Enhancing Data Collection in Vehicular Network Through Clustering Optimization

Chérifa Boucetta
Université de Reims Champagne-Ardenne
CReSTIC, Reims, France
cherifa.boucetta@univ-reims.fr

Frédéric Lassabe
UTBM, CNRS, Institut FEMTO-ST
F-90010 Belfort, France
frederic.lassabe@utbm.fr

Oumaya Baala
UTBM, CNRS, Institut FEMTO-ST
F-90010 Belfort, France
oumaya.baala@utbm.fr

***Abstract*—In this paper, we present a novel approach to enhance data collection in Vehicular Ad-Hoc NETworks (VANETs). VANETs are a growing area of interest due to their unique characteristics and challenges, such as rapidly changing topology and frequent network disruptions. Efficient data collection is a critical issue in vehicular networks and has therefore become a focus of research. To address this challenge, we propose a stable clustering optimization solution based on adaptive multiple metrics. The cluster head selection is done based on both mobility metrics, such as position and relative speed, and Quality of Service (QoS) metrics, such as neighborhood degree and link quality. The proposed solution has been tested and evaluated through simulations using a vehicular mobility simulator in a realistic urban environment. The results show that the proposed approach provides more stable clusters with higher QoS, and allows for the selection of the appropriate cluster head to collect data from the vehicles and forward it to the destination.**

***Index Terms*—Clustering, VANET, Data Collection, intelligent transportation system (ITS), Quality of Service (QoS).**

INTRODUCTION

Intelligent Transportation Systems (ITS) are used to monitor traffic flow, provide real-time information to drivers, and control traffic signals. They have significantly contributed to improving road safety, reduce congestion and enhance the efficiency of transportation systems. Within ITSs, Vehicular Ad-hoc Networks (VANETs) play a crucial role in providing communication services among nearby vehicles and infrastructures through Dedicated Short-Range Communication (DSRC) to enhance safety and provide real-time information to drivers.

In VANET network, there are three types of messages: Beacon, alert and service. The message Beacon is used for neighbor identification, discovery and control. Alert messages are used for traffic management and services which are intended for localization and discovery sites. Several challenges need to be addressed in vehicular network such as communication reliability, scalability, mobility management, security and privacy, etc. In essence, in a dense network, the communication overhead can become high, leading to reduced network performance. For that, clustering helps to reduce the communication overhead and increase the efficiency of information dissemination, leading to improved network performance [2].

Clustering consists on grouping vehicles into clusters based on their proximity to each other. It involves the selection of a cluster head (CH) which is responsible for managing the communication between the vehicles within the cluster and forwarding information to other clusters [13]. The selection of the cluster head can be done based on various parameters such as the distance between vehicles, the direction of vehicles, and the available resources of each vehicle. Several clustering algorithms have been proposed in the literature. Yu, et al. [9] proposed a clustering algorithm for VANETs that considers both time and distance for cluster formation. The algorithm is evaluated through simulations and compared to other clustering algorithms in terms of cluster stability, overhead, and delay. The authors in [17] proposed a dynamic clustering algorithm for VANETs that is based on vehicle density. The algorithm divides the network into clusters of varying sizes based on the density of vehicles in each area. It uses a dynamic threshold to adjust the cluster size and a gossip-based protocol to exchange information among vehicles. The algorithm was evaluated using simulations, and the results showed that it outperformed existing clustering algorithms in terms of stability and scalability. Another scheme is presented in [10]. The authors described an improved clustering algorithm for VANETs that considers both vehicular density and distance for cluster formation. The algorithm is evaluated through simulations and compared to other clustering algorithms in terms of cluster stability, overhead, and delay. Shi et al. proposed in [15] an improved clustering algorithm based on distributed DBSCAN for VANET by addressing the challenges of high communication overhead and low clustering accuracy in conventional DBSCAN algorithms. Another approach is to use reinforcement learning techniques, where the clustering algorithm can learn from the behavior of the vehicles and adapt its grouping accordingly [8], [6]. This can be particularly useful in scenarios where the vehicular network is dynamic and constantly changing. While these clustering schemes have demonstrated promising results in improving communication and resource utilization in VANETs, there are still some limitations. The most of these schemes rely on pre-defined fixed parameters for cluster formation and maintenance, which may not be suitable for dynamic and unpredictable scenarios. Additionally, while these clustering schemes improve communication efficiency and reduce congestion, they may also increase latency due to the extra overhead of cluster formation and maintenance. This paper proposes a real-time dynamic, adaptive, and flexible clustering scheme for vehicular networks, which can adjust clusters according to changing network conditions. The selection of cluster heads is based on two factors: similarity and degree of nodes. Similarity measures how closely a node's characteristics match those of its neighboring nodes, while the degree of a node represents the number of links it has to other nodes. The proposed scheme is evaluated through experiments conducted on a realistic urban application developed using C++. The study includes an analysis of the variation of the number of clusters and the stability of the cluster. Visualizations are used to present the experimental results.

