



Geographical Information Based Clustering Algorithm for Internet of Vehicles

Rim Gasmi¹(✉), Makhlouf Aliouat¹, and Hamida Seba²

¹ Faculty of Sciences, LRSD Laboratory, Ferhat Abbas University Setif 1, Setif, Algeria

gasmi34000@gmail.com, aliouat_m@yahoo.fr

² LIRIS, CNRS Lyon 1, University of Lyon, Lyon, France
hamida.seba@univ-lyon1.fr

Abstract. Nowadays, Internet of Vehicles (IoV) are considered as the most important promoter domain in the Intelligent Transportation System (ITS). Vehicles in IoV are characterized by a high nodes' mobility, high nodes' number and high data storage. However, IoV suffer from many challenges in order to achieve robust communication between vehicles such as frequent link disconnection, delay, and network overhead. In the traditional Vehicular Ad hoc NETWORKS (VANETs), these problems have often been solved by using clustering algorithms. Clustering in IoV can overcome and minimize the communication problems that face vehicles by reducing the network overhead and ensure some Quality of Service (QoS) to make network connectivity more stable. In this work, we propose a new Geo-graphical Information based Clustering Algorithm "GICA" destined to IoV environment. The proposal aims to maintain the cluster structure while respecting the quality of service requirements as the network evolves. We evaluated our proposed approach using the NS3 simulator and the realistic mobility model SUMO.

Keywords: IoV · IoT · Clustering · Smart vehicles · Robust communication

1 Introduction

In the last decade, the world has seen a great development in the vehicle industry, which in turn increases the number of vehicles vastly. This development has surely made the lives of peoples easier and has even accelerated the economic growth around the world. In the other hand, it has also caused some challenges, such as traffic congestion, traffic accident, energy consumption and environmental pollution [1]. These issues led researchers to create a new concept called Vehicular Ad hoc Network (VANET) that allows vehicles to communicate with each other using Vehicle to Vehicle communication mode (V2V), with road Infrastructures using Vehicle to Infrastructure communication mode (V2I) [2].

© Springer Nature Switzerland AG 2021

É. Renault et al. (Eds.): MLN 2020, LNCS 12629, pp. 107–121, 2021.

https://doi.org/10.1007/978-3-030-70866-5_7

The emergence of the Internet of Things (IoT) opens the door to the conventional VANETs towards a new paradigm called the Internet of Vehicles (IoV). IoV is considered as a new Intelligent Transportation System paradigm [3]. IoV extends VANETs communication types, networks technologies and applications. This development creates new interactions at the road level among vehicles, humans and infrastructures. However, IoV still suffer from several drawbacks such as the high vehicles dynamicity that increases the frequent link disconnections. The high number of vehicles connected to IoV in smart city, leads to a high growth in network overhead. Clustering algorithms are the most used method to reduce network overhead and maximize network stability [4]. The concept of these algorithms is based on dividing network into several groups and each group is headed by a chosen group member. This new network scheme decreases the number of messages exchanged, where the network topology is already known and node does not need to broadcast route request to all network nodes in order to find a new route. In this paper, we propose a new Geographical Information based Clustering Algorithm “GICA” destined to IoV environment. The proposed algorithm aims to find the stable nodes in the network in order to be selected as Cluster Head.

2 Related Work

Several works have been proposed to address scalability issues in vehicular networks. Self-Organized Clustering Architecture for Vehicular Ad Hoc Networks (SOCV) [5], is a one-hop clustering protocol that ensures safe network communications in VANETs. The authors create a dynamic virtual backbone in the network that includes Cluster-Heads and cluster-gateways nodes. These nodes guarantee efficient message propagation in the network.

Destination Based Routing (DBR) algorithm for context-based clusters in VANETs [6] combines two mechanisms. The First one is based on direction, relative velocity, interest list and final destination. The second one is a destination-based routing protocol used to enhance inter-cluster communication. Using context-based clustering that includes interest list of vehicles, the overall end-to-end communication is significantly improved.

QoS Based Clustering for Vehicular Networks (QoSCluster) [7] is a multi-metric clustering algorithm. The authors propose two kinds of metrics to calculate the weight value: QoS Metrics that include average link expiration time, average link bandwidth and average time to completion, and Stability Metrics that include the level of connectivity, relative mobility and the average relative distance. The authors compared the proposed scheme to DMCNF, the obtained results show the effectiveness and robustness of QoSCluster.

Mobility adaptive density connected clustering algorithm (MADCCA) [8] uses traffic density to elect cluster heads. The Key parameters used in MADCCA are vehicle’s velocity and node density.

