Annals of Telecommunications (2021) 76:581–603
https://doi.org/10.1007/s12243-020-00831-x

A review of clustering algorithms in VANETs

Mengying Ren[1] · Jun Zhang[2] · Lyes Khoukhi[3] · Houda Labiod[2] · Véronique Vèque[4]

Received: 19 May 2020 / Accepted: 17 December 2020 / Published online: 2 February 2021

Abstract

In vehicular ad hoc network (VANET), lots of information should be delivered on a large scale in a limited time. Meanwhile, vehicles are quite dynamic with high velocities, which causes a large number of vehicle disconnections. Both of these characteristics lead to unreliable information transmission in VANET. A vehicle clustering algorithm, which organizes vehicles in groups, is introduced in VANET to improve network scalability and connection reliability. However, different clustering techniques and algorithms are required for different scenarios, such as information transmission, routing, and accident detections. This paper explores the vehicle clustering techniques from the aspects of cluster head selection, cluster formation, and cluster maintenance procedures. Meanwhile, context-based clustering algorithms are summarized, and the hybrid-clustering algorithms are highlighted. The paper also summarizes the existing clustering performance metrics and performance evaluation approaches.

Keywords VANET · Cluster · Algorithm

1 Introduction

Vehicular ad hoc network, known as VANET, is a self-organizing network formed by a collection of moving vehicles [1, 2]. With the rapid development of automotive manufacturing, vehicles are becoming more and more intelligent and powerful. Vehicles can communicate with other vehicles directly through wireless technologies in a V2V (Vehicle to Vehicle) manner or indirectly in a V2I (Vehicle to Infrastructure) or I2V (Infrastructure to Vehicle) manner [3]. Moreover, vehicles can also connect with motorcycles via V2M (Vehicle to Motorcycle) communication and pedestrians via V2P (Vehicle to Pedestrians) communication. Collectively, these wireless connections are referred to as V2X (vehicle to everything) communication, which can support numerous use-cases, including both safety and non-safety applications.

For all of these applications, the most important and investigated one is the safety application. ETSI, the European Telecommunications Standards Institute, has provided a classification for road safety applications in the standard [4]. To support these services in VANET, the information should usually be transmitted with low latency and high accuracy. However, the large number of mobile vehicles and the dynamic network topology of VANET cause scalability and unstable connection problems. One efficient solution to solve these challenges is clustering, which virtually organizes vehicles into groups, named clusters [2].

In past years, many protocols were proposed based on clustering algorithms in mobile ad hoc networks (MANETs), like [5, 6], etc., aiming to save network resources and to increase network efficiency. According to the predictable mobility and predefined road topology of

✉ Mengying Ren
renmengying.0804@gmail.com

Jun Zhang
jun.zhang@telecom-paris.fr

Lyes Khoukhi
lyes.khoukhi@ensicaen.fr

Houda Labiod
houda.labiod@telecom-paris.fr

Véronique Vèque
veronique.veque@u-psud.fr

1 University of Technology of Troyes, Troyes, France

2 Telecom Paris, Institut Polytechnique de Paris, Palaiseau, France

3 Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, Caen, France

4 Université Paris Saclay, CNRS Centrale Supélec, Laboratoire des Signaux et Systèmes, Gif-sur-Yvette, France

VANETs, researchers found that previous clustering algorithms in MANETs were no longer suitable for VANETs. Due to the time of achieving the clustering procedure, additional control overheads may be added. Therefore, an effective clustering algorithm should form a minimum number of clusters, as well as dynamically maintain the cluster structure without causing large overhead over the network. Moreover, an efficient cluster maintenance scheme is required to avoid unnecessary cluster re-formations.

This paper gives an overview of the existing clustering algorithms in VANETs from the following aspects: cluster topology formation, cluster head selection, cluster maintenance, and cluster performance evaluation. A complete description of the mechanisms in each clustering step is provided in the paper. Moreover, detailed comparison and discussion of various approaches and their performance metrics are presented.

2 Overview

Clusters in VANETs are virtual groups of vehicles. Each cluster has at least one cluster head (CH) that is the leader of the group. A CH is usually followed by several vehicles, called cluster members (CMs). Cluster topology can be classified into two categories based on the distance between CH and CM: one-hop clusters and multi-hop clusters. One-hop clusters are usually constructed based on CH's transmission range (TR). CHs add their one-hop neighbors into clusters. Each CM can communicate directly with its CH; two different CMs can communicate directly with each other or indirectly via their CH. Multi-hop clusters were proposed later than one-hop clusters, in which not all CMs can communicate with their CH directly.

2.1 Previous surveys

In recent years, several surveys were published, summarizing the existing clustering algorithms applied in vehicular ad hoc networks (VANETs). In the year 2012, the researchers have synthesized the current clustering algorithms for VANETs from the year 2000 to 2012 in [7]. The authors described the clustering algorithms independently and compared the following contents: clustering metrics, radius (cluster topology), location services, cluster density, and simulators. The authors stated that there is a lack of a fair comparison of different clustering algorithms, because the simulation scenarios, network simulators, and performance metrics change substantially in the different clustering algorithm. However, this survey is considered incomplete without providing the classification and detailed comparison and analysis.

In the year 2014, another survey [1] has been published, covering more research works. The authors presented a detailed classification of clustering in VANETs based upon various vital parameters. The survey divided the existing clustering algorithms as follows: predictive clustering, backbone-based clustering, MAC-based clustering, traditional clustering, hybrid clustering, and secure clustering. The authors described the algorithms precisely in each category and provided a comprehensive comparison and discussion. Compared to the previous survey, this work is much more complete and meaningful; however, the parameter comparisons among different algorithms are not persuasive enough only with the fuzzy words "low," "medium," and "high."

The most recent survey of VANETs clustering techniques was published in the year 2016 [8]. Compared with other existing surveys, it applied a complete taxonomy of VANETs clustering techniques, especially from the aspects of applications and the problem of evaluating and performance comparison of clustering algorithms, which have not been covered in previous surveys. Moreover, the authors precisely summarized the strategies for each clustering procedure step, including cluster head selection, cluster formation, and cluster maintenance. It provides a lot of comprehensive analysis and comparisons; however, the authors did not analyze the scenarios that the algorithms applied, and there is also a lack of performance metrics comparison, which is an essential part when comparing clustering algorithms. Table 1 compares the existing surveys from different aspects.

2.2 Main contributions

Compared with previous research work, this survey focuses on the following main contributions:

- We summarize and compare the existing surveys by highlighting their advantages and limitations. As far as we know, this is the first survey of clustering algorithms that compares the current related surveys;
- We provide an overview of clustering algorithm development in VANETs from the year 1999 to the year 2020, which has never been summarized before;
- We explore more recent research works than previous survey articles, providing a new classification according to the context, and highlighting the hybrid clustering algorithm since it plays a more vital role with the development of network access technologies;
- We summarize and classify the existing clustering techniques in terms of clustering process: cluster head selection, cluster formation, and cluster maintenance;
- We present a comprehensive analysis of the performance evaluation methods of clustering algorithms,

Table 1 Comparison of previous surveys

Comparison aspects		Survey 2012 [7]	Survey 2014 [1]	Survey 2016 [8]	This survey
Overview	Number of algorithms	19	43	45	66
General	Algorithm classification		✓	✓	✓
	Application			✓	✓
	Architecture				✓
Clustering techniques	Cluster head selection	✓	✓	✓	✓
	Cluster formation			✓	✓
	Cluster maintenance			✓	✓
Performance evaluation	Performance metrics				✓
	Simulator			✓	✓
	Traffic scenario				✓
	channel model			✓	✓
Conclusion and perspective		✓	✓	✓	✓

including performance metrics, simulation tools, and traffic scenarios. A new classification of clustering performance metrics is proposed;

– We conclude the challenges and the future development of clustering algorithms in VANETs.

