two main phases is necessary for complete deployment, and implementation of our introduced scheme in IoV and Vehicular Adhoc network (VANET). In Figure 3 we have visualized major components of our proposed architecture.

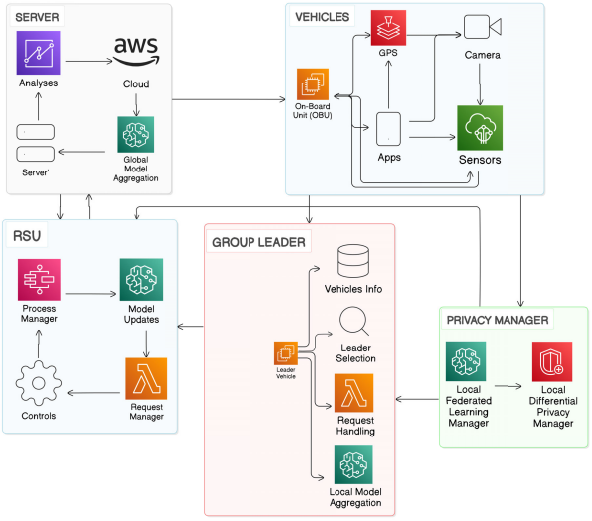


FIGURE 2. Proposed system architecture.

A. SELECTION OF GROUP LEADER

The inclusion of a group leader in the framework makes it somehow complex, causes extra costs of maintenance, and deployment, but also helps in the optimization of the overall framework. As IoV has different types of applications usage of a group-leader-based shadowing scheme may not be suitable for every usage in IoV systems. It also depends upon the suitability of usage in IoV applications. In the previous group leader-based shadowing scheme, if vehicles increase in the network, then that scheme will become less efficient. And that the previous scheme totally depends upon the group leader. If the group leader fails or is compromised, then the whole scheme group will not work as described. Because the group leader is the main central point the scheme didn't work without group leader inclusion as the privacy and security of the VANET were also compromised.

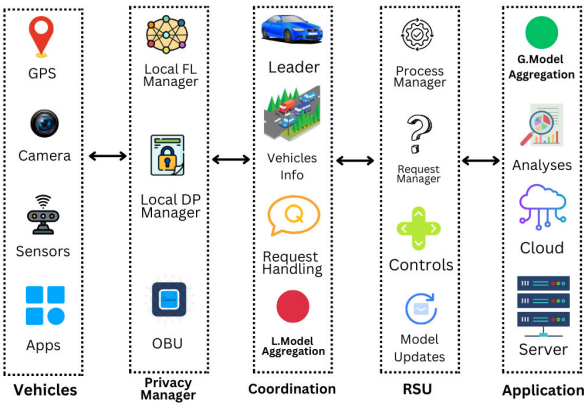


FIGURE 3. Components of proposed architecture.

In Figure 4 the process of group leader selection, we have made some optimizations with the comparison of the previous group leader-based shadowing scheme. The previous scheme used a very complex method for group leader selection and

it took some time also because it needed more calculations to do in the selection process. Previous shadowing-based scheme uses Direct trust, indirect trust, centeredness, and OBU resources in the process of group leader selection. We have optimized this to only take these three parameters; the rules obeyed, centeredness, and OBU resource features of vehicles for the leader selection process. It makes our scheme group leader selection simpler, and optimized.

In our approach, we simplified the group leader selection criteria by focusing on centeredness, rules obeyed, and OBU resources. This was done to reduce the computational complexity and improve efficiency. The previous method relied on direct and indirect trust calculations, which were time-consuming and prone to manipulation. By using straightforward and measurable parameters, our method is faster, more robust, and better suited to the dynamic nature of vehicular networks. Table 5 shows the comparison of the previous and our proposed method of group leader selection.

Algorithm 1 Process of Group Leader Selection

Start
while For a time T_m, collect all the messages from vehicles V_i **do**
  Initialize $SelGL_i = C_i + RO_i + OR_i$ for each vehicle V_i
  Broadcast $SelGL_i$ to all other vehicles
  Receive $SelGL_j$ from all other vehicles V_j
  for each vehicle V_i **do**
    if received message contains a leader announcement **then**
      Compare $SelGL_i$ with $SelGL_j$ for all $j \neq i$
      if $SelGL_j > SelGL_i$ **then**
        Assign $GL_i = V_j$ (vehicle V_j as the new leader)
      else
        Maintain current leader $GL_i = V_i$
      end if
    end if
  end for
  Consensus step: Each vehicle V_i broadcasts its selected leader GL_i
  Final leader V_{GL} is determined as:

$$V_{GL} = \arg\max_{V_i \in V} SelGL_i$$

  Announce the vehicle with the highest $SelGL$ score as the group leader
  Repeat the process after a specific interval to ensure dynamic adaptation
end while $= 0$

Making cells is necessary to let every vehicle know about their range and other nearby vehicles. The size of the cell may also be dependent on the range of vehicles. As if the vehicles have more range of transmission power and other related features, then we can make cells wide enough. This can also be calculated by the dedicated short-range communication (DSRC) standard. This process repeats after some time to update the cells of the vehicle they are in, so everyone now communicates with their own cell, not the others one to avoid the communication collision and make the scheme more effective, dynamic, and feasible.

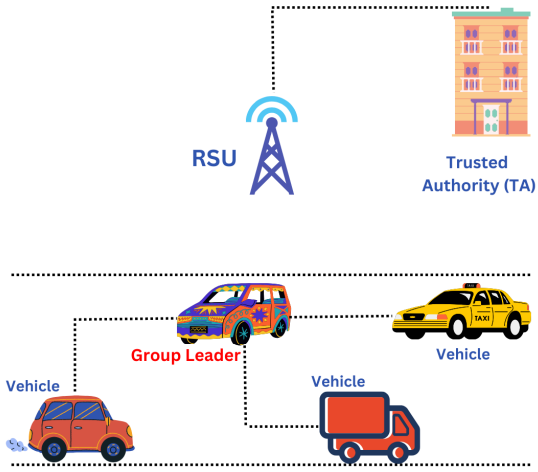


FIGURE 4. Process of group leader selection.

This process starts with broadcasting a message from every vehicle in the cell to other vehicles in the very beginning, as they have to choose a leader. So it's very important to first know each other, then selecting a leader is easy for them. When first each broadcasts the message, then it returns with the other vehicle's information. First, every vehicle assigns itself a leader, but we just want only 1 leader in a cell.

Algorithm 1 shows all the processes for selecting a group leader in a cell. Also, algorithm 1 contains the nodes (vehicles) in the network, and from these nodes, only solo nodes will be selected and take charge of leader activities to perform. For this, as the broadcast messages are received, algorithm 1 starts collecting this response from different nodes. This process only happens in a specific amount of time. As specified in algorithm 1, when broadcast responses are collected and if they contain the leader information, then the algorithm checks all other responses and if didn't find any leader information, it assigns the first leader message received as a leader. And this process repeats, if sometimes a response doesn't contain the leader information, then the response receiving node will assign itself the leader.