An efficient hierarchical clustering protocol (EHCP) for multi-hop communication is introduced in [9]. The goal of this proposal is to ensure effective resource

use and to enhance the network lifetime. The authors assume that the vehicles are connected to the Internet through road-side unit gateways. By using Internet connection, each vehicle can collect information about its neighboring nodes then it executes the clustering algorithm to elect the appropriate Cluster Heads. To elect Cluster Heads, the EHCP uses two different parameters: link expiration time and the vehicle's relative degree.

Distributed Clustering Algorithm Based on Dominating Set (DS) for Internet of Vehicles, called DCA-DS is presented in [10]. The authors use the node span parameter to select the Cluster Head. The node span parameter is defined as the number of neighboring nodes that belong to no cluster, including the node itself. DCA-DS uses simple heuristic method as well as a greedy strategy. The node that has the highest span value is added to the DS, which later takes the role of Cluster Head and the rest of its neighbors become Cluster Members. The process is iteratively repeated until all the nodes are clustered.

A Segmented Trajectory Clustering-Based Destination Prediction mechanism is proposed in [11]. The authors propose destination prediction-based trajectory segmentation algorithm to segment each original trajectory to different sub-trajectories. The sub-trajectories are clustered based on the average nearest point pair distance to reveal the common characteristics or similar tracks. Then the authors use a deep neural network based on the history trajectories in order to predict destinations.

3 Motivation

The majority of clustering algorithms existing in literature are generally based on one-hop clustering. This means that the communication process is only between a cluster member and its Cluster Head (1-hop distance at most). The use of such method minimizes the Cluster Head coverage area, thereafter, several clusters are established. The high number of formed clusters in the network leads to a cluster overlapping problem and decreases the network performance. Moreover, vehicular networks are characterized by a high topology change caused by a high vehicles' mobility and a limited driving directions. Most proposed schemes in literature do not take in consideration the combination between vehicle mobility and destination information to form clusters. Moreover, many proposed clustering approaches use high number of control messages in order to exchange vehicles parameters, which causes network overloading. This leads to several collisions, especially if the proposed system uses multi-metric mechanism that needs a high number of messages. In this work, we propose a geographical information-based clustering algorithm, called "GICA", destined to an IoV environment. GICA is based on three metrics: destination, mobility and position. It uses a multi-hop clustering mechanism to connect cluster members with their Cluster Head. The proposed algorithm combines the two main characteristics of vehicular networks direction and mobility in order to select the best nodes in the network to be Cluster Heads. Furthermore, we take advantage of beacon messages that are already used to update the parameters used to select Cluster Heads.

4 Network Model

In this section, we present the Geographical Information based Clustering Algorithm “GICA” destined to IoV environments. GICA aims to find the stable nodes in the network to elect them as Cluster Heads. It is based on the following assumptions:

- We suppose that the digital map is divided into numerous regions, and each vehicle knows the organization of the city map.
- Each vehicle in the network is equipped with a GPS device and a digital map that allow it to locate its position in the map.
- Each vehicle can estimate its speed.
- Each vehicle has its own final destination matrix that is used to save and calculate the vehicles that share the same final region as it.
- All nodes that are in the same region will pass by the same regions in order to arrive to their final destination regions.

In order to find the best vehicles in the network to be elected as Cluster Heads, we propose to divide the city map into different regions. The use of such method allows to find the most durable nodes in term of time, where the node that has the higher number of neighbors, that share the same next region as it, will be selected as Cluster Head. This selected node will not lead to frequent link disconnection because they will keep a long communication period with its neighbors. Figure 1 shows an example of clusters-based city map divided into 6 regions.

Our network architecture is organized in clusters. Each vehicle can take one of the following states:

- **Not-Decided (ND):** is the initial status taken by each node or when the cluster head gives up its role.
- **Cluster Head (CH):** it is the coordinator of the group (cluster). Its role is to communicate information between nodes (intra-cluster communications), or between different clusters (inter-cluster communications).
- **Cluster Member (CM):** it is a core member node that benefits different services from the CHs.
- **Gateway:** is a relay node. It is the common access point for two or more cluster heads.