I. System Model & Preliminaries

In this section, we present the network model and then define the cluster head selection criteria .

A. Network model

Our study focuses on a vehicular network that is modeled by an undirected graph, denoted as $G(V, E, t)$, which is generated every t seconds. Here, V represents the set of vehicles, while E denotes the set of edges that connect these vehicles [5]. To visualize the connectivity graph, we utilize a C++ Geographic Information System (GIS) application developed by our team called libgeologic. This application is based on the city of Belfort in northeastern France and represents a real environment. The city map is divided into cells of equal size, and several data sources, including GIS shape files and socio-economic information, are used to define an attraction weight for each cell. To simulate the mobility of vehicles in urban environments, we use a mobility module called V-MBMM, which is capable of accurately simulating vehicle displacement [3]. In addition, we define a radio propagation module, V-PROPAG (Vehicular Radio Propagation Model) [1], that calculates the radio coverage area of each vehicle in the network. It takes into account the density of obstacles and the environmental characteristics such as buildings, forests, and mountains, etc. Fig. 1 is an example of a radio wave propagation pattern. The V-MBMM model generates mobility traces and determine the nodes positions and movements. Whereas, the V-PROPAG model uses as input the data of the terrain categories to calculate the attenuation of the signals.

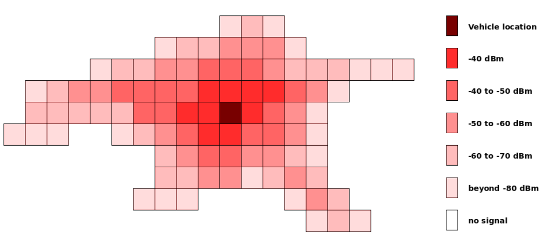

Fig. 1. Example of radio wave propagation pattern

B. Cluster-Head selection criteria

Vehicular networks are characterized by various topologies, high speeds, and unstable communication links. Therefore, ensuring the stability of clusters is crucial for facilitating efficient data exchange. A cluster is considered stable if it can maintain its structure over a period of time, despite the high mobility of vehicles. When setting up clusters, several parameters need to be considered, with the most important being speed and distance. Thus, we propose several criteria that we consider essential in selecting the CH.

a) Similarity: Similarity refers to the measure of how much two vehicles or clusters have in common based on specific criteria, such as location, speed, direction, etc. Many similarity-based clustering algorithms are proposed in the literature to group vehicles that exhibit similar behavior or are located in close proximity to each other [11], [7], [14] and [4]. This criteria enables efficient data sharing and facilities communication between vehicles that have similar driving patterns, preferences, or interests. In the proposed scheme, we extend the definition of the similarity introduced in [11]. We denote $S(i, j)$ the similarity between two nodes i and j. It is defined as follow:

$$S(i,j) = \frac{\|P_i^{'} - P_j^{'}\| - \|P_i - P_j\|}{N_i} \tag{1}$$

$$P_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix} \tag{2}$$

$$P_i^{'} = \begin{pmatrix} x_i + v_{x,i}T \\ y_i + v_{y,i}T \end{pmatrix} \tag{3}$$

where:

- P_i, P_j: current position of the nodes i and j (two-dimensional vector)
- P_i', P_j': expected position of the nodes i and j (two-dimensional vector)
- v: speed of the vehicle
- T: the time interval.
- N_i: degree of vehicle i. It presents the number of neighboring nodes and used to reduce the number of clusters.

b) Mobility: This criteria plays a significant role in selecting a CH in VANETs. It is based on the speed of the vehicle, the direction, the distance and the connectivity. Let $M(i, j)$ the mobility of nodes i and j.

$$M(i,j) = a\|P_i - P_j\| + b\|V_i - V_j\| \tag{4}$$

Where P_i, P_j and V_i, V_j are the current positions and the speed of nodes i and j, respectively. We define two coefficients a and b.

