Distributed Clustering in Vehicular Networks

Leandros A. Maglaras Dimitrios Katsaros
Department of Computer & Communication Engineering, University of Thessaly, Volos, Greece

Abstract—Clustering in VANETs is of crucial importance in order to cope with the dynamic features of the vehicular topologies. Algorithms that give good results in MANETs fail to create stable clusters since vehicular nodes are characterized by their high mobility and the different mobility patterns that even nodes in proximity may follow. In this paper, we propose a distributed clustering algorithm which forms stable clusters based on force directed algorithms. The simulation results show that our Spring-Clustering (Sp-Cl) scheme has stable performance in randomly generated scenarios on a highway. It forms lesser clusters than Lowest-ID and it is better in terms of Cluster stability compared to Lowest-ID and LPG algorithms in the same scenarios.

I. INTRODUCTION

For exchanging information about the current driving situation - traffic or weather conditions, hazard areas or road conditions - vehicles form a spontaneous network, known as a vehicular ad hoc network (VANET), even though the aid of fixed infrastructure [1] can be used. Due to the distributed network nature many messages are generated describing the same hazard event, hence, these messages can be combined to a single aggregate message through clustering. Since VANETs have a very limited capacity, it is desired that the number of messages can be reduced e.g. using aggregation. To reduce the number of aggregators, single messages are not broadcasted through the whole network, however, they are contained in a given area around the hazard event location. Only vehicles inside this area receive single messages and aggregate them. The vehicles outside this area are informed about the hazard event by the aggregate messages only. To further reduce the number of messages in a network, aggregate messages can be aggregated also.

In order to perform aggregation, several clustering techniques have been introduced, while other clustering algorithms for MANETs are also used (Figure 1). Cluster leaders also called clusterhead are assigned special operations like regulation of channel use, data aggregation and message routing between cluster members and between clusters.

Exchange of information between vehicles can be either V2V or vehicle-to-roadside (V2R). Forming V2V-based VANETs has some advantages as compared with the V2R-based VANETs. First, the V2V-based VANET is more flexible and independent of the roadside conditions, which is particularly attractive for the most developing countries or remote rural areas where the roadside infrastructures are not necessarily available. Also, V2V-based VANET can avoid the fast fading, short connectivity time, high frequent hand-offs, and so forth caused by the high relative-speed difference between the fast-moving vehicles and the stationary basestations. However, the link qualities in V2V communications can also be very bad due to multipath fading, shadowing, and Doppler shifts caused by the high mobility of vehicles. V2V communication can be used as the basic means of communication between vehicles and Roadside units may help in places of high vehicle density.

In our clustering scheme only V2V communication between vehicles is considered. The combination of V2V with Roadside units in urban areas with high traffic, where RSUs take control of nearby clusters acting as clusterhead is a matter of future enhancement for our Spring-clustering scheme.

II. VANET CLUSTERING ALGORITHMS

One of the many challenges for VANETs is the dynamic and dense network topology, resulting from the high mobility and high node-density of vehicles [2] especially in urban environments. This dynamic topology causes routing difficulties. A clustered structure can make the network appear smaller and more stable in the view of each vehicle.

A well-known mobility-based clustering technique is MOBIC [3], which is an extension of the Lowest-ID algorithm [4]. In Lowest-ID, each node is assigned a unique ID, and the node with the lowest ID in its two-hop neighborhood is elected to be the cluster head. This scheme favors nodes with lower identifiers to become CHs without taking in mind mobility patterns of the nodes.

In MOBIC, an aggregate local mobility metric is the basis for cluster formation instead of node ID. The node with the smallest variance of relative mobility to its neighbors is elected as the cluster head. The relative mobility for a certain node is estimated by comparing the received power of two consecutive messages from each neighboring node which is...
not a so easy task in high dense environments. If we consider certain scenarios where attenuation of radio signals inevitably exists, the power of signals for estimating mobility can be very limiting and may provide inaccurate measurements. In cluster maintenance also a clusterhead is not guaranteed to bear a low mobility characteristic relative to its members. As time advances the mobility criterion between cluster members is somewhat ignored. If mobile nodes move randomly and change their speeds from time to time, the performance of MOBIC may be greatly degraded.