In another case, if the leader's responses are more than 1, then it compares all leader's information like we have defined centeredness, rules obeyed and OBU resources, and having more positive numbers node will be selected as a group leader node and the message with leader selected info will be broadcasted so every node can know about the newly elected leader. The reason to do all of this process is to make sure the selection of the group leader node will be on a more positive number and the selection of a solo node as a group leader.

TABLE 4. Critical parameters of the proposed method.

Parameter	Description	Explanation/Values
C_i (Centeredness)	Measures how centrally a vehicle is located in its respective cell.	$C_i = \sqrt{x_i^2 + y_i^2} - r$, where (x_i, y_i) is the vehicle's position, and r is the reference distance from the center of the cell. Lower values indicate better centrality.
RO_i (Rules Obeyed)	Represents the compliance of a vehicle with traffic rules.	$RO_i = \sum_{j=1}^{m} W_j R_{ij}$, where W_j is the weight of rule j, and $R_{ij} = 1$ if the rule is followed, otherwise $R_{ij} = 0$. Higher values indicate better compliance.
OR_i (OBU Resources)	Denotes the computational and storage resources of the vehicle's On-Board Unit (OBU).	$OR_i = P_{\text{comp},i} + C_{\text{storage},i}$, where $P_{\text{comp},i}$ is the computational power, and $C_{\text{storage},i}$ is the storage capacity of the vehicle's OBU.
$SelGL_i$ (Selection Score)	Combined score for selecting the group leader.	$SelGL_i = C_i + RO_i + OR_i$. The vehicle with the highest score is selected as the group leader.
ϵ (Privacy Parameter)	Controls the amount of noise added for differential privacy.	Lower ϵ provides higher privacy but may reduce utility. Used in the Laplace mechanism for model updates and LBS requests.
Δf (Sensitivity)	The maximum change in a function's output when a single data point in the input dataset changes.	Scales the noise in the Laplace mechanism: Noise $\sim$ Lap$(\Delta f/\epsilon)$. Depends on the specific function being used.
D (Dataset)	Represents the local dataset available on each vehicle for federated learning.	Denoted as D_i for each vehicle i.
θ (Model Parameters)	Parameters of the global model trained via federated learning.	θ_t represents the model at iteration t, and θ_{global} represents the aggregated model after updates from all vehicles.
AS_A (Anonymity Set Size)	Set of locations indistinguishable from a vehicle's true location due to privacy-preserving mechanisms.	Size increases with travel time or traffic density. Larger sizes indicate better privacy.
H (Entropy)	Measures uncertainty in a vehicle's location within the anonymity set.	$H = -\sum_{i=1}^{\lvert N\rvert} p_i \log_2(p_i)$, where p_i is the probability of a vehicle being the target. Higher values indicate stronger privacy protection.
P_A (Tracking Success Ratio)	Probability that a vehicle's anonymity set size equals 1 (uniquely tracked).	$P_A = \Pr(\lvert\text{AS}\rvert = 1)$. Lower values indicate better privacy.

TABLE 5. Comparison of previous method and proposed method for group leader selection.

Metric	Previous Method	Proposed Method
Trust Calculation Delay	High	Low
Susceptibility to Attacks	High (trust manipulation)	Low (verifiable criteria)
Adaptability	Low	High

Selecting a group leader in a cell of nodes (vehicles) is very important to do the tasks of a cell more effectively and collaboratively. Figure 5 shows how the group leader is selected in a cell. This way, the performance of a cell is increased in a VANET as every node knows who to communicate with for which purpose. As demonstrated in Algorithm 1, with respective of time, the comparison of the currently selected group leader node in a cell will be made to other nodes as well, if the best leader node then the current leader node will be replaced with the other node having more centeredness, rules obeyed and OBU resources and message with the new leader selection information broadcasted in the cell so everyone know the new leader. This process of checking the leader repeats after a certain period of time. Only vehicles with more power and they also consider them to be a leader will only broadcast the message in a cell to become a group leader.

In a vehicular network, from a set of vehicles within a specific area or cell, the vehicle with the maximum score based on centeredness, compliance with traffic rules, and onboard unit (OBU) resources is selected as the group leader. Let we prove it how a vehicle is selected as a group leader on the bases of different properties and selection criteria.

Let:

- A set of n vehicles $V = \{V_1, V_2, \ldots, V_n\}$ in a cell.
- i $\in N$, $i = \{1, 2, 3, \ldots, n\}$.

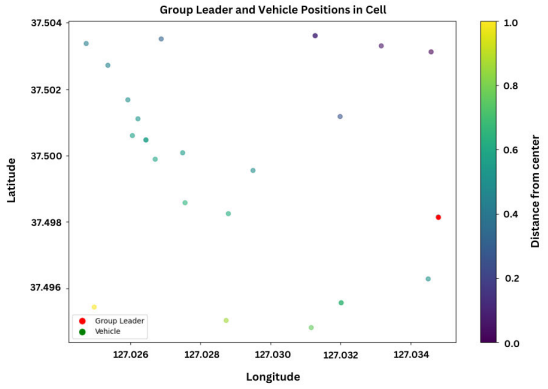


FIGURE 5. Group leader selection in a cell.

- V_i is the attributes given.
- As general form $V_i = \{V_1, V_2, \ldots, V_n\}$
- For each vehicle V_i, its position is (x_i, y_i).
- C_i is the Centeredness of each vehicle V_i, which can be calculated as according to the distance formula:

$$C_i = \sqrt{x_i^2 + y_i^2} - r$$

- Where i $\in$ N
- And, r is the reference distance from the center of the cell.
- RO_i is Rules Obeyed by each vehicle V_i, which can be calculated as:

$$RO_i = \sum_{j=1}^{m} W_j R_{ij}$$

- where W_j is the weight assigned to rule j, $R_{ij} = 1$ if vehicle V_i follows rule j, otherwise $R_{ij} = 0$, and m is the total number of rules.

- OR_i OBU Resources of each vehicle V_i, which can be calculated as:

$$OR_i = P_{\text{comp},i} + C_{\text{storage},i}$$

- where $P_{\text{comp},i}$ is the computational power and $C_{\text{storage},i}$ is the storage capacity of the onboard unit.

Theorem 1: The vehicle V_{leader} with the highest group leader selection score $SelGL_i$ is selected as the group leader, where selection criteria based on this formula:

$$SelGL_i = C_i + RO_i + OR_i$$

Proof: For each vehicle V_i (where $i \in \{1, 2, \ldots, n\}$), the selection score $SelGL_i$ is:

$$SelGL_i = C_i + RO_i + OR_i$$

Here:

- C_i represents centeredness (lower values are better).
- RO_i represents rules obeyed (higher values are better).
- OR_i represents OBU resources (higher values are better).