5 Cluster Head Election Metrics

To elect the clusterheads, we use the following metrics:

Destination Metric (D): is defined by the number of vehicles that share the same final region as a specific vehicle. To compute D, we use a matrix, called the final destination matrix, to save the nodes that share the same final region. All the nodes in the network are identified by an ID from 1 to n with n is the number of nodes in the network. The final destination matrix A size is $n * n$ such

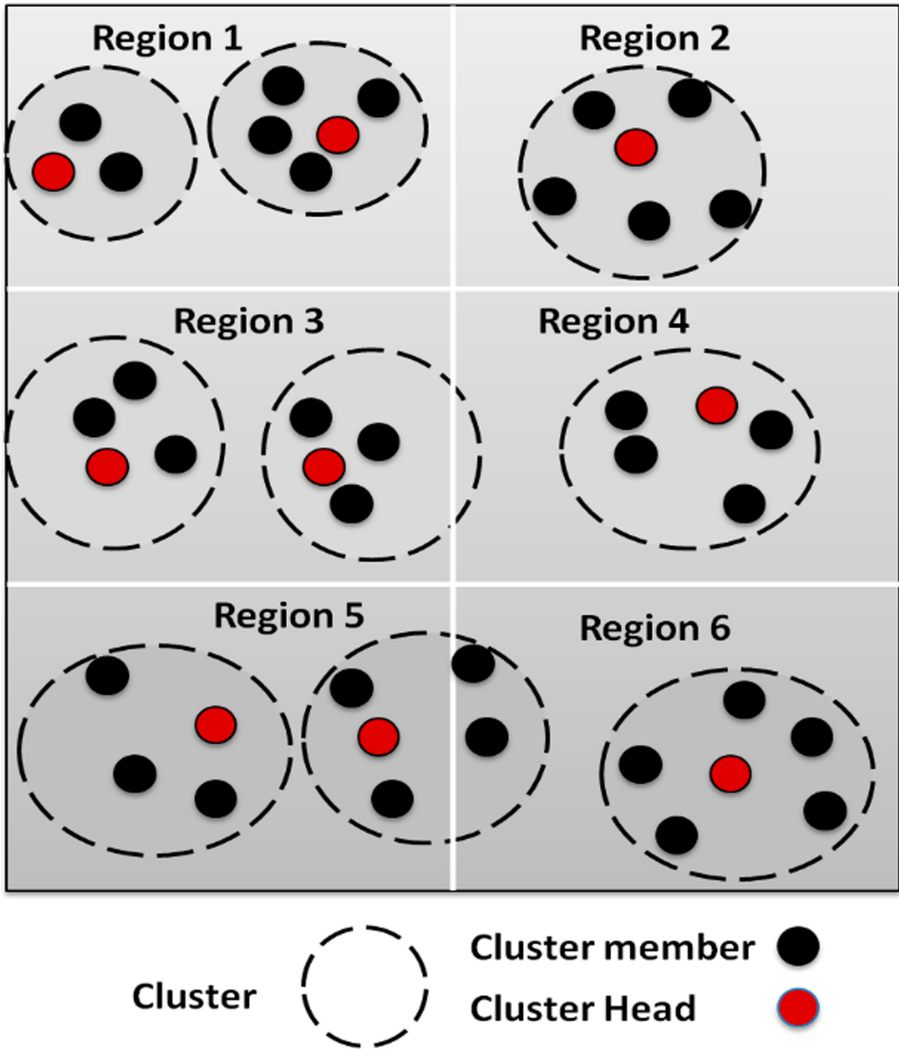


Fig. 1. clusters-based city map divided into 6 regions

that $A[i][j] = 1$ if vehicle i shares the same final region as vehicle j and $A[i][j] = 0$ otherwise. The sum of columns from 1 to n in line i indicates the number of vehicles that share the same final region as vehicle i . So, the value of D is computed as follows:

$$D = \sum_{j=1}^n A[i][j], \tag{1}$$

Table 1 shows an example of final destination matrix with n equal to 12.

Table 1. Final destination matrix.

	01	02	03	04	05	06	07	08	09	10	11	12
01	1	0	1	0	0	1	0	1	1	0	1	1
02	0	1	0	1	0	0	0	0	0	0	0	0
03	1	0	1	0	0	1	0	1	1	0	1	1
04	0	1	0	1	0	0	0	0	0	0	0	0
05	0	0	0	0	1	0	1	0	0	0	0	0
06	1	0	1	0	0	1	0	1	1	0	1	1
07	0	0	0	0	1	0	1	0	0	0	0	0
08	1	0	1	0	0	1	0	1	1	0	1	1
09	1	0	1	0	0	1	0	1	1	0	1	1
10	0	0	0	0	0	0	0	0	0	1	0	0
11	1	0	1	0	0	1	0	0	0	0	1	1
12	1	0	1	0	0	1	0	0	0	0	1	1