3 History of VANET clustering algorithms

The VANET clustering algorithms started to be investigated in the early 1990s, and the number of research works is increasing in recent years. Figure 1 provides a blueprint for the development of clustering algorithms since the early 1990s. The *X*-axis indicates the year, and the *Y*-axis shows the corresponding algorithm name. Meanwhile, the figure presents a relation between these existing clustering algorithms. The black arrow indicates that the following algorithm is compared with the original one in the simulation, and the blue arrow shows that the following algorithm follows the framework of the original one but without comparison.

In Fig. 1, about 51 clustering algorithms are observed from the early 1990s. Around 30% of them were derived from the idea of the earliest Lowest-ID (LID) algorithm [9], originally proposed for MANETs to increase the communication efficiency among mobile nodes. In the year 2005, the authors summarized the existing clustering algorithms in MANETs in [10]. In this article, MANET clustering algorithms were classified based upon different metrics, including DS-based (Dominating Set based), low-maintenance, mobility-aware, energy-efficient, low-balancing, and combined-metrics-based clustering.

Because of the common characteristics between VANETs and MANETs and the increased popularity of VANETs, some MANETs clustering algorithms were deployed to adapt to the particular characteristics of vehicular communications. It is easy to observe that after the year 2005, plenty of clustering algorithms were proposed for VANETs. Furthermore, most of these were derived from the existing MANET clustering algorithms, including DMAC [11], MOBIC [12, 13], K-ConID [14], and PC [15].

Since 2010, numerous VANET applications were explored, and researchers started to design clustering algorithms based on the requirements of solving specific problems instead of purely increasing cluster stability. Besides one-hop cluster topology, multi-hop cluster topology was widely accepted (e.g., K-hop [16], HCA [17]), because of the increased application requirements and the limited range of wireless transmission. Meanwhile, the development of cellular technologies, including UMTS and LTE, expanded the deployment of clustering algorithms in vehicular networks. Hybrid clustering architecture, combining V2V and V2I communication manners, is becoming a new trend, making more applications available in vehicular networks. The rest of this article will provide a more detailed discussion.

Based on the clustering algorithms shown in Fig. 1, the most popular algorithms that were particularly designed for VANETs are selected: PPC [24], ALM [21], k-hop [16], APROVE [20], C-DRIVE [28, 29], CBMAC [25], CCP [26, 27], C-RACCA [22], UFC [18], and MBDC [19]. The criteria of popularity are based on the number of citations and the frequency of this algorithm been compared with other algorithms. We observe the frequency of the article cited from the year 2007 to the year 2017 and analyze the

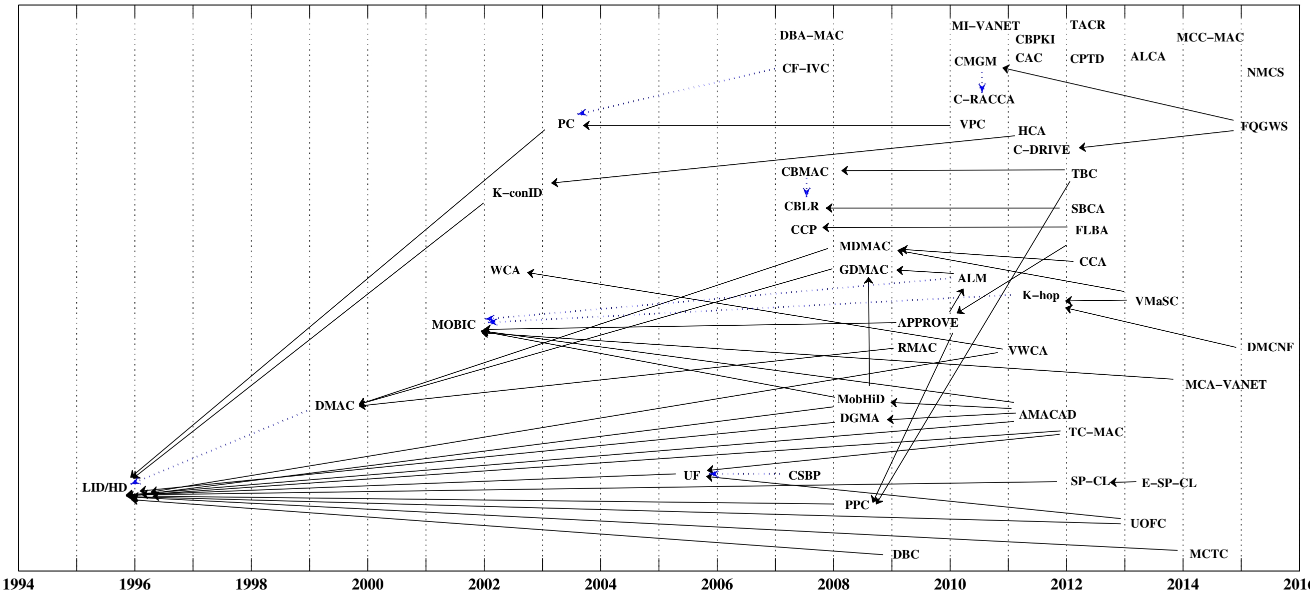

Fig. 1 The development of clustering algorithms in VANETs. (Dashed line: without comparison; Solid line: with comparison)

development trend of clustering algorithms. In Table 2, it is easy to observe that multi-hop cluster topology [16] is becoming popular in recent years.

4 Clustering algorithms in VANETs

Vehicular clusters' construction is a dynamic procedure due to the high mobility of vehicles and intermittent communications. Vehicles have to obtain the necessary information from their neighboring cars, including their identities, positions, and velocities. The potential cluster head will be selected based upon the criteria, such as relative mobility, received signal strength, and link lifetime. Clusters will be established by adding potential vehicle members. In [8], a basic flow of clustering algorithm was introduced, indicating the general steps of clustering algorithms.

This section presents a summarization of the approaches and the criteria of each of the clustering steps, including cluster head selection, cluster formation, and cluster maintenance. For each method, there is a discussion of the advantages and disadvantages.

4.1 Cluster head selection

In the clustering algorithm, the most vital part is to select a reliable leader, which can achieve the highest stability among its vicinities. In most clustering algorithms, CHs are selected at the beginning of the clustering process. Then, the cluster formation process is controlled by the selected CHs.

Table 2 Number of citations

Algorithm	Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Total
UFC [18]	2017	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1	1	9	20	4	35
MBDC [19]	2017	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	8	15	34	12	69
APPROVE-2 [20]	2014	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	10	23	27	21	26	19	2	128
K-hop [16]	2011	N/A	N/A	N/A	N/A	N/A	N/A	1	11	10	15	27	21	16	21	8	130
ALM [21]	2010	N/A	N/A	N/A	N/A	0	5	12	6	14	22	19	17	13	19	5	133
C-RACCA [22]	2010	N/A	N/A	N/A	N/A	3	11	17	20	16	16	17	23	9	13	4	149
APPROVE-1 [23]	2009	N/A	N/A	N/A	1	3	10	30	18	28	21	31	31	18	17	6	215
PPC [24]	2008	N/A	N/A	2	9	14	21	10	13	24	20	22	25	14	21	5	200
CBMAC [25]	2007	N/A	0	0	5	7	5	9	15	11	6	12	4	10	4	5	93
CCP-2 [26]	2007	N/A	0	1	16	15	31	29	33	40	28	37	40	33	17	5	325
CCP-1 [27]	2006	0	5	8	7	6	16	13	8	15	18	11	5	8	7	1	128

This subsection presents a classification of CH selection approaches and analysis.