To select the group leader with the maximum score of selection properties than all other vehicles, we first perform that properties calculation by their formula's.

As we know C_i Centeredness of each vehicle V_i can be calculated as:

$$C_i = \sqrt{x_i^2 + y_i^2} - r$$

And RO_i is Rules Obeyed by each vehicle V_i can be calculated as:

$$RO_i = \sum_{j=1}^{m} W_j R_{ij}$$

And OR_i OBU Resources of each vehicle V_i can be calculated as:

$$OR_i = P_{\text{comp},i} + C_{\text{storage},i}$$

So, as we know every property and how it can be calculated: $SelGL_i$:

$$V_{\text{leader}} = \arg \max_{i \in \{1,2,\ldots,n\}} SelGL_i$$

The vehicle with the highest score has the best balance of:

- Centeredness (ideal positioning),
- Rules obeyed (reliability),
- OBU resources (technological capability).

A high $SelGL_i$ indicates the vehicle is centrally located, follows more rules, and has sufficient resources.

Thus, the vehicle V_{leader} with the maximum $SelGL_i$ is the optimal choice for the group leader.

1) CENTEREDNESS

The centrality of a vehicle within the network can be determined by various factors such as its position, connectivity, and influence. A more centralized vehicle, which has better connectivity with other vehicles and is strategically positioned within the network, is more likely to be selected as a group leader. It is the location of vehicles in the respective cell. The vehicles are divided into cells on the basis of their GPS location and the vehicle will instantly know its cell when the cell division is made. A vehicle that is in the center of the cell and has good reach to other vehicles in the cell will be better for our group leader. The more the location of the vehicle in the center, the more the chances to be selected as group leader. This will be a dynamic because vehicles move with speed and change their locations. We can also check the centeredness of a vehicle by Eq. 1 which is a distance formula:

$$\sqrt{x^2 + y^2 - r} \tag{1}$$

Figure 4 shows how the distance between vehicles is calculated and the vehicle which is most in the center of the cell will be selected as the group leader with other features considered too.

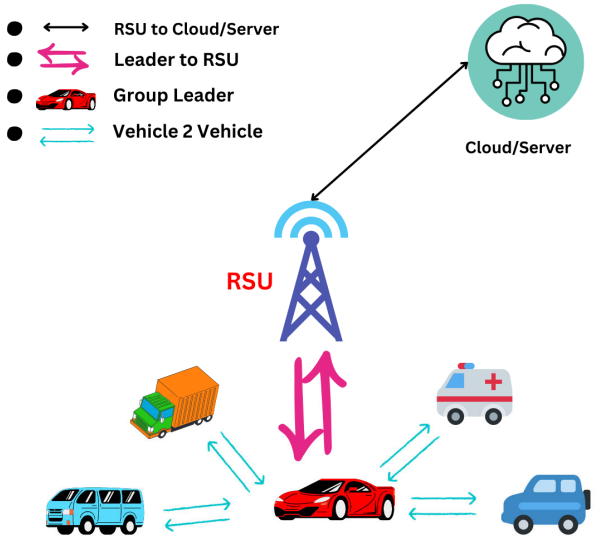


FIGURE 6. Proposed federated learning scheme.

2) RULES OBEYED

A cell has a number of vehicles in it, and all vehicles won't follow all the traffic rules. Let $R_1, R_2, \ldots, R_n$ be the set of traffic rules, and let $W_1, W_2, \ldots, W_n$ be the weights assigned to these rules based on their importance. A vehicle with a greater number of rules followed and positive numbers from others have more chances of being selected as a leader. The Rules Obeyed (RO) score for each vehicle can be calculated as the sum of the weights of the rules it follows.

3) OBU RESOURCES

Onboard unit resources or OBU is a field with a Beacon message transmitted by a vehicle. Let P_{comp} denote the computational power (measured in some units) and C_{storage} denote the storage capacity (measured in some units) of a vehicle's onboard unit. A vehicle with more OBU resources,

represented by higher computational power and storage capacity, has a greater ability to handle requests and perform calculations in less time. We do this resource checkup for the sake of flexibility; if more vehicles are added to the cell, then the group leader with more power can handle it.

$$\text{Dist} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{2}$$

A group leader selection can be done by Eq. 3, incorporating the metrics of centeredness (C), Rules Obeyed (RO), and OBU Resources (OR).

$$\text{Sel}_{\text{GL}} = C + \text{RO} + \text{OR} \tag{3}$$

Eq. 2 shows the formula to calculate the distance of each vehicle in a cell. Finally, the group leader is chosen based on Eq. 3.

B. FEDERATED LEARNING

Introduced by Google in 2017, addresses privacy concerns associated with traditional centralized machine learning approaches, particularly when sensitive data is involved. Unlike conventional methods where training data is centralized, FL distributes the model training process across multiple devices, preserving data privacy while enabling model improvement.

In traditional machine learning, a global model θ is trained using a centralized dataset D through optimization of a loss function $J(\theta)$. This process involves transmitting the entire dataset to a central server, which poses privacy risks and scalability challenges.

However, in federated learning, the training data remains decentralized on individual devices D_i. The goal is to optimize the global model parameters θ without directly accessing raw data. The process can be formulated as follows: **Initialization**: Initialize a global model θ_0 and distribute it to participating devices. **Local Model Training**: Each device i trains the global model using its local dataset D_i by minimizing its local loss function $J_i(\theta)$: $\theta_{i+1} = \arg\min_\theta J_i(\theta)$ **Model Aggregation**: The updated models θ_{i+1} are sent to a central server for aggregation, where the global model θ is updated by aggregating the parameters: $\theta_{\text{global}} = \frac{1}{N}\sum_{i=1}^{N}\theta_i$ **Iterative Improvement**: Steps 2 and 3 are repeated for multiple rounds until convergence or a predefined stopping criterion is met.

FL offers several advantages, including: - **Privacy Preservation**: User data remains on local(edge) devices, that reduce the risk of data breaches. - **Reduced Latency**: Model updates occur locally, minimizing communication overhead and latency. - **Lower Power Consumption**: Devices perform local computations, reducing the need for centralized processing. - **Collaborative Learning**: Multiple devices contribute to model improvement, fostering collaboration and diversity in training data.

In Figure 6 our proposed scheme, we leverage federated learning to enhance the performance and privacy of location-based services (LBS) requests. By decentralizing model training and incorporating techniques such as differential privacy, we ensure both data utility and privacy, making our approach suitable for real-world applications. In Figure 10a, 10b, 10c, and 10d we have demonstrated the FL with different levels of noise addition on model updates using DP.