Many clustering methods have been introduced lately which aim at establishing stable clusters, where clusterhead reelection is reduced. In DDVC clustering (DLDC) [5] the Doppler shift of communication signals is used in order to create clusters. Affinity propagation is an algorithm for image processing, and APROVE has proved that its distributed case can be utilized for VANETs [6]. In Distributed group mobility adaptive clustering (DGMA) [7] group mobility information which contains group physical center’s coordinates, group size, group velocity is used for clustering. Density based clustering is based on a complex clustering metric which takes into account the density of the connection graph, the link quality and the road traffic conditions [8]. Blum et al. [9] proposed a Clustering for Open IVC Networks (COIN) algorithm where cluster-head election is based on vehicular dynamics and driver intentions; Zhang in [10] proposed a DSRC multi-channel-based clustering scheme.

Finally, a large number of sensor node clustering algorithms have been proposed in the literature, e.g., [11] and the references therein, but these are inappropriate for our environment since they assume stationary nodes.

III. SYSTEM OVERVIEW AND ASSUMPTIONS

The idea is based on force-directed algorithms. The force-directed assign forces among the set of edges and the set of nodes in a network. The most straightforward method is to assign forces as if the edges were springs and the nodes were electrically charged particles. The entire graph is then simulated as if it were a physical system. The forces are applied to the nodes, pulling them closer together or pushing them further apart.

Every node applies to its neighbors a force $F_{rel}$ according to their distance and their velocities. Vehicles that move to the same direction or towards each other apply positive forces while vehicles moving away apply negative forces. Components of the vector $F_{rel}$ along the east-west $F_x$ and north-south $F_y$ axes are calculated. In order to form stable clusters only vehicles that move to the same direction or towards each other are considered as candidate cluster members.

For a specific vehicle that the total magnitude of forces applied to it is negative no clustering procedure is triggered since all the surrounding nodes tend to move away from it. Calculating total force $F$ helps to avoid re-clustering in many situations - for example, when groups of vehicles move away from each other.

To illustrate this, consider figure 3 where three snap shots of a specific scenario are presented. It is assumed that nodes participating to a cluster are determined by nodes shape. Vehicles participating to a group are represented by a square and a free node is represented by a circle. In the left snapshot a group of vehicles that move along the east - west axis move towards a vehicle $i$ which is moving along the north-south axis. In the center snapshot the vehicles meet and in the right snapshot the group of vehicles move away from vehicle $i$. At the top of the figure nodes are reclustered when the vehicles meet. The total force applied to node $i$ the different time slots is positive for the two first moments and negative for the third. Vehicle $i$ according to $Sp-cl$ is encouraged to form a cluster with the group of vehicles that move along different directions, in order to exchange useful information.

In the occasion where vehicle $i$ meet the group of nodes as shown in figure 3 at the bottom of the figure no reclustering from free node $i$ would be triggered since the total force applied to it would be negative for all the time period. The negative relative force that is applied to node $i$, represents the fact that node $i$ is moving away from the group of nodes and thus the moment of meeting will be very short and changing cluster structure at this moment may lead to another re-clustering, immediately after nodes move outside their transmission range. Also this short time of period will not be enough for the exchange of useful data.

All nodes are equipped with GPS receivers and On Board
Units (OBU). Location information of all vehicles/nodes, needed for clustering algorithm is collected with the help of GPS receivers. The only communications paths available are via the ad-hoc network and there is no other communication infrastructure. The Maximum Transmission Range (R) of each node in the vehicular network environment is 250 meters.

A. Neighborhood identification

Neighbourhood identification is the process whereby a vehicle/node identifies its current neighbours within its transmission range. For a particular vehicle, any other vehicle that is within its radio transmission range is called a neighbour. All vehicles consist of neighbour set which holds details of its neighbour vehicles. The neighbours set is always changing since all nodes are moving. The neighbour set $N_i$ of vehicle $i$ is dynamic and is updated frequently. Every moving node keeps track of all current neighbors (their id’s) the current and the past distance.

Generally, neighbour node identification is realized by using periodic beacon messages. The beacon message consists of node Identifier (ID), node location, speed vector in terms of relative motion across the axes of $x$ and $y$ ($dx, dy$) total force $F$, state and timestamp. Node location is used in order to calculate the distance between the nodes. Each node informs other nodes of its existence by sending out beacon messages periodically.