Position Metric (P): is defined as the average distance between a vehicle and its neighbors. The node that has the lowest distance to its neighbors should be selected as a cluster head. Let (x_i, y_i) denote the position coordinate of vehicle i, (x_j, y_j) the position coordinate of the neighbor vehicle j of node i. Here, we calculate the mean position of the neighbors of node i as $x_m = \frac{\sum_{j=1}^n x_j}{N}$, $y_m = \frac{\sum_{j=1}^n y_j}{N}$, where N is the number of its neighbors. Let (x_{max}, y_{max}) denotes the vehicle position that has the longest distance to the mean position of the neighbors of vehicle i. The general distance P of node i is given by:

$$P = \frac{\sqrt{(x_i - x_m)^2 + (y_i - y_m)^2}}{\sqrt{(x_{max} - x_m)^2 + (y_{max} - y_m)^2}}, \quad (2)$$

Mobility Metric (M): In our proposed approach, The mobility metric is defined as the ratio between the node velocity and the average of its neighbors' velocities. The node that has the low mobility will not leave the cluster early, which indicates that the node is suitable to be selected as CH. Let us assume that $P_1(x_1, y_1)$ is the position of node i at time t_1 and $P_2(x_2, y_2)$ is the position of node i at time t_2 . d_i is the distance traveled by node i over time Δt ($\Delta t = t_2 - t_1$):

$$d_i = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}, \quad (3)$$

Thus, the velocity of node i over Δt , is computed as:

$$v_i = d_i / \Delta t, \quad (4)$$

The mobility metric is defined as follows:

$$M_i = \frac{v_i}{\sum_{j=1}^n v_j / n}, \quad (5)$$

Where n is the number of neighbors.

To select the CHs a combination of the previous defined parameters is proposed:

$$W = f_1 D + f_2 \frac{1}{M} + f_3 \frac{1}{P}. \quad (6)$$

Where f_1 , f_2 and f_3 are weight factors and $f_1 + f_2 + f_3 = 1$

6 Geographical Information Based Clustering Algorithm “GICA”

6.1 Initial Network Phase

This phase defines the first period where no clusters have been formed in the network. Each vehicle starts with the initial status (ND), for each Interval Time it broadcasts a Hello message to its neighbors in the aim of exchanging with them its updated information to be used to elect the Cluster Heads such as mobility, position, region, etc. The Hello message exchange process is also used to inform the neighbors that the corresponding vehicle is still present and did not leave the transmission range. After receiving the Hello message, each vehicle modifies and updates its neighborhood table. The corresponding vehicle adds the new discovered vehicles that did not already exist in its neighborhood table and updates the information of those that are already saved in its neighborhood table. The received Hello message is also used to recalculate mobility and position parameters used to calculate the weight value. Moreover, each vehicle updates its final region matrix by adding the new vehicles that share the same final region as the corresponding vehicle or by delating the vehicles that leave the communication range or change their final region. The Hello message includes the vehicle ID, current status, current position, current mobility, final region, weight value and CH ID (Fig. 2).

Algorithm 1: Initial network phase

```

Node-status ND;
Wait for a Hello message;
Receive(Hello message);
Get position(x,y);
Get final region();
Calculate destination(D)metric based on Eq.1;
Calculate mobility(M)metric based on Eq.2;
Calculate position(p)metric based on Eq.(3&4&5);
Calculate node weight based on Eq.6;
Exchange Hello message;

```

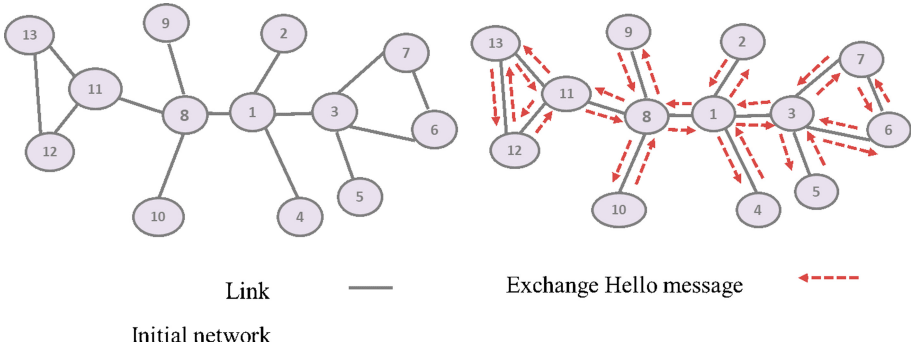


Fig. 2. Initial network phase

6.2 Clustering Phase

Our clustering election process is based on k-hop cluster structure, where the CH is selected based on the calculated weight value at k-hop distance. After receiving all the needed information to calculate the weight value (destination, mobility, position). Each vehicle calculates its weight value and compares it with those of its neighbors. The vehicle that owns the highest weight selects itself as CH and sends a message CHMsg containing its ID and its weight value to all its neighbors. When a ND vehicle receives a CHMsg message, it changes its status to CM in order to assume its role as a cluster member and updates the ID of its CH. After that, each vehicle sends an AcceptMsg to its CH in order to add it to its membership table. In the case where a vehicle receives numerous CHMsg messages from its neighbors, it chooses the CH that has the highest weight, then it records all the IDs of the CHs that invited it to join their cluster according to the order of their weight. On receiving an AcceptMsg the CH adds the vehicle in its membership list (Fig. 3).