In the proposed approach, similarity and mobility are important in CH election. Hence, we combine them to choose the most stable CH. In Equation 5, we define the weight of CH i in a cluster composed of υ vehicles.

$$W_i = \max_{i \in \upsilon}(\mu N_i + \lambda \Delta M_i) \tag{5}$$

where, ΔM_i is the mobility average of i over T and λ and μ are two coefficients (λ, $\mu \in [0, 1]$).

C. Cluster Criteria

Maintaining the stability and number of clusters is crucial for improving the overall efficiency of the vehicular network and ensuring reliable communication in the presence of high mobility and dynamic changes.

a) Stability of the cluster: A stable cluster can minimize packet loss and provide a high-quality communication link for exchanging data, making it an important criterion for measuring the performance of clustering algorithms in VANET [12]. The more stable the cluster, the less load there is on maintenance and the more efficient the transmission of information. The selection of the CH is crucial for the stability of the cluster. The longer the service time of the CH, the more stable the cluster and the better the performance of the clustering algorithm. We calculate the average lifetime of a cluster as follow:

$$\Delta T = \frac{1}{L}\sum_{i=1}^{L}\Delta t_k \tag{6}$$

where,

ΔT_k is the average lifetime of the k^{th} cluster and L presents the number of clusters.

b) Number of Clusters: With the same distance and the same number of nodes, the number of clusters can directly affect the performance of the clustering algorithm. The average number of hops over the same distance increase with the number of clusters, which results in higher delays in information transmission and a much lower delivery rate. For these reasons, we aim to reduce the number of clusters in the proposed scheme.

II. The Proposed Approach

A. Clustering Algorithm

The clustering algorithm used for data dissemination in VANET categorizes the nodes into three types, each with its own unique characteristics and responsibilities.

- **Cluster Head**: is responsible for managing the cluster, coordinating the communication within it and forwarding the collecting data to other clusters or to the central server.Each cluster has one cluster head.
- **Member**: is a regular vehicle that is part of the cluster. It can be a relay node.
- **Undefined**: The status of nodes at the beginning of the clustering algorithm or during updates may not be defined until their role is determined by the algorithm.

There are 6 main steps in the algorithm as depicted in Fig. 2:

- Initialization
- Broadcast hello
- Neighbors list
- Compute vehicles' similarity
- Find the most stable node as cluster head
- Update vehicles' status

In the following paragraphs, we will outline the different steps of the algorithm.

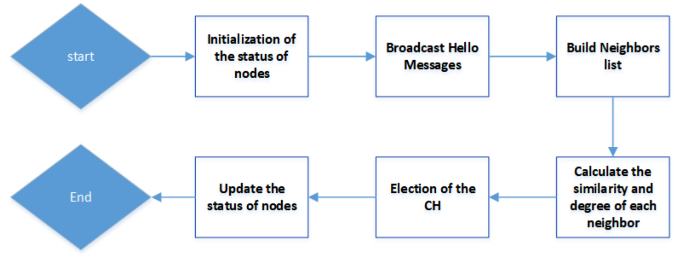

Fig. 2. Proposed Clustering Algorithm Flowchart

a) Initialization: Initially, all nodes are designated as **undefined**. The mobility parameter is initialized to 0. In order to determine the speed of each node, we use the mobility module of libgeologic tool introduced previously.

b) Broadcasting hello and making neighbors list: Each vehicle broadcasts a hello message to advertise itself to neighboring vehicles on the network. When vehicles within range of another node (red shape in figure 1) receive this message, they add the sending vehicle to their neighbor list. Each node forms a cluster with its neighbors. The Hello message contains information such as vehicle ID, location, and speed.