All nodes within the transmission range of source/packet carrier node will announce their presence by sending beacon messages frequently. After the reception of a beacon, each node will update its neighbour set table. For a neighbour that already exists in its neighborhood only the current and the past distance are updated. If a node position is changed, then it will update its position to all neighbours by sending a beacon signal. If a known neighbour, times out, it will be removed from the neighbour set table. The total number of neighbors of a given vehicle is called the “active neighborhood set” $N_i$ of the vehicle.

B. Clustering process and protocol structure

As described in the previous section the beaconing thread is responsible for exchanging informations between neighboring nodes. Another task of this thread is processing and proper use of the messages received from other nodes. Each node constantly updates knowledge about neighboring nodes. Each node $i$ using the information of the beacon messages calculates the pairwise relative force $F_{relij}$ for every neighbor applied to every axes $j$ using the coulomb law.

$$F_{relij} = \kappa_{ij} \frac{q_i q_j}{r_{ij}^2}, \quad F_{relijy} = \kappa_{ijy} \frac{q_i q_j}{r_{ij}^2}$$

where $r_{ij}$ is the current distance among the nodes $k_{ij}$ ($k_{ijy}$) is a parameter indicating whether the force among the nodes is positive or negative depending on whether the vehicles are approaching or moving away along the corresponding axis and $q_i$ and $q_j$ may represent a special role of a node (e.g. best candidate for Cluster head due to being close to an RSU, or due to following a predefined route (bus)).

In coulombs law a positive force implies it is repulsive, while a negative force implies it is attractive. In our implementation a positive force symbolizes the fact that the specific pair of nodes is approaching or is moving towards the same direction while a negative force is applied to nodes that move to different directions. Every node computes the accumulated relative force applied to it along the axes $x$ and $y$ and the total magnitude of force $F$. According to the current state of the node and the relation of its $F$ to neighbor’s $F$, every node takes decisions about clustering formation, cluster maintenance and role assignment.

A node may become a clusterhead if it is found to be the most stable node among its neighborhood. Otherwise, it is an ordinary member of at most one cluster. When all nodes first enter the network, they are in non-clustered state. A node that is able to listen to transmissions from another node which is in different cluster can become a gateway. We formally define the following term: relative mobility parameters $k_{ijx}$ and $k_{ijy}$.

**Definition 1:** Relative mobility parameters $k_{ijx}$ and $k_{ijy}$ between nodes $i$ and $j$, indicate whether they are moving away from each other, moving closer to each other or maintain the same distance from each other. To calculate relative mobility, we compute the difference of the distance at time, $t$ and the possible distance at time, $t + dt$ for every axis.

Relative mobility at node $i$ with respect to node $j$ is calculated as follows:

We calculate the distance at every axes between the nodes at time $t$ and the possible distance at time $t + dt$ according to,

$$D_{cxij} = x_i - x_j, \quad D_{fxij} = x_i + dx_i - x_j - dx_j \quad (2)$$

$$D_{cyij} = y_i - y_j, \quad D_{fyij} = y_i + dy_i - y_j - dy_j \quad (3)$$

The relative movement $dx$ and $dy$ of every vehicle along the axes $x$ and $y$ are calculated by the vehicles OBU according to previous data received from the GPS with respect to the traffic ahead (figure 4). According to mobility in every axis relative mobility $k_{ijx}$ and $k_{ijy}$ are calculated according to :

$$if \quad D_{cxij} \leq D_{fxij} \quad then \quad k_{ijx} = -a_x dt. \quad (4)$$

$$if \quad D_{cxij} \geq D_{fxij} \quad then \quad k_{ijx} = a_x dt. \quad (5)$$

where $a_x$ and $a_y$ are given by

$$if \quad D_{cxij} \leq D_{fxij} \quad then \quad a_x = D_{fxij} - D_{cxij} \quad (6)$$

$$if \quad D_{cxij} \geq D_{fxij} \quad then \quad a_x = \frac{1}{D_{cxij} - D_{fxij}} \quad (7)$$

Parameters $a_x$ and $a_y$ indicate the significance of the force applied between the vehicles by reflecting the ratio of divergence or convergence among moving nodes. In equation 6 $a_x$ is proportional to the divergence among nodes, since the faster it takes place the more negative the force must be. In equation 7 $a_x$ is proportional to the reverse difference of the distance among the nodes, since nodes that approach
each other in a fast pace wont probably stay in contact for a sufficient amount of time in order to form cluster and exchange information.