Algorithm 2: Clustering

```

Wait for a Hello message;
Receive (Hello message);
Calculate its weight;
Get neighbor vehicle weight();
while Clustering phase is active do
    Compare neighbor vehicle weight with its weight ;
    if its weight is the highest among its neighbors weights then
        | Broadcast CHMsg message;
    else
        | Do nothing ;
    end
end

```

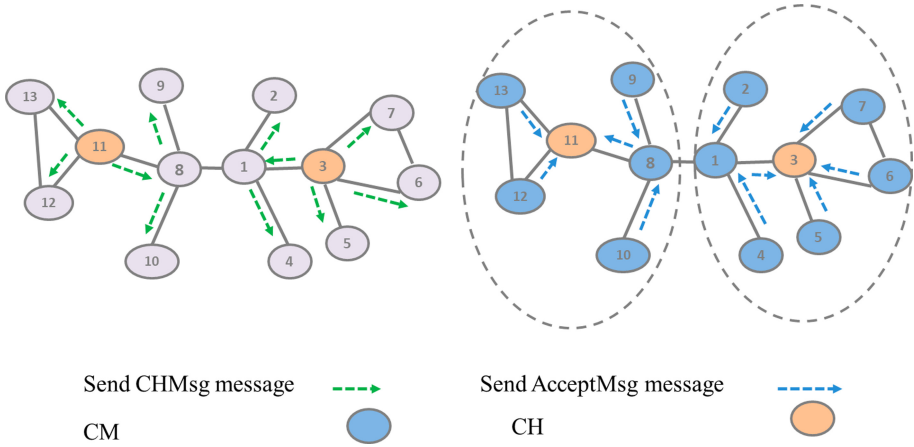


Fig. 3. Clustering phase

6.3 Gateways Selection Process

Each Cluster Head is responsible to choose its gateway nodes, after adding all the nodes and declare them as CM, the CHs broadcast an Invite message to their neighbors, the nodes that receive more than one CHMsg from different CHs are considered as gateway candidates. The gateway candidates respond with GatAccept message. To select the best gateway nodes the CH chooses the candidates that have the highest weight value (Fig. 4).

Algorithm 3: Gateway selection process

```

Wait for CHMsg message;
Receive CHMsg;
if the receiver node receives more than one CHMsg from different CHs then
    the receiver node is gateway candidate;
    Wait for GatMsg message;
    Receive GatMsg;
    Send GatAccept message to CH;
end
    
```

6.4 Maintenance Phase

Three state transitions are possible for a vehicle in the clustering process:

- An ND node wants to join an existing cluster to become CM: in this case, the vehicle broadcasts Hello messages periodically. If the ND node hears a Hello message from a CH neighbor, it sends a JOINMsg message to the CH neighbor. This allows the ND node to join the cluster. If the ND node receives multiple responses from different CHs, the one with the highest weight will be chosen.
- A CM leaves the cluster: If the CH does not receive a Hello message from a specific CM, then this CM is considered to have quitted the current cluster and will be deleted from the list of members.

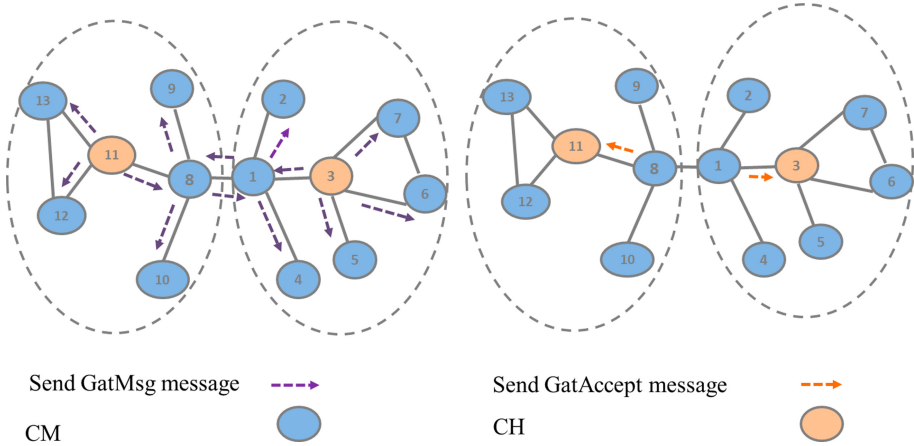


Fig. 4. Gateway selection process

- A CH leaves the current cluster: if during a specified period of time, the CMs do not receive Hello messages from their CH, this CH is supposed to have left the current cluster. The CM joins other clusters saved during the cluster formation phase by sending JOINMsg. Otherwise, if no CH is saved, it broadcasts JOINMsg to discover new CHs (Fig. 5).