c) Calculate the parameters: After defining the clusters, the similarity and degree of nodes in each cluster are computed using Equations 1 and 4 as presented in section I. Following this, each vehicle calculates the average velocity of its neighbors within a time interval T. In the practical case, we represent the speed and position by the vectors in 3 dimensions and we take into account the direction to calculate the similarity parameter.

d) Find the stable node: The node with the highest parameter obtained from Equation 5 in each cluster is considered as the most stable node. This node is identified based on the parameters computed in step 3. It is expected to have a longer service time and can act as a Cluster Head for data dissemination.

e) Update the status of Vehicles: If a node is already classified as either a $head$ or a $member$ of an already existing cluster, its neighbors will be assigned to this cluster. If the node is in an $undefined$ state, a new cluster will be created.

Algorithm 1 describes our clustering model. It is repeated in order to keep clusters consistent with nodes layouts. Since nodes are vehicles, they are:

- Moving: it will change clusters composition by breaking and making links between vehicles, therefore modifying neighborhoods.
- "Appear": all vehicles do not move at the same time, therefore, a vehicle only participates to the clustering process if it is turned on which makes it look like it appeared in the nodes list.
- "disappear": contrary to vehicles which just turned on, a vehicle also can be turned off, thus being removed from the process as it disappeared from the nodes list.

The implementation uses a boolean state to keep track whether the vehicle is on (active) or off (inactive). For clarity, the algorithm only considers vehicles turned on and doesn't check its active state.

Algorithm 1 Clustering algorithm

Require: $Population$: a list of vehicles and their information at time t
  for all $vehicle$ **in** $population$ **do**
    $vehicle.State \leftarrow undefined$
    $vehicle.StabilityParameter \leftarrow computeStability()$
  end for
  for all $Vehicle$ **in** $population$ **do**
    $neighbors \leftarrow getNeighbors()$
    $most\_stable \leftarrow findMostStableVehicle(neighbors)$
    if $most\_stable.State = undefined$ **then**
      $most\_stable.State \leftarrow head$
      create $newCluster$
      add $newCluster$ **to** $clusterList$
      for all $object$ **in** $neighbors$ **do**
        add $object$ **to** $newCluster$
        $object.State \leftarrow member$
      end for
    else
      $cluster \leftarrow findVehicleCluster(most\_stable)$
      for all $object$ **in** $neighbors$ **do**
        add $object$ **to** $cluster$
        $object.State \leftarrow member$
      end for
    end if
  end for

B. Data dissemination and use case application

Our clustering algorithm can be used in a drone-based application for data collection and transmission. In this application, drones equipped with sensors communicate with cluster heads to collect data on traffic flow, road conditions, and environmental parameters. Collected data is then transmitted to a central server. This data can then be used to optimize route planning for emergency services or to alert drivers about potential hazards or congestion ahead.

The cluster head is responsible for coordinating the route planning and scheduling for the vehicles within its cluster. The data collection task can be considered as a vehicle routing problem (VRP) [16], which is a type of combinatorial optimization and integer programming problem. The goal is to determine the best trajectory for the drones, taking into account the battery level and storage capacity for data. By traversing all cluster heads, the drone can collect information from the entire network. This approach enables the drone to gather as much information as possible in a single round trip.

The proposed clustering scheme can be used in VRP to minimize the travel distance and improve the data collection time. By optimizing trajectories, our approach can help to reduce the overall energy consumption and improve the efficiency of data collection. The objective function is to minimize the total distance traveled by the drones while collecting data from the vehicles in the network. This function would take into account the distance between the drone's starting point and each head node in the network, as well as the distance between each pair of head nodes that the drone needs to visit in order to collect data from all the vehicles. The function could also consider the capacity of the drone's battery and the amount of data that it can store, in order to optimize the route planning.