Algorithm 2 Applying Federated Learning on Model Updates Process

Initialization
Central server initializes the global model θ_0
Distribute θ_0 to all participating devices D_i
while not converged or stopping criterion not met **do**
  for each device D_i **do**
    Receive global model θ_t from the central server
    Update $\theta_i^{t+1} = \arg\min_\theta J_i(\theta)$ using local dataset D_i
    Send updated model θ_i^{t+1} to the central server
  end for
  Central server aggregates the updated models:

$$\theta_{global}^{t+1} = \frac{1}{n}\sum_{i=1}^{n}\theta_i^{t+1}$$

  Distribute the updated global model θ_{global}^{t+1} to all devices
end while $= 0$
End

In a vehicular network, federated learning (FL) is employed to enhance model training while preserving data privacy. Vehicles equipped with onboard units (OBUs) participate in a decentralized training process. The goal is to train a global model θ across multiple devices without centralizing the data, thereby ensuring privacy and efficiency.

Let:

- A set of n devices $D = \{D_1, D_2, \ldots, D_n\}$.
- Each device D_i has its local dataset D_i.
- The objective is to train a global model parameter θ without accessing raw data from individual devices.
- The federated learning (FL) process optimizes the global model θ while preserving data privacy and scalability.

Theorem 2: The global model is trained iteratively using federated learning without compromising data privacy.

Proof:

- The central server initializes the global model θ_0 and distributes it to all devices.
- Each device D_i receives the initial model θ_0 and performs local updates using its dataset D_i:

$$\theta_i^{t+1} = \arg\min_\theta J_i(\theta)$$

  where $J_i(\theta)$ represents the local loss function for device D_i.
- After local training, each device D_i sends its updated model θ_i^{t+1} to the central server. The server aggregates

these updates as:

$$\theta_{\text{global}}^{t+1} = \frac{1}{n}\sum_{i=1}^{n}\theta_i^{t+1}$$

This aggregation allows the global model to be updated based on collective knowledge from all devices.

- The process is repeated by redistributing the updated global model $\theta_{\text{global}}^{t+1}$ to all devices for further local training. This iteration continues until convergence is achieved or a stopping criterion is met.

So, In this way, the global model θ is optimized in a decentralized fashion, which enhances privacy and scalability. Federated learning minimizes the global loss function $J(\theta)$ while preserving the privacy of local datasets.

C. DIFFERENTIAL PRIVACY

Federated learning facilitates decentralized model training by leveraging data stored on edge devices such as computers and phones. The process entails sending models to these devices for local training, ensuring data privacy as the data never leaves the device. Updated model parameters are aggregated by a central server using algorithms like Federated Averaging. As demonstrated in Algorithm 3. While this approach safeguards client data to a certain extent, concerns persist regarding the privacy of the updated parameters transmitted from the device to the central server. To address this, differential privacy can be employed, adding noise to the parameters prior to transmission. This ensures a high level of privacy and security for client information. Figure 7 and Figure 8 show the effect of different levels of differential privacy on accuracy and noise in data.

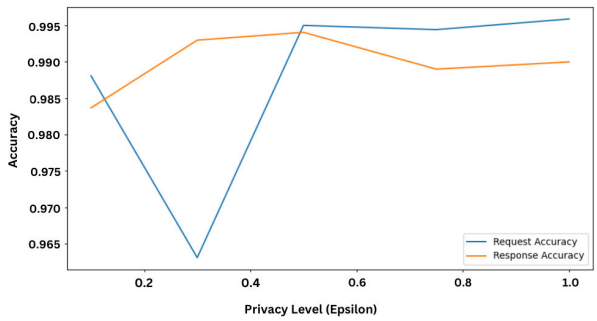


FIGURE 7. Accuracy of LBS request with different levels of DP noise.

DP provides a rigorous framework for quantifying and preserving individual privacy in statistical analysis. One common method to achieve DP is the Laplace mechanism. Let $f(D)$ represent a function computed over a dataset D. The Laplace mechanism adds random noise to the output of $f(D)$ to achieve differential privacy:

$$f(D) + \text{Lap}(\frac{\Delta f}{\epsilon})$$

where Δf denotes the sensitivity of the function f, and ϵ represents the privacy parameter that quantifies the level of privacy protection. The Laplace distribution, denoted by $\text{Lap}(\mu, b)$, is defined by its probability density function:

$$\text{Lap}(x|\mu, b) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right)$$

Another approach involves employing algorithms that group similar data together, sharing summaries of data groups instead of individual data points. Let $\mathcal{G}$ denote the grouping function that maps individual data points to a group. DP makes it sure that the output of $\mathcal{G}$ remains indistinguishable even when individual data points are modified. Formally, for any pair of neighboring datasets D and D' that differ by a single data point, and for any subset S of possible outputs, the following condition holds:

$$\Pr[\mathcal{G}(D) \in S] \leq e^{\epsilon} \cdot \Pr[\mathcal{G}(D') \in S]$$

where ϵ represents the privacy parameter.

Incorporating DP into FL ensures that the aggregation of model updates at the central server does not compromise individual data privacy. By introducing noise or grouping mechanisms, federated learning achieves the delicate balance between privacy and model utility, thereby enabling effective model training while preserving data privacy.

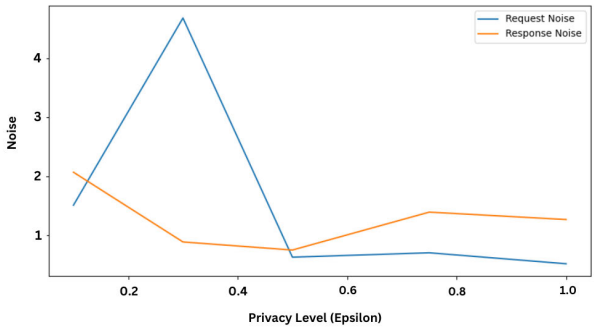


FIGURE 8. Noise effect with different levels of privacy.

In our proposed scheme, differential privacy (DP) is employed to protect the privacy of location-based service (LBS) requests sent from vehicles to the group leader, as well as the responses returned by the group leader. The Laplace mechanism is applied to achieve differential privacy, ensuring that the inclusion or exclusion of any vehicle's data does not significantly affect the overall output.

Let:

- A set of n vehicles $V = \{V_1, V_2, \ldots, V_n\}$.
- Each vehicle V_i has its LBS request data R_i.
- A group leader G processes these requests and returns responses while applying differential privacy.

Theorem 3: Differential privacy can be applied to both LBS requests and responses using the Laplace mechanism to protect individual data points while maintaining the utility of the aggregated information.