4.1.1 First Declaration Wins

The First Declaration Wins (FDW) mechanism was first proposed in the passive clustering algorithm [15] in the year 2002. FDW is a cluster head election rule which does not require any metric information. FDW is based on the idea of contention, in which a vehicle, firstly claiming to be a cluster head, dominates the other vehicles within its transmission range.

In CF-IVC [30], the authors proposed a clustering formation protocol for inter-vehicle communication, based on the passive clustering (PC) model proposed in [15]. FDW rule has been applied in this mechanism. The first node, relaying the received packet with its CH claim, wins the competition. Any neighboring nodes that receive the cluster head information will change their state to be CMs.

Another protocol that applied FDW rule for CH election is PassCAR [31]. It proposed a passive clustering aided routing protocol for VANETs. If two nodes are in the CH_{READY} state, the FDW mechanism ensures that only one node will become the CH. Another node in the transmission range of CH will become the CM.

A contention-based CH selection algorithm has been introduced in a Unified Framework of Clustering algorithm (UFC) [18]. In the initial state, every vehicle calculates a contention timer based on its neighbors' information. All of the vehicles start to count down the contention timer at the same time, and the first one finishing the countdown will become the CH and broadcast a CH announcement message.

Discussion The FDW CH selection method can reduce the exchanged number of packets. Initial clusters can be constructed very fast without any additional restrictions. However, the stability of the fast-constructed clusters cannot be guaranteed since there are no criteria to check the link stability between CMs and their CH. The method proposed in UFC [18] solve this problem by filtering out the unstable neighboring vehicles before the CH selection process.

4.1.2 First vehicle in the moving direction

C-DRIVE [32] was firstly proposed in the year 2009, aiming at estimating the density of vehicles on a given road segment. The authors supposed that each car is equipped with a digital map to determine the direction in which it will travel. The first vehicle entering the region in a particular direction will become CH. Each CH is responsible for computing the density of vehicles in its cluster; this information will help the traffic signal to adjust the signal timings and thus to manage the traffic more effectively.

MC-DRIVE [29] is a modified C-DRIVE [32] method, to estimate vehicles' density at the intersection scenario. The first vehicle entering the road is considered as a temporary CH that will lead the cluster formation process.

In [33], the authors proposed a multi-hop cluster-based routing protocol, named CONVOY, for highway scenario. The CH is simply the vehicle moving in the first place of the convoy. The CH leads the cluster formation and controls the size of the vehicle convoy.

Another typical vehicle cluster method is called platoon. Vehicle platooning is a technique where highway traffic is organized into groups of close-following vehicles called platoon or convoy. The idea of organizing traffic in platoons to dramatically increase road capacity was originally proposed in [34] by PATH for Intelligent Vehicle Highway System (IVHS). In 2015, a platoon management protocol for Cooperative Adaptive Cruise Control (CACC) was proposed in [35]. A platoon leader is the first vehicle in the platoon with at least one follower. With CACC technique, the headway can be reduced to further improve the platooning efficiency.

Discussion The approach to select the first vehicle in the moving direction as the CH is easy to realize. This method is more suitable for the intersection area. Once the CH is chosen, it will directly add the moving vehicles behind it as the CMs. However, the selected CH could frequently be overtaken on the road. How to update the CH information and reduce the control overhead is a big challenge.

4.1.3 Weighted sum

In VANETs, CH selection is usually based on multiple mobility metrics, such as speed, position, node degree (number of neighbors), and link lifetime. Each protocol has its cluster stability definitions. The cluster head selection is based on the following two conditions: a vehicle's mobility pattern and the relative mobility respecting to vehicle's neighbors. Mobility-based metrics are widely used in recent VANETs CH selection, such as speed, position, link lifetime, and node degree (usually the number of one-hop neighbors).

According to the research, it is easy to observe that most of the CH selection methods are based on the weighted sum of various parameters. The weight-based clustering method was originally proposed for MANETs in [13]. The CH is selected based on the combined weight value, including different metrics values, and the weighting factors are distributed to each of these metrics.

In [36], the authors proposed a compound clustering scheme based on two groups of parameters: Lowest-ID or Highest Degree, and the traffic specific position or Closest Velocity to Average algorithms, to find a more stable clustering method. A utility function was proposed, presented as the weighted sum (utility) for each vehicle. The vehicle with the highest utility is selected to be the CH. The implementation uses a weight factor 85% to the Lowest-ID or Highest-Degree method and 15% to the traffic-specific information of position or velocity. However, the authors did not specify how to define the weight factor.

DBC [37] proposed a new multilevel clustering algorithm for VANETs based on vehicle's connection level estimation, link quality level estimation, and traffic condition estimation. Connection level estimation is to estimate the vehicle density. Link quality is represented by signal-to-noise (SNR). The notion Group Membership Lifetime (GML) counter is proposed, and it indicates the vehicle's link duration with other group members. GML is used to check the reliability of the vehicle. The vehicle with the most stable communication with other group members gains the highest weight and becomes CH. However, the calculation of weight value has not been introduced in the paper.

In [38], the authors proposed a lane-based clustering algorithm (TC-MAC-1). The main idea is to select a CH based on the lane where most of the traffic appears. It is based on the assumption that each vehicle knows its exact driving lane on the road via a lane detection system and in-depth digital street map. With the knowledge of lane information, vehicles moving on the road can be easily separated into three types: Left Turn, Right Turn, and No Turn. Like the previous weighted sum method, each vehicle computes and broadcasts its CH level, based on the information of network connectivity level, average distance level, and average velocity level. Here, the lane weighting factor is determined based on the ratio of the number of lanes for each traffic flow to the total number of lanes on the roadway.

The authors proposed VWCA clustering algorithm in [39], based on the original WCA [13] and entropy-based clustering algorithm [40]. The authors defined the entropy value of local networks, where the "local networks" denotes a vehicle's neighborhood list. Entropy value has been proved to reduce the frequency of cluster re-affiliation [39] effectively. Besides, the authors proposed three novel parameters for the weight calculation: vehicle's distrust value, relative moving direction, and the number of neighbors. Each vehicle calculates its combined weight value in a distributed manner. The one with the minimum combined weight value is selected to be CH. The authors evaluated the influence of weighting factors on cluster performance. The weighting factors can be dynamically chosen according to their purpose, which is considered a big improvement.

The authors of [41] proposed a destination-based clustering protocol (AMACAD). Each vehicle calculates a weight value between one of its neighbors, $F_{v,z}$, based on relative distance, relative speed, and a relative destination. CH is selected according to the sum of $F_{v,z}$. The influence of weight factors on the number of constructed clusters and CH lifetime were tested. Similar to VWCA [39], the weight factors are dynamically assigned according to the different scenario.

In [42], the authors proposed a fuzzy-logic based CH selection method (FLBA). Each vehicle calculates a stability factor SF based on its average speed difference respecting all of its neighbors. CH selection is based on a weighted stability factor SF_w, represented by the weighted sum of the vehicle's current SF_w and the previous SF_{w-1}. The vehicle with the highest SF_w is selected as CH. The weight factor, called the smoothing factor, is predetermined as 0.5; however, the authors did not explain the reason for this value.

In [43], the authors proposed a multi-agent-based clustering protocol (MDDC). A stability metric was introduced to determine the CH. The elected CH should have a higher connectivity degree (number of neighbors), less average speed, and longer travel time on the lane. The leading vehicle computes the stability factor for all the cluster members. The stability metric was represented by a weighted sum of these parameters. The CM with the highest stability metric is chosen as the CH. The weighting factors are determined by the initiator vehicle. However, the selection of weighting factors has not been described in detail.