The following is an example of a hypothetical scenario where drones are deployed to collect data for environmental monitoring, such as air quality index or traffic management purposes, such as congestion and accidents. The scenario consists of five clusters (Fig. 3) that the drones need to visit to collect data. The trajectory of the drones should be adjusted dynamically based on real-time data about traffic and road conditions to ensure efficient data collection. Fig. 3 shows the normal case where the drone has enough resources to visit all the CHs. However, in the second scenario in Fig. 4, the drone does not have enough battery and it must return to the recharge station. Therefore, it was only able to serve 4 clusters. In this case, the drone's battery life is a critical factor that must be taken into consideration when planning its trajectory. If the drone runs out of battery while collecting data, it would not be able to complete its mission, resulting in incomplete data collection and potentially rendering the entire mission useless

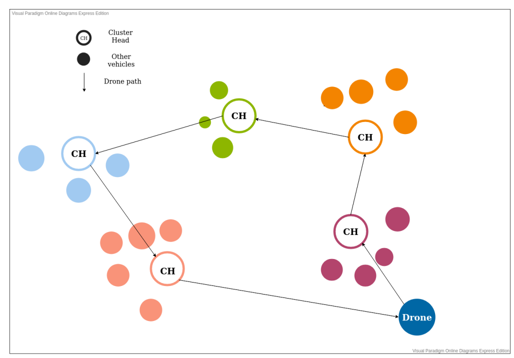

Fig. 3. Hypothetical scenario 1

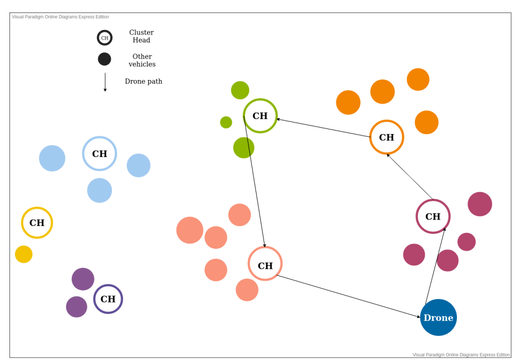

Fig. 4. Hypothetical scenario 2

III. Simulation and Performance Evaluation

This study involves the simulation of a vehicular network, consisting of 200 vehicles, on a $2000m \times 2000m$ area moving between $7am$ and $8am$. To determine path loss, obstacles between transmitters and receivers are taken into account, even those outside the roads. The clustering module is implemented on the libgeologic simulator, which allows the visualization of the obtained clusters. It is a GIS (Geographic Information System) application written in C++ and represents the city of Belfort in the northeastern France and is implemented by our team. We use the mobility model and the propagation model presented in our previous work [5].

A. Visualization

Figures 5 and 6 illustrate the visualization of clusters, where vehicles are grouped in different colors representing distinct clusters. Vehicles within a cluster can exchange information with each other, enabling communication between them. In the context of the vehicle-drone application, information from vehicles within a cluster can be transmitted to a drone through its designated CH. Both figures have blue dots that represent either static vehicles or vehicles that do not belong to any cluster (undefined vehicles). If a cluster head changes, the color of the cluster also changes accordingly over time. The rate of color transformation represents the speed at which a cluster changes its head vehicle. In cluster 1, there is no color change, indicating that the cluster head does not change despite the mobility of vehicles. However, in cluster 2 (orange cluster), the cluster head changes over time, thus is the color.

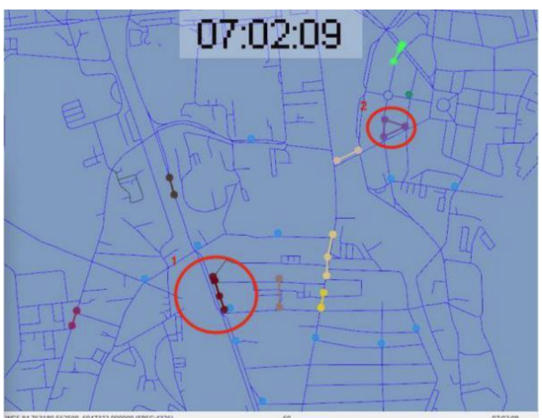

Fig. 5. Visualization at 7:02:09 am

B. Performance Evaluation

We evaluate the average cluster lifetime, the variability in the number of clusters based on the number of vehicles, and the cluster stability.

Fig. 7 and Fig. 8 illustrate the relationship between the number of vehicles and the number of clusters over the course of the simulation. It is observed that an increase in the number of vehicles leads to a corresponding increase in the number of clusters, while a decrease in the number of vehicles results in a decrease in the number of clusters. It is noteworthy that the maximum number of clusters, which is 37, is achieved when the number of active vehicles is 165. However, it

Fig. 6. Visualization at 7:02:26 am

should be noted that this parameter is highly dependent on the distribution of vehicles.