Proof:

- We have to prove that for any two neighboring datasets D and D' that differ by a single data point, and for

any subset S of possible outputs, the differential privacy condition will be maintained:

$$\Pr[G(D) \in S] \leq e^{\epsilon} \cdot \Pr[G(D') \in S]$$

- We have set of n vehicles $V = \{V_1, V_2, \ldots, V_n\}$.
- And each vehicle V_i has its LBS request data R_i.
- Each vehicle V_i sends an LBS request R_i to the group leader. The request is processed by a function $f(R_i)$, and Laplace noise is added to the output to ensure differential privacy:

$$\tilde{f}(R_i) = f(R_i) + \text{Lap}\left(\frac{\Delta f}{\epsilon}\right)$$

  where Δf is the sensitivity of f and ϵ is the privacy parameter.
- The group leader processes the differentially private requests and generates a response S, processed by a function $g(S)$. Laplace noise is added to the output of $g(S)$:

$$\tilde{g}(S) = g(S) + \text{Lap}\left(\frac{\Delta g}{\epsilon}\right)$$

  where Δg is the sensitivity of g and ϵ is the privacy parameter.
- As all the requests go to group leader G, which processes these requests and returns responses while applying differential privacy.
- For any two neighboring datasets D and D' that differ by a single data point, and for any subset S of possible outputs, the differential privacy condition is maintained:

$$\Pr[G(D) \in S] \leq e^{\epsilon} \cdot \Pr[G(D') \in S]$$

  This ensures that the outputs for D and D' remain indistinguishable, preserving privacy.

Hence proved that, by applying differential privacy to both LBS requests and responses, the scheme ensures that individual data points are protected while maintaining the utility of the overall data.

Algorithm 3 Differential Privacy for LBS Requests and Responses

Initialization
Set privacy parameter ϵ
for each LBS request from a vehicle **do**
  Compute request function $f(D)$
  Add Laplace noise to the request:

$$f(D) + \text{Lap}\left(\frac{\Delta f}{\epsilon}\right)$$

  Send the differentially private request to the group leader
end for
for each response from the group leader **do**
  Compute response function $G(D)$
  Add Laplace noise to the response:

$$G(D) + \text{Lap}\left(\frac{\Delta G}{\epsilon}\right)$$

  Send the differentially private response to the vehicle
end for
End $= 0$

V. EXPERIMENTS AND RESULTS

A. EXPERIMENTAL SETUP

A desktop system is used for all the practical experiments of this proposed scheme. System specifications were these: Intel Core i5-4570 with Base speed of 3.20GHz, with 4 cores and 4 logical cores having 6.0 MB L3 cache. 16 GB of RAM with 1600 MHz speed and 2.5 inch 256 GB Liteon SSD with Windows 10 pro operating system were used. Python version 3.8x was with SUMO version 1.18x.

For the Group Leader Selection, we need to have data on vehicles. So, for vehicle data collection, we have done a simulation of vehicles within the selected area, of Seoul, South Korea as it was selected by the base paper technique. Then using the SUMO Urban Mobility tool, we gathered vehicle data. For Differential privacy and Federated learning processes, I have used the EMNIST handwritten letters dataset.

B. SIMULATION PROPERTIES

SUMO simulation tools were used for traffic simulation in the selected area. We used the Open Street Map for area selection. The area for simulation we selected is Seoul location South Korea with spans of 2.5*1.5 km. We selected this area because it contain more than 30 interactions which were enough to use for simulation and experimentation on the data being gathered. Different properties of vehicles were gathered and used in the simulation process, these include the position of vehicles(posm), timestamp at it starts or stops, speed of vehicles (speed), vehicle ID, lane ID, latitude(xm) and longitude(ym). In addition, other properties metrics were used like cell id, centeredness of vehicles in the specified cell, acting frequency, transmission power, traffic rules obeyed, processing power, and storage available of each vehicle. We followed the specific method for the group leader selection process, and we used some specific vehicle features like centeredness of vehicles in the cell, resources available of OBU, and traffic rules obeyed.

We have done the simulation of 100, 200, and 300 devices to check the time taken for the scheme on different numbers of vehicles to be implemented. Figure 9 shows SUMO tool visualization during the simulation of vehicles in a selected area. For 100 vehicles the implementation of the proposed scheme took 3.5 seconds without the group leader and 2.0 seconds with the group leader. For 200 vehicles it takes 3.0 seconds with the group leader 4.5 seconds without the group leader 5.5 seconds without the group leader and 4 seconds with the group leader for 300 vehicles respectively. The proposed scheme shows the importance of group leaders

FIGURE 9. A view of SUMO tool during vehicle simulation in selected area.

with less time in implementation and more time without a group leader because then each vehicle must have to do its own task which is time taken obviously. The selected group leader has more resources and power to perform the task required for the proposed model. This way, our proposed scheme with an optimized group leader selection process and task management improves the scalability and performance of the model. Table 6 shows the performance of our proposed method with other techniques.

TABLE 6. Demonstration of different schemes based on performance.

Technique	Location Accuracy	Low Traceability	Real Time Accuracy	Privacy Preservation
Mx-zone	No	Yes	No	Yes
Silent Period	No	Yes	Yes	No
Spatial Obfuscation	Yes	No	No	No
Cui Shadowing	Yes	Yes	No	Yes
Proposed Scheme	Yes	Yes	Yes	Yes

C. ANONYMITY SET SIZE

Anonymity set size refers to how effectively an adversary can distinguish a specific vehicle from other vehicles in the vicinity. It considers the probabilities associated with the target vehicle and potential vehicles that could be mistaken for it. Let's denote the actual route taken by a vehicle VA as LocA, and the set of all possible locations that could be confused with LocA as LocN. The anonymity set of LocA, denoted by ASA, comprises all such potential locations where the adversary might mistakenly identify the target vehicle. The size of the anonymity set, denoted by |ASA|, indicates the number of items in this set. Mathematically, the anonymity set can be expressed as:

$$ASA = \{Loc_N : P(A, N) \neq 0\} \tag{4}$$

In our proposed technique, the anonymity set size of vehicle VA increases with its travel time. At first, when travel time t = 0, the size of |ASA| is small. As vehicle VA moves, it creates a shadow effect, making it harder for attackers to tell the real route from the shadows. This makes the size of |ASA| grow over time. In areas with high traffic, the anonymity set size increases significantly.

For example, in busy urban areas with 400 vehicles, the anonymity set size reaches about 1.5 bits. At first, with 200 vehicles, the size is around 1.08 bits, but it rises quickly as the number of vehicles reaches 400. Beyond this, the set continues to grow. A larger anonymity set size means a more effective privacy scheme, assuming each vehicle is equally likely to be targeted.

Our proposed scheme performs better than the base paper scheme in both sparse and dense conditions. Figure 10b shows how our scheme either outperforms or matches other schemes in terms of anonymity set size.

D. ENTROPY OF ANONYMITY SET SIZE

Entropy is a measure of uncertainty that shows the relationship between the actual location of a vehicle, V1, and the locations of other vehicles, Vn, that create shadows. Even though it may not always be practical to assume all vehicles are equally likely to be targeted, entropy is a useful way to measure the strength of the privacy scheme. The higher the probability that vehicles are equally targeted, the higher the entropy.