The authors proposed an affinity propagation (AP) [44] based clustering algorithm (APROVE) for VANETs in [20]. Every vehicle calculates the responsibility and availability value in a distributed manner. The proposed method is based on position, velocity, and the prediction of vehicles near future position. For each clustering interval, the car calculates its CH convergence value $\mathrm{CH}_{\mathrm{convgi}}$, represented as the sum of the responsibility array and availability array. The vehicle with a positive convergence value will become a CH.

Discussion Weighted sum-based CH selection method has been widely applied during the development of clustering algorithms in VANETs. The selected metrics could vary under different contexts with the assigned weighting factors. The researchers could introduce any metrics if they have an adequate reason. For example, the trust value proposed in VWCA [39]. It guaranteed a comprehensive consideration of various parameters that may influence the clustering performance; however, the weighting factor assignment is still an open issue, which may considerably affect the clustering performance. Only a small part of the research works provided a reasonable approach in weighting factor selection.

4.1.4 Aggregate relative mobility

The aggregate relative mobility-based CH selection approach was first proposed for MANETs in [12], called MOBIC. The authors proposed a relative mobility metric to select the CH. The Received Signal Strength (RSS) between two successive packet transmissions from a neighboring node can represent the relative mobility between two nodes. Each node aggregates the variance of relative mobility value of all its neighbors, and the one that has the lowest aggregate relative mobility value among its neighbors should become CH.

The idea of the aggregate relative mobility in MOBIC [12] is impressive; however, calculating two nodes' relative mobility through the Received Signal Strength is considered unreliable. In [21], the researchers proposed an aggregate local mobility (ALM) clustering algorithm to prolong the cluster lifetime for VANETs. Similar to MOBIC [12], ALM was also based on the aggregate relative mobility; however, it replaced the Received Signal Strength by the relative distance between two nodes.

In [16], the authors proposed another aggregated relative mobility-based clustering algorithm for VANETs, named K-hop. Different from ALM [21], the authors introduced the ratio of packet delivery delay of two consecutive packets to calculate a vehicle's N-hop relative mobility. Moreover, the previous one-hop cluster structure was extended to the *k*-hop cluster structure. The aggregate relative mobility value is the summary of the relative mobility times a weight value for all neighbor nodes in N-hop. The vehicle that has the smallest aggregate relative mobility is selected as the CH.

Discussion Comparing with previous CH selection approaches, the aggregate relative mobility-based approach is dynamic. The relative mobility between two nodes depends on the information of both the last and the current time interval. Theoretically speaking, this method is more precise with vehicles' historical data; however, the appropriation of selecting the time interval is important to ensure the accuracy of vehicles' relative mobilities.

4.1.5 Dominating set based

Dominating set (DS) is a typical technique in the clustering algorithm. A survey about clustering for MANETs [45] has summarized DS-based clustering. The idea of finding a connected dominating set (CDS) for MANETs comes from the fact that any vehicle can communicate with another vehicle in the same CDS. The researchers adjusted the DS-based clustering algorithms to VANETs, and the dominating set nodes are considered as CHs, ensuring the communication among CHs.

In [24], a position-based clustering algorithm (PPC) was proposed to form a stable backbone in a highly dynamic vehicular environment. The authors applied the idea of the minimum dominating set (MDS). The proposed cluster is a 2-hop cluster, and the cluster size is controlled by the predefined maximum cluster radius. If a vehicle has a long travel time and small speed deviation on the road segment, it will have a higher priority to be the CH. To ensure that only one vehicle is selected in the cluster, a hash function has been introduced.

A Hierarchical Clustering Algorithm (HCA) was proposed in [46], which also selected the CHs based upon the concept of DS. The authors formulated the problem of creating 2-hop clusters and scheduling them as distributed G^2 dominating set issue. HCA aimed to find a randomized approach that can benefit both from a small dominating set and fast convergence. A CH is a member of dominating set G^2, and it manages and synchronizes the shared channel access for all other nodes in the formed cluster.

In [47], a routing protocol for VANETs (SCRP) was proposed. The objective is to choose routing paths with minimum end-to-end delay (E2ED) for non-safety applications in urban VANETs. The backbone vehicles are selected as the CDS. These nodes are selected according to vehicles' stability factor, considering the average distance and the speed relationship between a vehicle and its neighbors. The stability factor is also represented as a weighted sum value. The vehicle with the lowest stability factor is added to the backbone.

Discussion Formulating the clustering algorithm for a selection of dominating set is not a new solution both in MANETs and VANETs. The CHs are the vehicles in the dominating set to ensure the connection among CHs; meanwhile, to reduce the redundant clusters.

4.1.6 Other cluster head selection approaches

Apart from the mentioned CH selection approaches, there exist some other interesting clustering algorithms. Ant colony system (ACS), initially proposed in [48], has been extended to solve the clustering problem. In the year 2012 and 2015, researchers have proposed two ACS-based clustering algorithm respectively: TACR [49] and ASVANET [50]. TACR used an ant colony routing technique based on trust, and the CH selection was based on a vehicle's direction, position, and relative speed. Different in the ASVANET algorithm, the CH selection happens after the construction. The node closest to the center of the cluster is selected to be CH. However, the performance of

ASVANET did not show any improvement compared with the Lowest-ID [9] algorithm.

Similar to ASVANET, another center-based CH selection method was proposed in MBDC [19]. The authors introduced the concept of temporary CH, which is the first vehicle moving forward, to assist the cluster construction. The temporary CH will deliver the CH role to the vehicle that is closest to the center of the cluster according to the predefined cluster's radius. However, this method requires a long time for the cluster construction, and the cluster information may be incorrect.

In the year 2020, researchers proposed a novel hesitant fuzzy (HF) multi-criteria ranking framework [51] to select proper cluster-heads. A multiple criteria decision-making (MCDM) method is applied. Vehicles can have a set of possible values, defined as an HF set (HFS). These values are calculated based on the speed difference, position difference, and node connectivity. Instead of choosing CH based on single criteria, [51] provided a multi-criteria-based CH selection method, which shows improved CH stability.

An innovative fuzzy-based CH selection scheme is proposed in [52] to prolong VANET lifetime. Similar to multi-criteria CH selection methods, researchers applied 7 criteria, including vehicle speed, vehicles' velocity, spectrum price, QoS, number of neighbors, centroid, and signal strength. Among these elements, the spectrum price is a new one compared to the existing CH selection criteria. Then, Sugeno fuzzy integral is applied to aggregate a set of criteria to a single one for CH selection. The simulation results show an improved cluster lifetime.

Discussion The CH selection process is the most crucial part of the clustering algorithm. Instead of using a single metric at the beginning, such as the node identification in Lowest-ID [9], more and more CH selection methods are essentially depending on multiple mobility metrics, including relative speed, relative distance, number of neighbors, and link duration between two vehicles. The difference is the approach when combining these primary metrics. Besides, the strategy of CH selection also depends on the context. For instance, in the intersection scenario, like C-DRIVE [29], it is more efficient to choose CHs according to their moving directions.

4.2 Cluster construction

This section will discuss the cluster construction from the aspect of cluster topology. After the CH selection process, clusters are formed according to multiple predefined criteria, including transmission range, cluster radius, and the maximum number of member vehicles. Although the cluster formation criteria vary in each algorithm, there are two cluster topologies of the formed cluster: single-hop cluster and multi-hop cluster, as shown in Fig. 2.