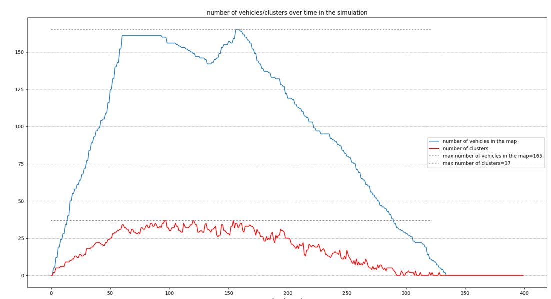

Fig. 7. number of vehicles/clusters

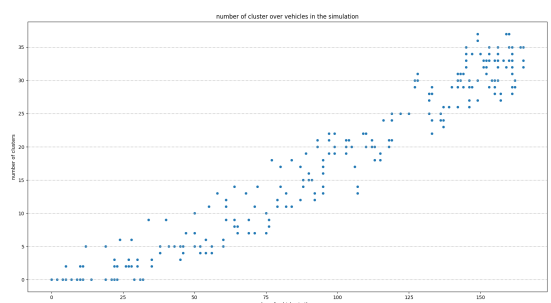

Fig. 8. number of cluster over vehicles

The cluster stability is evaluated based on the lifetime of the cluster head. Fig. 9 shows the stability of clusters as we vary the number of vehicles. At 200 vehicles, the average lifetime of all clusters is measured at $\Delta T = 5.199$. Fig. 9 illustrates that the stability of clusters decreases with an increase in the number of vehicles due to the loss of many connections caused by the distribution of the vehicles. Our proposed clustering algorithm exhibits higher stability than the conventional DBSCAN algorithm. This is because the

DBSCAN algorithm relies on a dynamic parameter to identify clusters, which is determined by a central point and may change frequently. In contrast, our algorithm utilizes a fixed parameter to measure clusters, leading to greater stability.

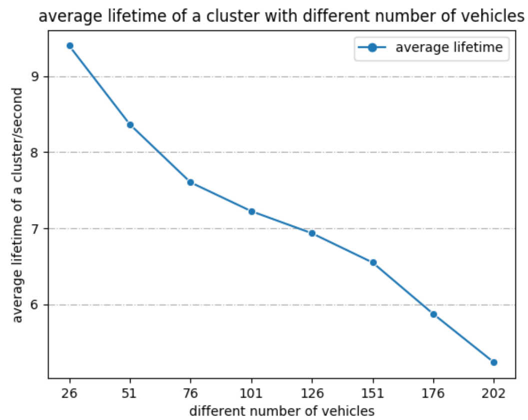

Fig. 9. Average lifetime of a cluster as a function of the number of vehicles

We consider 200 vehicles and vary the values of λ and μ in Equation 5. We consider the following values:

- λ=0,3; μ=0,7
- λ=0,5; μ=0,5
- λ=0,7; μ=0,3

We present results in table I. We observe that the degree of nodes plays a crucial role in determining the stability of the cluster. Specifically, as the proportion between the coefficient of average speed and the node degree coefficient increases, the lifetime of a cluster decreases. This indicates that the degree of nodes has a greater impact on the stability of clusters in the system compared to other factors such as the average speed of the nodes.

TABLE I
CLUSTER HEAD WEIGHT

λ	μ	**Cluster Head Weight**
0.3	0.7	5.45s
0.5	0.5	5.24s
0.7	0.3	5.14s

IV. CONCLUSION

Clustering in VANETs is a promising approach to enhance the performance of vehicular communication networks and improve the user experience. It helps to increase the efficiency of information dissemination and reduce the communication overhead. This paper presents an optimized clustering solution that achieves stability using adaptive multiple metrics. The cluster heads is selected based on mobility metrics, including position and relative speed, and Quality of Service (QoS) metrics such as neighborhood degree and link quality. We evaluated our solution through simulations in a realistic urban environment using a vehicular mobility simulator. The results demonstrate that our approach generates more stable clusters enabling the selection of an appropriate cluster head to collect and forward data from vehicles to the destination. As a future work, we would like to develop a cluster-based framework for security protocols in VANET that is both efficient and robust, utilizing cutting-edge techniques such as advanced machine learning, neural networks, and AI-based method.