To measure location privacy, we use entropy as a metric for global anonymity. We denote the anonymity set as N, where each vehicle is represented by an index i. The total number of vehicles is |N|, and pi is the probability that vehicle i is the target. The entropy H of an individual vehicle within the anonymity set N can be calculated using the following equation:

$$H = -\sum_{i=1}^{|N|} p_i \log_2(p_i) \tag{5}$$

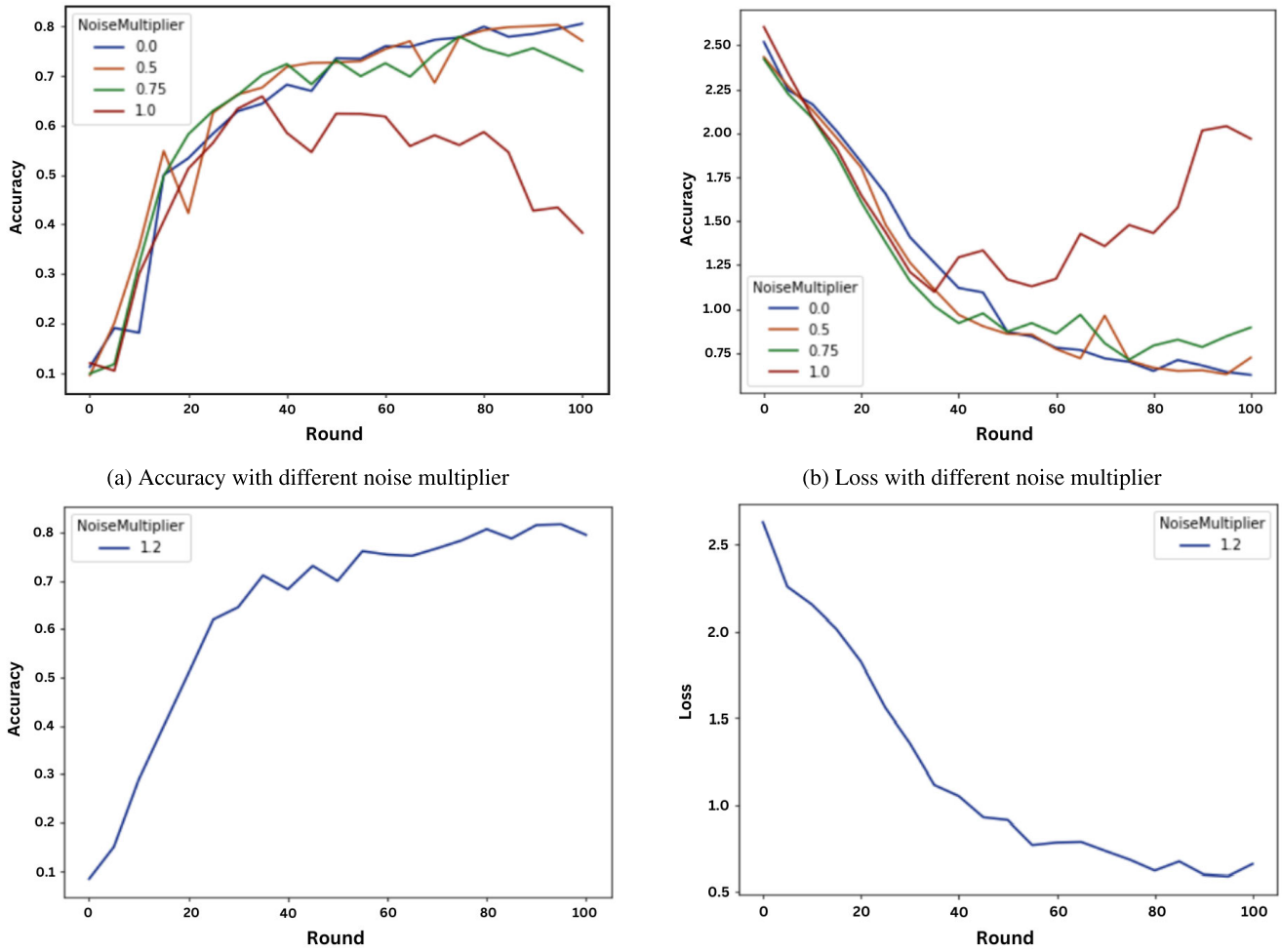
Looking at the figure, initially, the entropy is around 1.1 bits when the size of the vehicle set is 100. However, as more vehicles, up to 400, enter the network, the entropy of the anonymity set progressively increases. Notably, there is a sudden and sharp rise in entropy as the number of vehicles in the set increases, as illustrated in Figure 10c. In terms of this metric, our proposed scheme demonstrates superior performance compared to the base paper scheme.

E. TRACKING SUCCESS RATIO

The tracking success ratio refers to how effectively a vehicle can be continuously tracked, with 90% of its trace available to potential adversaries. Continuous tracking is essential for compromising privacy, as adversaries require the complete trajectory of a vehicle, including known endpoints, for de-anonymization purposes. Let's denote the tracking success ratio as PA, which is the probability that the anonymity set size |ASA| for a vehicle VA is 1. Mathematically, it is defined as: PA = Pr(AS = 1)

$$PA = Pr(|AS| = 1) \tag{6}$$

PA will be one at a time only when anonymity set also 1. As the number of elements in the anonymity set increases,



(a) Accuracy with different noise multiplier

(b) Loss with different noise multiplier

(c) Accuracy with desired 1.2 noise multiplier

(d) Loss with desired 1.2 noise multiplier

FIGURE 10. **Analysis of proposed FL scheme on different noise levels.**

PA decreases. The tracking success ratio measures the threat level posed by network attackers.

In Figure 10d, we compare our proposed scheme with the base paper scheme for different sets of vehicles. For example, with 100 vehicles, the base paper scheme has a tracking ratio of 20%, while our proposed scheme has a tracking ratio of 37% in the same scenario.

Similar improvements are observed for other sets of values, indicating enhanced privacy protection for vehicles. Furthermore, the computational overhead of both schemes is shown in Table 5. Comparative analysis reveals that our framework offers security features with minimal compute and communication overheads, as demonstrated through experimental and analytical findings.

F. SECURITY ANALYSIS

1) THREAT MODEL

In our tests, we looked at various potential threats to see how strong our system is for the IoV and VANET. We focused on two main types of attacks:

Sybil Attack: A bad actor creates many fake identities to mess up network operations or get unauthorized access.

Message Tampering: This means including wrong location information or changing the routing process to damage the system's integrity.

2) SYSTEM RESILIENCE

Our proposed scheme employs several strategies to mitigate these threats:

Group Leader Selection: The process of selecting a group leader within each cell of vehicles ensures that only trusted nodes with sufficient resources and adherence to traffic rules are assigned leadership roles. This minimizes the potential for Sybil attacks by ensuring that malicious entities are not designated as leaders.