4.2.1 Single-hop cluster

During the research, it is easy to observe that the majority of clustering algorithms establish single-hop clusters. Cluster formation depends on the information of the vehicle's one-hop vicinities. For example, in PPC [24], cluster volume is limited by the cluster radius, which is determined by the average velocity of vehicles, the number of nodes in a cluster, and the number of lanes on the road. In MC-DRIVE [29], a predetermined distance threshold was designed to limit the cluster construction. The distance threshold is calculated with the speed and transmission range of the vehicle approaching the intersection.

In VWCA [39], Adaptive Allocation of Transmission Range (AART) algorithm was proposed. Different from other existing clustering methods in which the transmission range is constant, the AART algorithm allows vehicles to adjust the transmission range dynamically based on current vehicle density. The adaptive transmission range (100–1000 m) is based on the intra-cluster communication standard, Dedicated Short-range Communication (DSRC) standard [53]. The proposed approach not only reduces the number of clusters, especially the single-node clusters but also uses bandwidth resources efficiently. AART algorithm has also been applied in another clustering method proposed in [54], a data envelope analysis (DEA)-based clustering algorithm. In DEA, a vehicle's transmission range can dynamically change based on vehicle density, between 100 and 1000 m according to DSRC standard [53].

In [19], the authors proposed a mobility-based scheme for dynamic clustering in VANETs (MBDC). The cluster formation depends on CH's transmission range. Hence, the established clusters are single-hop clusters.

Discussion Single-hop cluster topology can reduce the cluster formation time and decrease cluster management overhead since fewer information exchanges are required. Many existing clustering algorithms construct single-hop clusters directly according to the CH's transmission range

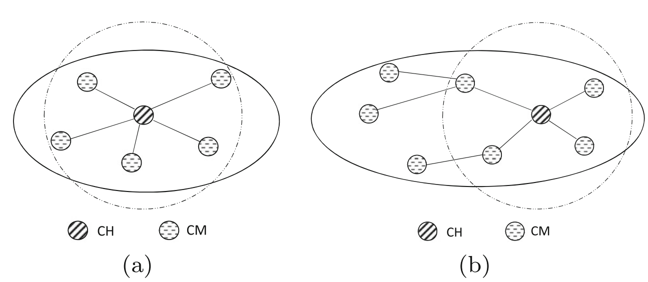

Fig. 2 Cluster topology

or the limited cluster radius. Therefore, the number of members in a cluster will be determined by local traffic density. When vehicle density is very high, collisions could happen in the cluster and would cause a low packet delivery ratio. When vehicle density is very low, a vehicle may not find any neighbors to form a cluster and stays single. Both of these two situations will cause worse cluster performance and should be avoided. The proposed AART algorithm can adjust the transmission range to local vehicle density. Moreover, limiting the maximum and the minimum number of vehicles in a cluster can also solve this problem.

4.2.2 Multi-hop cluster

In recent years, it is easy to observe that the multi-hop cluster topology is becoming the trend in the cluster design. Researchers believe that the structure of multi-hop clusters are more stable with fewer vehicle disconnections and re-affiliation. Moreover, some applications require information transmission on a large scale, and multi-hop cluster topology can increase the information transmission efficiency.

The first multi-hop clustering scheme designed for VANETs was K-hop [16]. Similar to MOBIC [12] and ALM [21], the main idea of cluster construction is based upon aggregate relative mobility. The authors introduced N-hop relative mobility between two nodes. The cluster size is limited by the number of hops between the CH and its farthest member vehicle.

The proposed HCA algorithm [17] is also a multi-hop clustering algorithm. It aims to establish more stable communications among vehicles for time-sensitive message delivery. The maximum distance between CH and CM is two hops. HCA is proposed for fast topology control, which is suitable for VANETs. Meanwhile, it also introduced a channel access method using synchronization. Different from other clustering algorithms, HCA does not rely on the location services. The proposed algorithm was compared with K-ConID [14]. Although HCA showed fewer cluster switches than K-ConID, the message delivery delay was not evaluated, which was inconsistent with its objective.

CCA algorithm [55] is a multi-hop clustering algorithm that is derived from the concept of network criticality. It was claimed as the first work to apply a localized robust graph metric (network criticality) to guide the cluster formation process. Cluster formation relies on local N-hop neighbors' information and link expiration time (LET). The authors evaluated the performance of CCA by setting the number of hops to 1 and two, respectively. The simulation results showed better cluster stability when the number of hops was 2.

The previously mentioned CONVOY [33] is also a multi-hop cluster. The CH controls the cluster size by the predetermined cluster length. The authors have evaluated the effect of cluster length on cluster performance and defined the cluster length as 2 km.

In [56], the authors designed a multi-hop clustering scheme (VMaSC), simulated under a realistic traffic scenario by Simulation of Urban MObility (SUMO) [57]. The scheme aims to provide more stable clusters and to reduce the number of CH in the network. The CH election relies on the calculated relative mobility concerning its neighbors. The cluster size is controlled by the predetermined number of hops. Compared with previous multi-hop clustering algorithms, VMaSC provided a more complete and impressive performance evaluation. It was compared with K-hop [16] for 1-hop, 2-hop, and 3-hop respectively, and showed a better performance regarding CH duration, CM duration, and CH change number.

Another multi-hop clustering algorithm DMCNF was proposed in [58]. The authors designed a neighborhood follow strategy. Each vehicle chooses a one-hop stable neighbor to follow, and the node with more followers and smaller average relative mobility is passively selected to be CH. Similar to HCA [17], each vehicle only needs to update its one-hop neighbors' information, which can reduce the overhead. The authors compared its performance to K-hop ($K = 3$) [16], and DMCNF showed lower cluster overhead. DMCNF did not predefine the number of hops of the cluster; however, the maximum number of hops is four according to the simulation results.

Discussion According to the research, multi-hop clusters show higher cluster stability, especially regarding the number of CH changes, CM re-affiliation, and cluster lifetime. However, multi-hop cluster formation and cluster maintenance are more complicated, which will cost significant control overheads and long cluster formation time. During the research, it can be observed that some algorithms have predetermined the cluster size through the maximum number of hops from CH to CM and the cluster diameter. Simulation results of VMaSC [56] showed that cluster performance became worse when the maximum number of hops is bigger than three. In DMCNF [58], the maximum number of hops was tested to be four and most vehicles were presented in 1-hop and 2-hop clusters without any cluster size limitation. Moreover, according to the research, none of the clustering algorithms has evaluated the cluster construction delay. A significant cluster formation time may cause unexpected information transmission delay.

4.3 Cluster maintenance

Due to the dynamic topology of VANETs, frequent vehicle disconnections and re-connections may cause severe packet loss. Cluster maintenance is indispensable to reduce frequent vehicle re-clusterings and finally achieve a more stable clustering performance. Clustering maintenance includes vehicle leaving, vehicle joining, and cluster maintenance. The maintenance methods will be discussed in this section.

During the clustering process, each vehicle periodically broadcasts Beacon messages with their necessary traffic information to inform their vicinities of updating information. Typically, the CH contains a list of all information about its member vehicles. When it loses the connection with a member vehicle, it deletes this member's information from the list. The disconnected car attempts to find a new cluster to join. On the other hand, when a CH receives a request to join information from a car, it checks whether this car could be its member or not, and sends back a confirmation message, or simply ignores the request.

Compared with the vehicle leaving and joining process, the cluster merging process is more complicated. Cluster merging happens when two or more clusters can be represented by one merged cluster, which can reduce the number of clusters and increase clustering efficiency. Cluster merging conditions are different in clustering algorithms. Usually, cluster merging is triggered when two CHs are in the transmission range of each other like ALM [21]. Some methods, like PPC [24] and CONVOY [33], apply a distance threshold to control cluster merging. To avoid frequent cluster reformations, many methods propose a contention timer to defer the cluster merging process. Another discussed issue in cluster merging is how to choose the new CH for the merged cluster.