D. Cluster-head election parameters

Vehicles use beacon messages in order to broadcast information to neighboring nodes such as Identifier (ID), node location, speed vector in terms of relative motion across the axes of x and y (dx, dy) total force F, state and timestamp. Using this information as stated above nodes calculate the forces applied to each other according to position and relative mobility. The mobility information of the neighbors is needed for the vehicle to initiate the cluster formation request, while cluster-head election information for any node is limited to the nodes that are within range. After receiving information of all neighboring vehicles, node i calculates

\[
F_x = \sum_{j \in N_i} F_{rel ij x} \quad \text{and} \quad F_y = \sum_{j \in N_i} F_{rel ij y} \quad (9)
\]

which is the total force along axes x and y applied to it. The total value of forces (norms) is calculated for every node according to :

\[
F = |F_x| + |F_y| \quad (10)
\]

Total force F is used to determine the suitability of a vehicle to become clusterhead according to the following criteria:

- The suitability value of the vehicle is calculated by considering the mobility information of its neighbors (parameters \(k_{ij x}\) and \(k_{ij y}\))
- Nodes having higher number of positive neighbors (\(F_{rel ij x} \geq 0\) \(F_{rel ij y} \geq 0\)), maintaining closer distances to their neighbors, should have be qualified to be elected as cluster-heads.

E. The cluster formation algorithm

In order to execute the algorithm, each vehicle is assumed to maintain and update the \(N_i\) set. At any time each vehicle \(i\) recalculates total F and according to total non-clustered members within range try to form a cluster and become the clusterhead.

If the node has the biggest positive force applied to it and there exist at least one free node in its neighborhood, it declares itself to be a clusterhead. In the opposite situation, where there exist a free node \(j\) with biggest total F in range the vehicle becomes a cluster member of \(j\). This algorithm leads to the formation of clusters which are at most two hops in diameter.

F. Cluster maintenance

The cluster maintenance procedure follows the following rules:

- For every free node.

When a standalone (non-clustered) vehicle comes within R distance from a nearby cluster-head, the cluster-head and the vehicle compare the total force \(f\) applied to them. If the relative force \(F\) of the clusterhead is bigger than that of the free node then the cluster-head will accept the vehicle and will add it to the cluster members list. If the clusterhead has smaller \(F\) then no action is triggered.
If there other nearby standalone vehicles it compares the values of $F$ in order to form a new cluster.

- For every member node.
  - If a member node at a certain time finds itself to have bigger $F$ than any of the surrounding clusterheads then it becomes a free node and tries to form its own cluster. When a cluster member moves out of the clusterhead’s transmission range, it is removed from the cluster members list maintained by the cluster-head and it becomes a free node again.

- For every clusterhead.
  - When two cluster heads come within each other’s transmission ranges and they stay connected over a time period CCI the cluster merging process takes place. The clusterhead with the lower $F$ gives up its cluster-head role and becomes a cluster-member in the new cluster.

IV. SIMULATION AND PERFORMANCE EVALUATION

An extensive simulation study was conducted to evaluate the performance of our protocol using a custom simulator. In our simulation, we consider different road traffic and different network data parameters. The simulation environment is a one direction 5-lane highway with a turn in order to evaluate the performance of the scheme.

The total length of the highway is 2 Km. The stationary LPGs created in each scenario are of size comparable to communication range of nodes, i.e. if the communication range of the vehicles is 80 meters each LPG is of 160 meters long as if the RSU has range of 80 meters.

A. The mobility model

The arrival rate of the vehicles follows the Poison process with parameter $\lambda$. The speed assigned to the vehicles is according to the lane it chooses to follow according to Table I.

<table>
<thead>
<tr>
<th>Lane</th>
<th>Speed km/h</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>80</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>120</td>
</tr>
<tr>
<td>4</td>
<td>140</td>
</tr>
<tr>
<td>5</td>
<td>160</td>
</tr>
</tbody>
</table>

TABLE I

SPEED PER LANE.

The density of the vehicles depends on parameter $\lambda$. The number of vehicles per lane is between (2 -15 v/km/Lane) depending on the speed being used and the value of parameter $\lambda$ according to Table II.

<table>
<thead>
<tr>
<th>parameter $\lambda$</th>
<th>v/km/Lane</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>8-15</td>
</tr>
<tr>
<td>5</td>
<td>5-9</td>
</tr>
<tr>
<td>7</td>
<td>3-6</td>
</tr>
<tr>
<td>9</td>
<td>2-5</td>
</tr>
</tbody>
</table>

TABLE II

DENSITY PER LANE.