Algorithm 4: Gateway selection process

```

vi, vj : vehicles ;
if Vi want to join a new cluster then
    Vi broadcast JOINMsg message;
    if Vi receives one AcceptMsg then
        | Vi ← CM ;
    else
        if Vi receives more than one AcceptMsg then
            | Vi chooses the CH that has the highest weight;
        end
    end
end
if a CH did not receive Hello message from vj for a period of time then
    | CH deletes vj from its list of CMs;
end
if a CM did not receive Hello message from its CH then
    | Send JOINMsg message to the saved CHs ; if no CH is saved then
    | Broadcast JOINMsg message;
    end
end

```

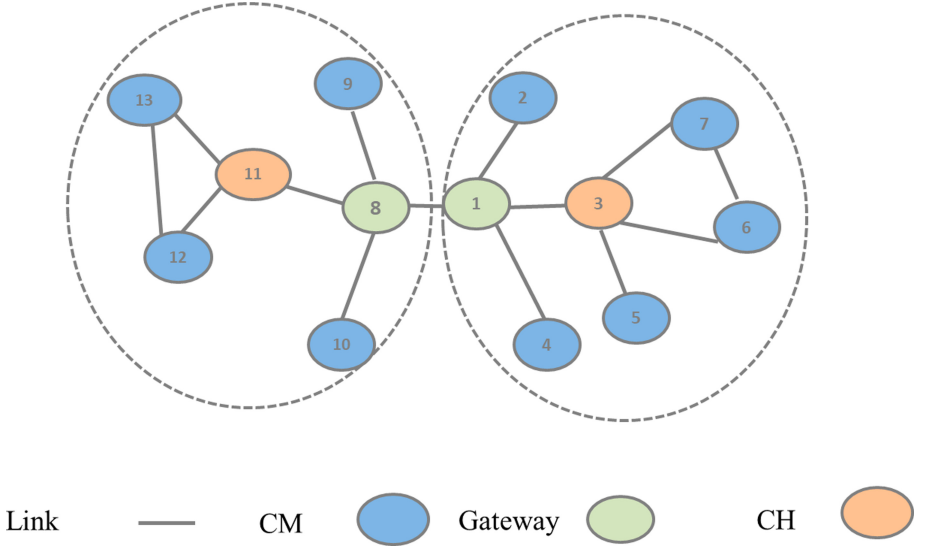


Fig. 5. Clusters formed

7 Simulation and Results

To evaluate the performance and efficiency of our scheme, we use Network Simulator 3 (NS-3) as a network simulator [12] and Simulation of Urban MObility (SUMO) [13] to generate mobility traces of vehicles and road traffic scenarios to imitate a real road network. For Media Access Control (MAC) layer, we used IEEE802.11p. We consider the channel bandwidth as 10 MHz. The proposed scheme is considered to be used in an urban area where the speed is randomly generated and its maximum value is fixed as 80 km/h. The rest of simulation parameters are given in Table 2.

Table 2. Simulation parameters.

Parameter	Value
Simulation time	300 s
Simulation area	1500 × 1500 m
Number of nodes	50–250 nodes
Number of regions	9 regions
Transmission range	250 m
Bandwidth	10 MHz
MAC/PHY protocol	IEEE802.11p
Vehicles' speed	30–80 km/h
Hello message	1 s

Our simulation results concern the following issues:

Average Duration of CH: It is the average period of time taken by an elected CH from its time of selection until it loses its role. Figure 6 presents the Average duration of CH of GICA and MADCCA clustering algorithms. According to the results shown in Fig. 6, for the lowest number of nodes (50 nodes), the average duration of CH of GICA is equal to 100 s, while the average duration of CH of MADCCA is 120 s. As the number of nodes gets higher, GICA algorithm records the highest average duration of CH compared to MADCCA algorithm, which means that the CHs elected using GICA algorithm maintain their CH status more than those elected by MADCCA. The good behavior shown by GICA is due to the effective CH election process that uses final region metric, which in turn ensures stable connections between the CHs and their CMs.

Average CM Duration: It is the time period elapsed from the moment a node joins a cluster and gets its CM status until it leaves the joined cluster. The CM lifetime is an important performance metric to evaluate the efficiency of the GICA and MADCCA protocols. Figure 7 introduces the average duration of CM GICA and MADCCA. The results show that GICA proves a good performance under different numbers of nodes compared to MADCCA. This can be explained by the weight value used by GICA that combines three metrics that ensure network stability (mobility, direction, position).