In PPC [24], cluster merging happens when the distance between two CHs is detected less than the dismiss threshold. The cluster with fewer members is dismissed to reduce communication overheads while its members join other clusters. Each node of this cluster begins a new registration to join other clusters. The effect of dismissing threshold on cluster reconfiguration rate has been evaluated in this work. However, directly merging a smaller cluster into a bigger one without considering the stability may lead to an unstable merged cluster and higher communication overheads, which is contrary to the intuition.

Recent cluster merging methods take the stability factor into consideration, which is similar to the CH selection method. In ALM [21], the CH which has lower aggregated local mobility (ALM) will become the merged CH. In APROVE [20], after cluster contention time (CCT), the CH with less CH convergence value will relinquish its CH status. In VWaSC-LTE [59], both CHs check the feasibility of becoming the new CH when merging happens. The merged cluster size is controlled by the predefined maximum number of CHs and CMs, and the maximum number of hops. The CH with less average relative speed will give up its CH role. In [60], a similar cluster merging approach is introduced. The CH with lower relative mobility will give up its CH role.

Discussion Due to the dynamic mobility pattern in VANETs, the cluster merging process is indispensable and happens frequently. A good clustering method should smoothly manage the cluster merging process, meanwhile, avoid unnecessary cluster re-affiliations. Most algorithms applied a contention timer to defer cluster merging, ensuring that only the relatively stable clusters can be merged. The merged CH is selected from the two merging CHs, which can guarantee most of the link connections with CMs. However, none of the previous clustering algorithms has analyzed the impacts of the cluster merging process on clustering performance. In [61], the authors summarized the existing cluster merging methods and proposed a leadership-based clustering merging scheme (LCM). Moreover, a fair comparison of different merging schemes was given.

5 Classification of clustering algorithms

5.1 Clustering for context-based applications

In recent years, clustering mechanisms are applied for specific VANET applications. Cluster nodes are treated as backbone nodes for information dissemination in VANETs. In this case, defining a proper clustering mechanism should be based on a specific context. Numerous context-based clustering algorithms have been proposed in recent years. Tables 3, 4, and 5 provide the classification of the existing clustering algorithms observed in this survey according to their context.

5.1.1 Information transmission

CB-BDP [67] proposed a cluster-based beacon dissemination process, aiming at providing a local map of vehicles' neighbors which could be used for safety applications. The idea of CB-BDP is to combine a cluster-based MAC and inter-cluster coloring scheme. The channel access can be synchronized between neighboring clusters. Firstly, the CH aggregates beacons in its cluster, using the intra-cluster aggregation protocol. Secondly, neighboring CHs exchange their cluster status using the proposed multi-hop inter-cluster communication protocol. Then, all of CHs broadcast the aggregated and exchanged information to all their CMs through the proposed intra-cluster dissemination protocol, providing each CM with a local map of its neighbors. The authors used a cluster coloring scheme and combined some connected clusters to form super-clusters, where each

Table 3 Summary of application-dependent clustering algorithms

Context		Algorithm	Year	Cluster performance		Network performance
				Macroscopic	Microscopic	
Application	Pure clustering	DMCNF [58]	2015	1, 2, 3	6	11, 16
		E-SP-CL [62]	2013	1, 3	7	
		UOFC [63]	2013	1, 3	6	
		VMaSC-1 [56]	2013	1, 2,	6	16
		CCA [55]	2012	1, 2, 3, 4	6, 7	
		SP-CL [64]	2012	1, 3	7	
		TBC-2 [65]	2012	1, 3	7	16
		FLBA [42]	2012	1, 2, 4		
		K-hop [16]	2011	1, 2	6	
		ALM [21]	2010	1, 3	7	
		DBC [37]	2009	2, 3, 4	6	
		MDMAC [66]	2008	3	6	16
		PPC [24]	2008			15
		AMACAD-2 [41]	2011	1, 2, 3	10	
		APROVE-2 [23]	2009	1, 2, 3	6	16
		UF [36]	2005		6	
		HCA [17]	2011	1, 3	7	
	Information transmission	CB-BDP [67]	2015			12
		CFT [68]	2017	2, 4		15
		SPC [60]	2016	1, 3	7	
		MCTC [69]	2014	1, 3, 4	9	
		MCA-VANET [70]	2014		6	
		TC-MAC-3 [71]	2013			
		MCMF [72]	2013			14, 15
		CSBP [73]	2007			14
		C-DRIVE [28]	2009			13
		LTE4V2X-3 [74]	2012			12, 13
	Routing protocol	CONVOY [33]	2013	1, 3	7, 9	
		PassCAR [31]	2013	3		12, 15
		TACR [49]	2012			12, 16
		CAC [75]	2011			12, 13, 15, 16
		MI-VANET [76]	2010			12, 15
		VPC [77]	2010			12, 14, 15
		RMAC [78]	2009	1		10
		CBLR [79]	2004			12, 14, 16
		PC [15]	2003			14, 15
	Traffic density estimation	MC-DRIVE [29]	2011	3		16
		CB-TIG [80]	2014	1, 3,		
	Security	ALCA [81]	2013	1, 2, 5		15
		VWCA [39]	2011	1, 2		12
		CBPKI [82]	2011	1, 3, 4, 5		
		SSC-HMT [83]	2018	2		12, 16
	Traffic safety	SRB [84]	2012			12, 14, 15
		C-RACCA [22]	2010			13, 14
	QoS	SBCA [85]	2012	1		12, 16
		CCP [26, 27]	2006, 2007	1, 4		13, 14, 15, 16
	Aggregation	CASCADE [86]	2015			12, 14, 15

Table 3 (continued)

Context	Algorithm	Year	Cluster performance		Network
			Macroscopic	Microscopic	performance
	SCB-INIA [87]	2015	4		16
Target tracking	PBC-TT [88, 89]	2014, 2017	1, 2,	6	12, 14, 16
Topology discovery	CPTD [90]	2012	3		11, 16
Traffic prediction	TC-OTP [91]	2012	3		
	CANI [92]	2019	1, 2, 3	8	16

The numbers in this table indicate the corresponding performance metrics in Table 7

super-cluster is colored independently. The CHs in the same super-cluster can always communicate with each other.

MCTC [69] is proposed for relay selection. The CH is selected as the vehicle that has the highest link connection value among its neighbors. The information exchanging is in a carry and forward fashion, taking into consideration vehicles moving in different directions. The Cluster Relay (CR) is selected by CH. The CR will carry and forward the information. The CM with the lowest relative speed is chosen as CR.

In 2014, the author proposed a multi-homing-based clustering method [70] for urban city scenario. It only relies on the vehicle's ability to send and receive wireless packets. Some redundant connections between vehicles are used to create clusters, aiming to enhance communication reliability. This algorithm is not location service dependent like HCA [17]. The author proposed a new clustering metric that considered redundant CH connections, enabling the support for multi-homing. The value of the metric is represented as an 8-bit unsigned non-overflowing integer counter with the initial value of zero. A higher value means higher vehicle interconnections.

In [60], the authors proposed two algorithms, sociological pattern clustering (SPC) and route stability clustering (RSC). It considers vehicles' social behaviors during the clustering process. The historical trajectories of vehicles can be stored in the Road Side Unit (RSU). The CH selection method is based on virtual force, which is defined based on vehicles' distances and their relative velocities, and in form similar to the Coulomb's law. For each vehicle, it computes the accumulated relative force along the X-axes and Y-axes, and the total node and the relation of its force to its neighbors. The vehicle with the highest force among its neighbors is considered as the most stable node and can become a CH. The social pattern includes the type of vehicles, the tall of the vehicle, the predicted time that the vehicle tends to stay on the main street, and vehicles whose driver behaves statistically smoothly. The relative mobility parameters are calculated based on a vehicle's current position and its future position. This protocol allows vehicles that are moving in opposite directions to be clustered.