The basic concepts of Lowest id are the following. Each node is given a distinct ID and it periodically broadcasts the list of its neighbors (including itself). A node which only hears nodes with ID higher than itself is a “clusterhead” (CH). The Lowest-ID node that a node hears is its clusterhead, unless the Lowest-ID specifically gives up its role as a clusterhead (deferring to a yet lower ID node). A node which can hear two or more clusterheads is a “gateway”. Otherwise, a node is a free node.

The basic architecture feature of the stationary LPG is to use a GPS-based grid to partition roadways into zip code areas that define LPGs. In stationary LPG all LPG areas are location based and well defined. Members of LPG dynamically change as vehicles move along the highway.

We compare the three methods under the same environment variables. Each simulation run repeated times with different random seeds and the collected data was averaged over those runs.

Snapshots of the simulation of the three methods are shown in Figures 7, 8 and 9.

Cluster stability

In order to evaluate the stability of the algorithm, we measure the stability of the cluster configuration against vehicle
mobility. In a high dynamic VANET, nodes keep joining and leaving clusters along their travel route. Good clustering algorithms should be designed to minimize the number of cluster changes of the vehicle by minimizing reclustering. This transitions among clusters are measured in order to evaluate the performance of the algorithm. The basic transition events the vehicle encounters during its lifetime:

- A vehicle leaves its cluster and forms a new one (becomes a clusterhead).
- A vehicle leaves its cluster (due to communication range) and joins a nearby cluster or becomes free.
- A cluster-head merges with a nearby more stable cluster.

We compare the average transition events of the vehicles for the Sp-Cl, Low-Id and LPG methods when different speeds and different transmission ranges are used. From Figure 10, we can see that the average transitions produced by our Sp-Cl technique is smaller compared to that produced by the Low-Id and LPG methods. This means our technique causes less number of cluster transitions for all different density topologies. Similar figures were produces for other transmission ranges.

The figures show that the average transitions of the vehicle decreases as the transmission range increases. This is because increasing the transmission range, increases the probability that a vehicle stay connected with its cluster-head.

**Number of clusters**

Due to high dynamics of the VANET, clusters are created (new clusters added to the system) and dissolved over time. The total number of clusters created over a period of time is a metric of the stability of the clustering method used. Good clustering algorithms should be designed to reduce the rate at which clusters are created and added to the system due to the mobility of the nodes. The ability of the clustering method to maintain cluster structure despite vehicles mobility defines its performance. In this simulation we counted the new clusters which are added to the system.

Figure 11 shows that the total number of clusters created by Sp-Cl is always smaller compared to that produced by the Lo-Id method and this number decreases as the transmission range increases. This is because the Sp-Cl method uses the accumulated forces among as a parameter to create the clusters. Thus, the clusters are more stable and have longer lifetime.

**Cluster lifetime**

The average cluster lifetime is an important metric that shows the performance of the clustering algorithm. The cluster lifetime is directly related to the lifetime of its cluster-head. The cluster-head lifetime is defined as the time period from the moment when a vehicle becomes a cluster-head to the time when it is merged with a nearby cluster.

The average cluster lifetime produced by the Sp-Cl and the Low-Id methods is evaluated in various topologies with different transmission ranges, and the results are illustrated in Figure 12.

**V. Conclusions**

Clustering can provide large-scale Vanets with a hierarchical network structure to facilitate routing operations. In this paper, we proposed a distributed clustering algorithm which forms stable clusters based on force directed algorithms. We proposed a mobility metric based on forces applied between nodes according to their current and their future position and their relative mobility.

The force applied between the vehicles reflects the ratio of divergence or convergence among them. We have simulated Sp-Cl and the results show that the performance of Sp-Cl is better than other existing algorithms. It also creates lesser and more stable clusters in order to achieve high scalability. The clusterhead change is relatively low and the overall performance of the method is stable to different topologies and transmission ranges.

**References**


Fig. 10. Average cluster change per vehicle for Sp-Cl, LPG and Low-Id methods for different transmission ranges (125m, 80m).

Fig. 11. The average total number of formed clusters for Sp-Cl and Low-Id methods for different transmission ranges (125m, 80m).

Fig. 12. The average cluster lifetime Sp-Cl and Low-Id methods for different transmission ranges (125m, 80m).