Overhead: We consider the messages needed for cluster formation and selection of the CH as a communication overhead. It is the amount of information (in bits) circulating in the network per unit time. As shown in Fig. 8, GICA records the

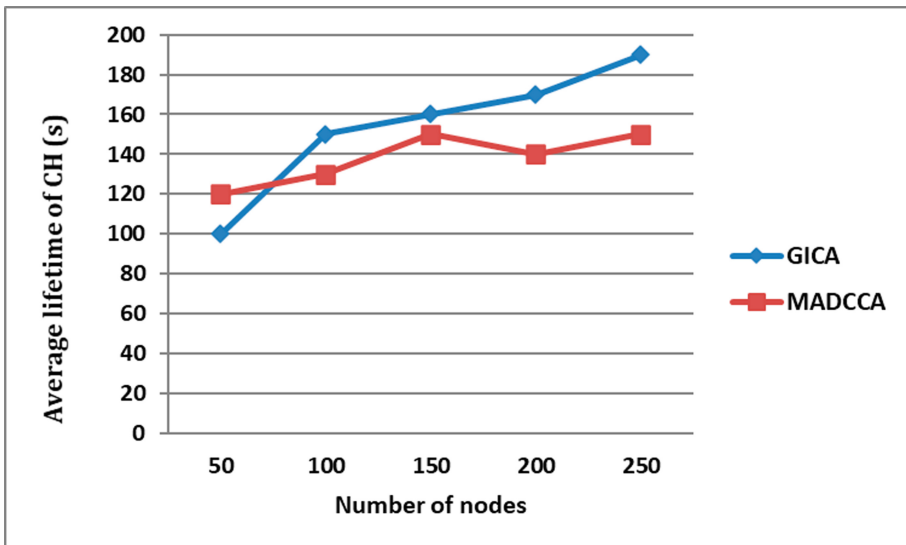


Fig. 6. The average duration of cluster head

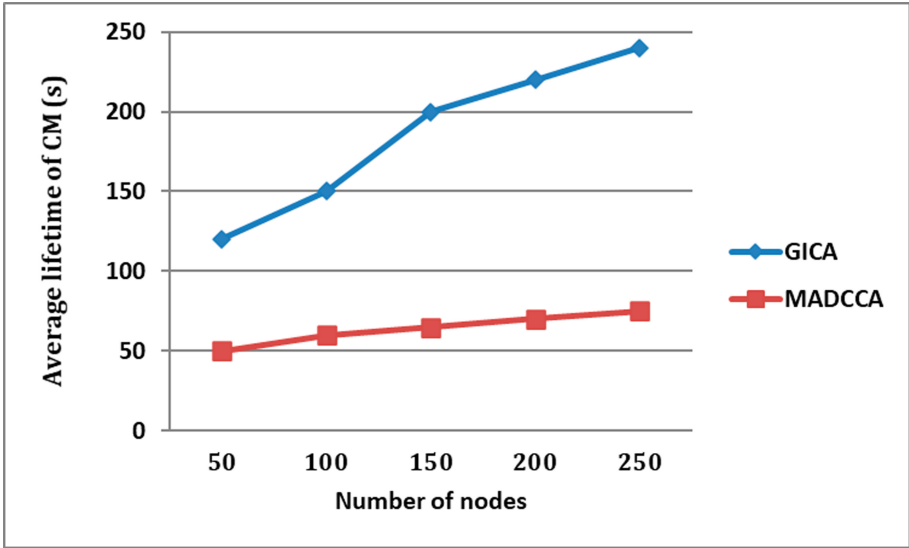


Fig. 7. The average duration of cluster member

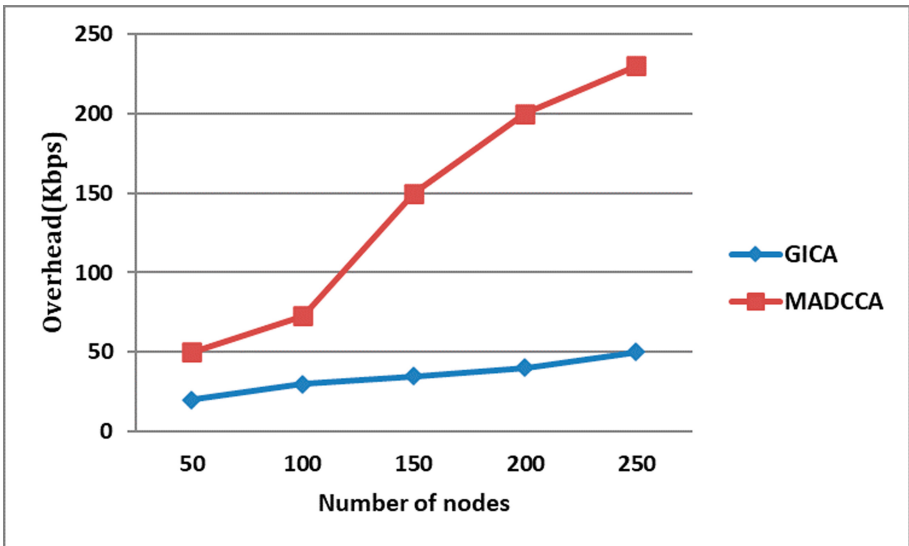


Fig. 8. Overhead of GICA and MADCCA protocols

lowest overhead comparing to MADCCA. This good performance is due to the stability of the nodes elected as CHs by our proposal that minimizes the link disconnection.

8 Conclusion and Perspective

In this paper we propose a new clustering protocol based on geographical information to maintain the communication stability in Internet of Vehicles. The proposed protocol is based on four elementary phases: initialization, Clustering, gateways selection and maintenance phase. To evaluate the performance of our protocol, several simulations were performed while varying the number of nodes using NS3 simulation tool. The obtained results show that our proposal outperforms the MADCCA [8] protocol in terms of Cluster Head lifetime, Cluster Member lifetime, and overhead which confirms that our scheme is well suited for Vehicular networks. To improve our solution in the short and long term, we expect to add more parameters in term of communication delay to minimize the response time.

References

1. Gasmi, R., Aliouat, M.: Vehicular Ad Hoc NETWORKS versus internet of vehicles - a comparative view. In: International Conference on Networking and Advanced Systems (ICNAS), Annaba, Algeria (2019)
2. Contreras-Castillo, J., Zeadally, S., Guerrero-Ibaez, J.: Internet of vehicles: architecture, protocols, and security. *IEEE Internet Things J.* **5**, 3701–3709 (2018)
3. Gasmi, R., Aliouat, M., Seba, H.: A stable link based zone routing protocol (SLZRP) for internet of vehicles environment. *Wirel. Pers. Commun.* **112**(2), 1045–1060 (2020). <https://doi.org/10.1007/s11277-020-07090-y>