CF-IVC [30] proposed a cluster-based inter-vehicle communication method which firstly classifies vehicles into different speed groups. Vehicles will only join a CH of similar velocity. To guarantee the collision-free data exchange among the vehicles during the intra-cluster or inter-cluster communication, the authors introduced a Code

Table 4 Summary of MAC clustering algorithms

Context	Algorithm	Year	Cluster performance	Network	
			Macroscopic	Microscopic	Performance
	CBMAC [25]	2007	1		12
	MCC-MAC [93]	2014			13
MAC	DMMAC [94]	2013	1, 2, 4		14, 16
protocol	TC-MAC-2 [95]	2012			13
	DBA-MAC [96]	2007			13, 14
	CCP [26, 27]	2006	1, 4		13, 14, 15, 16

The numbers in this table indicate the corresponding performance metrics in Table 7

Table 5 Summary of hybrid clustering algorithms

Context		Algorithm	Year	Cluster performance		Network
				Macroscopic	Microscopic	performance
Hybrid clustering	FloatingCar Data(FCD)	FCDOC [97]	2016			14
		GC-VDB [98]	2013	1,3,4		
		LTE4V2X [99]	2011			13,16
	Gateway selection	FQGwS [100]	2015			13,14,15,16
		CMGM [101]	2011			12,13,14,15,16
	Data transmission	VMaSC-LTE [59]	2015	1,2,3	6	12,14,16
		LTE4V2X [74]	2012			13,16
	Collision avoidance	CA-ICA [102]	2013			12,14
	Clustersize optimization	DCSO [103]	2016	4		12
	Uplink transmission	C-HetVNETs [104]	2015			13,14

The numbers in this table indicate the corresponding performance metrics in Table 7

Division Access scheme to assign orthogonal codes to the previously identified vehicles. Vehicles are classified into speed groups based on the velocities in their GPS. In CF-IVC, the exact locations of a vehicle's neighbors are not required when forming a cluster or sending a message.

5.1.2 Routing

A Robust Mobility-Aware Clustering (RMAC) routing approach was proposed in [78]. In RMAC, the CH is selected based on the mobility metrics, including location, speed, and direction. After the vehicle identifies its one-hop neighbors, the neighbors are sorted using the bubble sort algorithm. The clusters are overlapped and a CH can also be a CM of another cluster. Based on the constructed network via the RMAC algorithm, vehicles can exchange their neighbors' information and construct neighboring tables to support geographic routing.

A cluster-based location routing (CBLR) [79] protocol was proposed for V2V communication. It is based on non-positional routing that uses the location information provided by GPS. In CBLR, only the gateway nodes can retransmit the information.

In [70], the authors proposed a multi-homing clustering algorithm (MCA-VANET), a new clustering algorithm with redundant cluster head connections, enabling the support for multi-homing. Different from other clustering algorithms, all vehicles are claimed as CH in the initial state. The CH can change its status to a CM if it has enough CH connected. Therefore, each vehicle has at least one CH to be connected.

5.1.3 Traffic density estimation

The authors proposed a clustering algorithm for traffic density estimation at the intersection in [29]. Different from other clustering methods, MC-DRIVE introduced the imaginary points to control the cluster formation and CH election process. Clusters are formed based on the moving directions at the intersection area.

In [80], the authors proposed a hybrid approach for traffic density estimation on the road, with the assist of RSUs. Based on the estimated traffic density, the authors also introduced a new technique for traffic information generalization in low penetration terms of equipped vehicles. The approach showed higher accuracy on traffic density estimation.

5.1.4 Traffic safety

SRB [84] proposed a cluster-based broadcast protocol for safety applications in VANETs. It aims to reduce the effect of the broadcast storm problem by limiting the number of packet transmissions. By receiving the messages from neighbor nodes, the source node firstly detects the cluster, including neighbors within the distance threshold. CH is selected as the farthest vehicle within the distance threshold.

5.1.5 Traffic prediction

In [91], the authors proposed an online traffic prediction method based on traffic clustering. The authors tested the real traffic data and designed an online neural network-based traffic prediction algorithm, called affinity

Propagation. Each node pair calculates its similarity value. According to the tests, the authors concluded that the average relative error (ARE) decreased when the maximum hop increased from 1 to 4. When the number of hops was bigger than 4, the ARE value increased. Therefore, the maximum number of hops was fixed to 4 during the simulation.

In the year 2019, the authors proposed a Clustering Adaptation Near Intersection (CANI) approach [92] to improve the cluster stability for the intersection scenario. This approach is based on the prediction of vehicle behaviors near an intersection. The cluster formation is adaptive in real-time according to traffic prediction. The authors introduced a prediction model based on the Online Sequential Extreme Learning Machine (OS-ELM) machine learning model, which provides fast and continuous learning. Therefore, more accurate vehicle behavior can be predicted for the upcoming intersection.

5.1.6 Data aggregation

In [86], the authors proposed a method, called CASCADE, for accurate aggregation of highway traffic information. It aims to provide information about the upcoming vehicles. CASCADE uses data compression to provide aggregation without losing accuracy. With data compression, information can be transmitted more efficiently.

In [87], a secure cluster-based in-network information aggregation (SCB-INIA) algorithm was proposed. In SCB-INIA, the authors presented a new security mechanism for traffic efficiency applications that uses HyperLogLog estimators to create bandwidth-efficient integrity proofs. SCB-INIA is claimed to be able to achieve high protection against plausible attacker models, and that it is more bandwidth efficient than a comparably secure security mechanism that does not employ clustering. However, the authors only compared the overhead performance, which is considered inconceivable.

5.1.7 MAC clustering

In [105] and [25], the authors proposed a cluster-based medium access scheme for VANETs (CBMAC), to minimize the effect of hidden stations and to further lead to reliable data transmission. In CBMAC, CH takes over the responsibility to assign the bandwidth to its member vehicles in the cluster. The CH selection method is a weighted-based method, where the node with the lowest weighted sum among its neighbors will be chosen as a CH. Three metrics are considered in the calculation of the weighted sum: connectivity, relative velocity, and relative distance. The basic structure of a time division medium access (TDMA) frame includes three parts: the first part is for the CH to broadcast basic information; the second part is for the CH to assign slots for its CMs, and the last random-access part is for the CMs to transmit data.

DMMAC [94] proposed a distributed multichannel and mobility-aware cluster-based MAC (DMMAC) protocol. Vehicles organize themselves into more stable and non-overlapped clusters through channel scheduling and an adaptive learning mechanism integrated within the fuzzy-logic inference system (FIS). Every vehicle broadcasts its status message with a weighted stabilization factor. The vehicle with the highest weighted stabilization factor among its neighbors will elect itself as CH. The calculation of the weighted stabilization factor is based on the vehicle's average relative speed and the previous weighted stabilization factor. The triangular fuzzier is chosen to implement the FIS system, with the inter-distance and the relative velocity between two vehicles as the input parameters, and the vehicle's acceleration as an output value. In DMMAC protocol, each cluster uses a different sub-channel from its neighbors in a distributed manner to eliminate the hidden terminal problem. The proposed protocol can increase the system's reliability, reduce the time delay for vehicular safety applications, and efficiently cluster vehicles in highly dynamic and dense networks in a distributed manner.