V. ACKNOWLEDGMENT

We would like to express our gratitude to Martin Glodas, Haowen Ren and Mona Rahbari. They contributed to this research project during their Master's studies at UTBM, Belfort.

REFERENCES

[1] Kahina Ali Ali, Oumaya Baala, and Alexandre Caminada. Revisiting vehicular radio connectivity with propagation model. *Telecommunication Systems*, 11 2011.

[2] Elnaz Alizadeh, Aleksandr Ometov, Sergey Bezzateev, Anastasia Levina, Niko Mäkitalo, Sergey Andreev, and Yevgeni Koucheryavy. Dynamic clustering for intelligent transportation systems: A review. *IEEE Internet of Things Journal*, 8(6):4472–4503, 2021.

[3] Oumaya Baala, Cherifa Boucetta, Kahina Ali Ali, and Alexandre Caminada. More insights into communication issues in the internet of vehicles. 10 2018.

[4] Bidisha Bhabani and Judhistir Mahapatro. Clurma: A cluster-based rsu-enabled message aggregation scheme for vehicular ad hoc networks. *Vehicular Communications*, 39:100564, 2023.

[5] Cherifa Boucetta, Oumaya Baala, Kahina Ait Ali, and Alexandre Caminada. Performance of topology-based data routing with regard to radio connectivity in vanet. 06 2019.

[6] Aicha Dridi, Chérifa Boucetta, Abubakar Yau Alhassan, Hassine Moungla, Hossam Afifi, and Houda Labiod. Deep learning approaches for electrical vehicular mobility management: Invited paper. In *2019 International Conference on Wireless Networks and Mobile Communications, WINCOM 2019, Fez, Morocco, October 29 - November 1, 2019*, pages 1–6. IEEE, 2019.

[7] B. Hassanabadi, C. Shea, L. Zhang, and S. Valaee. Clustering in vehicular ad hoc networks using affinity propagation. *Ad Hoc Networks*, 13:535–548, 2014.

[8] Hyunbum Kim, Jalel Ben-Othman, and Lynda Mokdad. Intelligent terrestrial and non-terrestrial vehicular networks with green ai and red ai perspectives. *Sensors*, 23:806, 01 2023.

[9] Yubao Li, Fucheng Hu, Meng Guo, and Haojie Huang. A clustering algorithm for vanet based on time and distance. *Wireless Personal Communications*, 98(1):1233–1246, 2018.

[10] Bin Liu, Yunkai Wang, Dongpu Cao, Zhiyuan Chen, and Yan Liu. An improved clustering algorithm for vanet based on vehicular density and distance. *Wireless Personal Communications*, 87(4):1537–1554, 2016.

[11] Nitin Maslekar, Mounir Boussedjra, Joseph Mouzna, and Houda Labiod. A stable clustering algorithm for efficiency applications in vanets. pages 1188–1193, 07 2011.

[12] Ruchi Mehra, Rasmeet S. Bali, and Prabhsharan Kaur. Efficient clustering based olsr routing protocol for vanet. In *2016 Symposium on Colossal Data Analysis and Networking (CDAN)*, pages 1–7, 2016.

[13] Divyanshu Sharma and Bhupendra Singh Kaushik. Vanet clustering protocols: a comprehensive review. *Wireless Networks*, 26(2):1219–1245, 2020.

[14] Christine Shea, Behnam Hassanabadi, and Shahrokh Valaee. Mobility-based clustering in vanets using affinity propagation. pages 1 – 6, 01 2010.

[15] Liang Shi, Wenjie Li, and Gang Li. An improved clustering algorithm based on distributed dbscan for vanet. *Mobile Networks and Applications*, 23(3):632–641, 2018.

[16] Paolo Toth and Daniele Vigo. *The Vehicle Routing Problem*. Society for Industrial and Applied Mathematics, 2002.

[17] Lingling Zhang, Xuewen Chen, Wei Wang, and Zhiyong Wu. A dynamic clustering algorithm for vanets based on vehicle density. *Sensors*, 18(3):710, 2018.