Shadowing Scheme: By creating a shadow effect through the movement of vehicles, our scheme enhances privacy and makes it difficult for adversaries to discern the real route of a vehicle from the shadows. This effectively mitigates message tampering attacks by obfuscating the true location of vehicles.

Base Station Verification: While Road Side Units (RSUs) do not directly participate in detection, they serve as authorities capable of verifying a vehicle's origin. This adds an additional layer of security by ensuring that only authenticated vehicles are granted access to the network.

3) EVALUATION METRICS

To assess the efficacy of our system against these threats, we considered several key metrics:

False Positive Rate: This measures how often real vehicles are wrongly identified as bad. Lower false positive rates mean the detection system is more accurate and better at resisting Sybil attacks.

Anonymity Set Size: This shows how well an adversary can tell one vehicle from others nearby. A larger anonymity set size means better privacy protection and better defense against message tampering attacks.

Entropy of Anonymity Set Size: Entropy measures the uncertainty associated with the anonymity set, providing insight into the level of privacy afforded by the system. Higher entropy values signify increased privacy protection and make it more challenging for adversaries to track individual vehicles.

Tracking Success Ratio: This metric evaluates how effectively a vehicle can be continuously tracked, with a higher ratio indicating greater difficulty for adversaries in compromising privacy through continuous monitoring.

G. COMPARATIVE ANALYSIS

We tested our new method by comparing it to existing techniques. We looked at how well it chose group leaders, false positive rates, sizes of anonymous groups, entropy, and tracking success rates. Our method worked better, reducing implementation time and balancing privacy with accurate location services. Figure 11 shows the time taken for group leader selection in the original paper and our method. In conclusion, our method provides an efficient and secure way to communicate data for IoV and VANET. Using trusted group leader selection, shadowing schemes, and base station verification, we ensure data security, reliability, scalability, and privacy, making it suitable for the needs of modern vehicular networks.

VI. DISCUSSION

The proposed framework successfully combines Federated Learning (FL) with Differential Privacy (DP) and an optimized group leader selection mechanism to enhance privacy and scalability in IoV. Simulation results demonstrate significant improvements, including a 20% reduction in computational overhead in group leader selection and enhanced privacy protection, evidenced by larger anonymity set sizes and reduced tracking success ratios. However, the framework has limitations, the trade-off between privacy and utility due to DP noise, and scalability challenges in networks with thousands of vehicles. Also, the proposed method does not explicitly address scenarios like network partitions or abrupt vehicle failures, which could impact real-time operations. Future directions can include hierarchical group leader selection, failure detection and recovery, user-specific privacy control mechanisms, and different levels of privacy on different metrics. And using lightweight, energy-aware optimizations for FL and DP implementations, ML techniques for group leader selection, privacy parameters, and resource allocation dynamically, based on real-time and historical data.

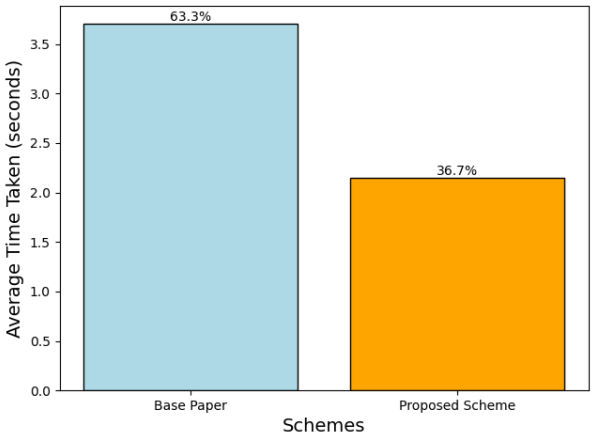


FIGURE 11. Time consumption comparison of different scheme.

VII. CONCLUSION

An optimized privacy preserved unique framework is introduced in our research, which protects the privacy of location, and LBS service, especially in the IoV. The proposed scheme uses Federated learning with differential privacy and an optimized group leader selection process to preserve location privacy before sending any data to the server while providing accurate location-based services (LBS). For maintaining privacy in routes followed and real-time location proposed scheme uses differential privacy. Our scheme maintains the balance between the accuracy of LBS requests in VANET and in the privacy of location to make it provide accurate results. Collection of data and model training will be on edge devices and only updated model aggregation will be shared to the group leader with the addition of differential privacy on each request and model update. Federated Learning ensures model training on edge devices while differential privacy provides privacy in VANET making the overall scheme scalable, distributed, secure, and more private in the context of IoV. It introduces an optimized performance while maintaining the privacy of data and LBS requests. Our proposed method achieves the best trade-off between utility, performance, and privacy.

ACKNOWLEDGMENT

The authors acknowledge TU Wien Bibliothek for financial support through its Open Access Funding Program.

REFERENCES

[1] S. Bayan and U. Mohammad, "A survey of data dissemination schemes in secure inter-vehicle communications," in *Proc. IEEE 13th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Mar. 2023, pp. 1224–1230.

[2] S. A. Moqurrab, A. Anjum, N. Tariq, and G. Srivastava, "Instant_anonymity: A lightweight semantic privacy guarantee for 5G-enabled IIoT," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 951–959, Jan. 2023.

[3] X. Wang, Z. Ning, M. Zhou, X. Hu, L. Wang, Y. Zhang, F. R. Yu, and B. Hu, "Privacy-preserving content dissemination for vehicular social networks: Challenges and solutions," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1314–1345, 2nd Quart., 2019.

[4] K. Levy and B. Schneier, "Privacy threats in intimate relationships," *J. Cybersecurity*, vol. 6, no. 1, pp. 1–13, Jan. 2020.

[5] I. Ullah, M. A. Shah, A. Khan, and G. Jeon, "Privacy-preserving multilevel obfuscation scheme for vehicular network," *Trans. Emerg. Telecommun. Technol.*, vol. 32, no. 2, p. 4204, Feb. 2021.

[6] M. Babaghayou, N. Labraoui, and A. A. A. Ari, "Location-privacy evaluation within the extreme points privacy (EPP) scheme for VANET users," *Int. J. Strategic Inf. Technol. Appl.*, vol. 10, no. 2, pp. 44–58, Apr. 2019.

[7] W. Liang, Y. Hu, X. Zhou, Y. Pan, and K. I. Wang, "Variational few-shot learning for microservice-oriented intrusion detection in distributed industrial IoT," *IEEE Trans. Ind. Informat.*, vol. 18, no. 8, pp. 5087–5095, Aug. 2022.

[8] S. M. Dibaji, M. Pirani, D. B. Flamholz, A. M. Annaswamy, K. H. Johansson, and A. Chakrabortty, "A systems and control perspective of CPS security," *Annu. Rev. Control*, vol. 47, pp. 394–411, Jan. 2019.

[9] N. Guo, L. Ma, and T. Gao, "Independent mix zone for location privacy in vehicular networks," *IEEE Access*, vol. 6, pp. 16842–16850, 2018.