4. Kerimova, L.E., et al.: On an approach to clustering of network traffic. *Autom. Control Comput. Sci.* **41**(2), 107–113 (2007). <https://doi.org/10.3103/S0146411607020071>
5. Caballero-Gil, C., Caballero-Gil, P., Molina-Gil, J.: Self-organized clustering architecture for vehicular ad hoc networks. *Int. J. Distrib. Sens. Netw.* **11**, 1–12 (2015)
6. Sethi, V., Chand, N.: A destination based routing protocol for context based clusters in VANET. *Commun. Netw.* **9**, 179–191 (2017)
7. Bellaouar, S., Guerroumi, M., Moussaoui, S.: QoS based clustering for vehicular networks in smart cities. In: Wang, G., Bhuiyan, M.Z.A., De Capitani di Vimercati, S., Ren, Y. (eds.) *DependSys 2019*. CCIS, vol. 1123, pp. 67–79. Springer, Singapore (2019). https://doi.org/10.1007/978-981-15-1304-6_6
8. Ram, A., et al.: Mobility adaptive density connected clustering approach in vehicular ad hoc networks. *Int. J. Commun. Netw. Inf. Secur.* **9**, 222 (2017)
9. Dutta, A.K., Elhoseny, M., Dahiya, V., Shankar, K.: An efficient hierarchical clustering protocol for multihop internet of vehicles communication. *Trans. Emerg. Telecommun. Technol.* **31**(5), e3690 (2019)
10. Senouci, O., Aliouat, Z., Harous, S.: DCA-DS: a distributed clustering algorithm based on dominating set for internet of vehicles. *Wirel. Pers. Commun.* **115**, 401–413 (2020). <https://doi.org/10.1007/s11277-020-07578-7>
11. Wang, C., Li, J., He, Y., Xiao, K., Hu, C.: Segmented trajectory clustering-based destination prediction in IoVs. *IEEE Access* **8**, 98999–99009 (2020)

12. Riley, G.F., Henderson, T.R.: The *ns-3* network simulator. In: Wehrle, K., Güneş, M., Gross, J. (eds.) *Modeling and Tools for Network Simulation*, pp. 15–34. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12331-3_2
13. Behrisch, M., et al.: SUMO-simulation of urban mobility: an overview. In: *The Third International Conference on Advances in System Simulation*, pp. 63–68 (2011)