[10] K. Sampigethaya, L. Huang, M. Li, R. Poovendran, K. Matsuura, and K. Sezaki, "CARAVAN: Providing location privacy for VANET," in *Proc. Embedded Secur. Cars (ESCAR)*, vol. 8, Jan. 2005, pp. 1–10.

[11] L. Huang, K. Matsuura, H. Yamane, and K. Sezaki, "Enhancing wireless location privacy using silent period," in *Proc. IEEE Wireless Commun. Netw. Conf.*, vol. 2, Mar. 2005, pp. 1187–1192.

[12] M. Li, K. Sampigethaya, L. Huang, and R. Poovendran, "Swing & swap: User-centric approaches towards maximizing location privacy," in *Proc. 5th ACM Workshop Privacy Electron. Soc.*, Oct. 2006, pp. 19–28.

[13] A. Wahid, H. Yasmeen, M. A. Shah, M. Alam, and S. C. Shah, "Holistic approach for coupling privacy with safety in VANETs," *Comput. Netw.*, vol. 148, pp. 214–230, Jan. 2019.

[14] H. Jiang, J. Li, P. Zhao, F. Zeng, Z. Xiao, and A. Iyengar, "Location privacy-preserving mechanisms in location-based services: A comprehensive survey," *ACM Comput. Surv.*, vol. 54, no. 1, pp. 1–36, Jan. 2022.

[15] Q. Huang, X. Xu, H. Chen, and L. Xie, "A vehicle trajectory privacy preservation method based on caching and dummy locations in the Internet of Vehicles," *Sensors*, vol. 22, no. 12, p. 4423, Jun. 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/12/4423

[16] N. Takbiri, A. Houmansadr, D. L. Goeckel, and H. Pishro-Nik, "Limits of location privacy under anonymization and obfuscation," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 764–768.

[17] J. Lim, H. Yu, K. Kim, M. Kim, and S.-B. Lee, "Preserving location privacy of connected vehicles with highly accurate location updates," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 540–543, Mar. 2017.

[18] X. Zhou, W. Liang, K. I. Wang, and L. T. Yang, "Deep correlation mining based on hierarchical hybrid networks for heterogeneous big data recommendations," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 1, pp. 171–178, Feb. 2021.

[19] J. Cui, J. Wen, S. Han, and H. Zhong, "Efficient privacy-preserving scheme for real-time location data in vehicular ad-hoc network," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3491–3498, Oct. 2018.

[20] P. Asuquo, H. Cruickshank, J. Morley, C. P. A. Ogah, A. Lei, W. Hathal, S. Bao, and Z. Sun, "Security and privacy in location-based services for vehicular and mobile communications: An overview, challenges, and countermeasures," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4778–4802, Dec. 2018.

[21] J. Ni, K. Zhang, X. Lin, and X. Shen, "Securing fog computing for Internet of Things applications: Challenges and solutions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 601–628, 1st Quart., 2018.

[22] Z. Lu, G. Qu, and Z. Liu, "A survey on recent advances in vehicular network security, trust, and privacy," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 2, pp. 760–776, Feb. 2019.

[23] G. Mehmood, M. Z. Khan, H. U. Rahman, and S. Abbas, "An efficient and secure session key establishment scheme for health-care applications in wireless body area networks," *J. Eng. Appl. Sci.*, vol. 37, no. 1, pp. 9–18, Jun. 2018.

[24] N. Takbiri, A. Houmansadr, D. L. Goeckel, and H. Pishro-Nik, "Fundamental limits of location privacy using anonymization," in *Proc. 51st Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2017, pp. 1–6.

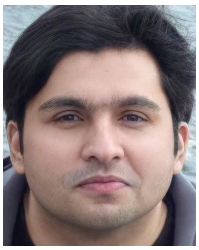
[25] S. A. Moqurrab, A. Anjum, A. Khan, M. Ahmed, A. Ahmad, and G. Jeon, "Deep-confidentiality : An IoT-enabled privacy-preserving framework for unstructured big biomedical data," *ACM Trans. Internet Technol.*, vol. 22, no. 2, pp. 1–21, May 2022.

[26] X. Li, Y. Zhu, and J. Wang, "Highly efficient privacy preserving location-based services with enhanced one-round blind filter," *IEEE Trans. Emerg. Topics Comput.*, vol. 9, no. 4, pp. 1803–1814, Oct. 2021.

[27] Y. Li, Y. Yin, X. Chen, J. Wan, G. Jia, and K. Sha, "A secure dynamic mix zone pseudonym changing scheme based on traffic context prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 9492–9505, Jul. 2022.

MUHAMMAD ADNAN is currently pursuing the Master of Science degree in information technology (MSIT), focusing on the Internet of Things (IoT), particularly the Internet of Vehicles (IoV) and the integration of differential privacy with federated learning. He is also with the Institute of Information Technology, Quaid-i-Azam University, Islamabad, Pakistan. His expertise includes full stack development, machine learning (ML), artificial intelligence (AI), federated learning (FL), differential privacy (DP), and Python scripting.

MADIHA HAIDER SYED received the Ph.D. degree in computer science from Florida Atlantic University, USA, in 2019. She is currently an Assistant Professor with the Institute of Information Technology, Quaid-i-Azam University, Pakistan. She was a recipient of the prestigious Fulbright Scholarship, in 2014, for the Ph.D. degree. Her research interests include cloud computing, security, privacy, software architecture, the IoT, cyber-physical systems, machine learning, and deep learning.



ADEEL ANJUM received the Ph.D. degree in computer science from Polytech Nantes, Nantes, France, in 2013. He is currently a Professor and the Director with the Department of Information Technology, Quaid-i-Azam University, Islamabad, Pakistan. He has several publications and authored a book on data privacy. His research focuses on AI-based data privacy. He was on the technical program committees of various international conferences.

SEMEEN REHMAN (Member, IEEE) received the Habilitation degree in the area of embedded systems from the Faculty of Electrical Engineering and Information Technology, TU Wien, in October 2020, and the Ph.D. degree with Karlsruhe Institute of Technology (KIT), Germany, in 2015. She served as an Assistant Professor at TU Wien. She is currently an Associate Professor at the University of Amsterdam (UvA), and a Privatdozentin at TU Wien. Her research focuses on dependable and energy-efficient embedded systems, approximate computing, security, and cyber-physical systems (CPS)/Internet of Things (IoT). She has co-authored a book, multiple book chapters, and over 90 publications in leading journals and conferences. She has received several prestigious awards, including Best Paper Awards at CODES+ISSS in 2011 and 2015, a Best Paper Award Nomination at DATE 2017, multiple HiPEAC Paper Awards, the DAC Richard Newton Young Student Fellow Award in 2015, and a Research Student Award from KIT in 2012. She has also served as a Topic Track Chair/co-chair and participated in the Technical Program Committees of various premier IEEE/ACM conferences on design automation and embedded systems.