5.2 Hybrid clustering for V2X applications

From the observation, most of the existing clustering algorithms construct clusters in a decentralized way. Vehicles self-organize a non-heterogeneous vehicular network based on IEEE 802.11p. However, according to the highly dynamic network topology, clustering in a decentralized way is not appropriate for VANET when compared to a centralized way, since it creates a large amount of overhead. In recent years, researchers start to focus on V2X communication type with the assist of cellular infrastructure instead of pure V2V communications. Table 6 summarizes the existing hybrid cluster architectures. It can be observed that most of the research works combine IEEE 802.11p with LTE cellular architecture, where the IEEE 802.11p interface is used for V2V communication, and the LTE interface is used for V2I communication. The CH is selected by the base station. Application information is transmitted from the base station to CHs, and CHs broadcast the information to their CMs. From another direction, CHs are responsible for collecting and aggregating the data from their CMs. Then, CHs deliver the aggregated information to the base station. The hybrid clustering approaches summarized in Table 6 serve different applications, including traffic data collection, information dissemination, gateway selection, and accident avoidance. The classification of hybrid clustering algorithms is provided in Table 5.

Table 6 Hybrid cluster architecture

Algorithm	Year	Application	Radius	V2V link	V2I link	Network simulator	Traffic simulator	Traffic scenario
FCDOC [97]	2016	Floating car data application off-loading	One-hop	802.11p	LTE	OMNET++, Veins	SUMO, OpenStreetMap	City map of Rome and New York
DCSO [103]	2016	Cluster size optimization to reduce packet loss	Multi-hop	802.11p	LTE	OMNET++, Veins	SUMO	Highway
VMaSC-LTE [59]	2015	Safety message dissemination	Multi-hop	802.11p	LTE	NS3	SUMO	Straight road
FQGwS [100]	2015	Gateway selection	One-hop	802.11p	LTE-A	NS2	VanetMobiSim	Multiple-lane highway
C-HetVNETs [104]	2015	A framework for performance analysis	One-hop	802.11p	LTE	N/A	N/A	Urban with intersections
GC-VDB [98]	2013	Data collection	One-hop	802.11p	LTE	OPNET	OpenStreetMap SUMO	Highway and urban
CA-ICA [102]	2013	Intersection collision avoidance	One-hop	802.11b	LTE	NS3	VanetMobiSim	Urban with intersections
CMGM [101, 106]	2011	Gateway selection	One-hop	802.11p	UMTS	NS2	N/A	Highway and urban
LTE4V2X [74]	2012	Data collection and data dissemination	Multi-hop	802.11p	LTE	NS3	VanetMobiSim	Highway
LTE4V2X [99, 107]	2011	Floating car data (FCD)	One-hop	802.11p	LTE	NS3	VanetMobiSim	Urban

In [107], the authors proposed a framework for a centralized heterogeneous vehicular network using LTE, called LTE4V2X. All vehicles are assumed to have both the LTE and the IEEE 802.11p interfaces. The clustering process is managed by the central eNodeBs. The framework is designed for floating car data (FCD) application and is claimed to be able to deploy other applications. In [99], the authors analyzed the cluster performance of the proposed LTE4V2X [107], especially under highway scenario. In the last version of LTE4V2X in [74], the authors extended the proposed LTE4V2X framework to a multi-hop communication. Meanwhile, the adaptation of the LTE4V2X framework for a data dissemination application has been introduced.

In [98], the authors proposed greedy-based clustering with velocity and direction restriction and cluster bonus (GC-VDB) scheme, aiming for the extended floating car data (xFCD) collection. In the GC-VDB scheme, the CH vehicle should fulfill the following conditions: (1) Vehicles with the highest number of reachable nodes via V2V communication; (2) Vehicles satisfy the velocity and angle requirements (the velocity and moving direction should be limited); (3) Vehicles with the highest CH score (CS). The CH can perform data compression called xFCD payload, which increases the information efficiency. The scheme was tested in both urban and highway scenarios. However, there was no comparison with other algorithms. The authors proposed a new simulation architecture that allows for the modeling of LTE cells under realistic user mobility. The simulation results have shown that the GC-VDB algorithm increases the cluster's lifetime, as well as decreases the total xFCD payload. In this case, the LTE utilization is significantly reduced.

In FCDOC [97], the authors used a VANET-based multi-hop dissemination logic to send control messages and elect CH, named designated nodes. The designated nodes are used to report vehicular data through LTE communications. The aggregated FCD can be achieved by the representative nodes through the LTE infrastructure. The simulation is based on a multi-layer simulation tool, constructed by SUMO, OMNET++, and Veins. This paper considers the real urban maps of the city centers of Rome and New York.

VMaSC-LTE [59] proposed a cluster-based architecture for VANET safety message dissemination. It is based on the work of VMaSC [56], a multi-hop clustering mechanism, proposed in 2013. The objective of this paper is to reduce the information transmission delay based on a hybrid architecture IEEE 802.11p and LTE. The authors designed a multi-hop clustering algorithm for this hybrid architecture. Compared with the previous pure VMaSC algorithm, the hybrid architecture has shown a higher packet delivery ratio and lower packet transmission delay. However, the overhead has not been analyzed.

In [102], a cluster-based architecture for intersection collision avoidance (CA-ICA) is proposed based on heterogeneous networks. Vehicles approaching the intersection start to broad CAMs; however, packet collisions may happen when the node density is very high. Hence, instead of broadcasting the CAMs directly, the vehicle equipped with an LTE interface will transmit CAMs to a base station and then be forwarded to vehicles on other roads. Since the CAMs should be broadcasted every 100 ms, the clustering algorithm is proposed to reduce the number of data transmissions. The vehicle approaching the intersection acts as the CH. It aggregates the cluster members' information in a single CAM and sends it to the corresponding base station via the LTE interface. Then, the base station will transmit CAM to the CH on the other roads, and these CHs broadcast this information directly to its members through an IEEE 802.11p interface. This heterogeneous network combines IEEE 802.11p and LTE, and it performs a higher packet delivery rate.

In [104], the authors proposed a framework of cluster-based heterogeneous vehicular networks (C-HetVNETs). The performance analysis models of intra-cluster and inter-cluster communications were proposed based on Markov queuing model. Floating car data (FCD) applications are carried in the framework based on the designed model. In C-HetVNETs, two-channel access interfaces are implemented, IEEE 802.11p for intra-cluster communication, and LTE for inter-cluster communication. The evolved NodeB (eNB) selects the vehicle that is closest to the center of the cluster to be CH. Clusters are formed only via one-hop of CH. Similar to the previous GC-VDB approach, CH will aggregate the data packets received from its CMs and can directly communicate with eNB. The analytical models are impressive; however, the authors did not mention which kind of simulators they used.

In DCSO [103], the authors proposed a new heterogeneous clustering algorithm for dynamic cluster size optimization. The authors analyzed the impact of the maximum number of hops between a specific vehicle and its CH on the average cluster size, the data aggregation performance, and the packet loss in the IEEE 802.11p network. Similar to the AATR approach [39], the maximum number of hops should be adjusted to vehicle densities. Then, the authors proposed a new heterogeneous clustering algorithm, delegating the CH selection to the cellular base station. The simulation results showed that DCSO resulted in larger clusters for the same maximum number of hops compared to VMaSC-LTE algorithm [59]. As a consequence, data compression at the CH is more efficient in DCSO.

Discussion With the rapid development of cellular network technologies, hybrid clustering via both V2X and V2X communication is becoming a trend in supporting more vehicular network applications. Information can be delivered by the base station in a large area with lower latency, instead of a pure multi-hop delivering method. However, the requirement of having both the IEEE 802.11p and cellular access interfaces on the vehicle is still a challenge.

6 Performance evaluation

Performance evaluations of clustering algorithms in VANETs are mostly based on network simulators because of the limitation of testing scales in real traffic environments. Moreover, using network simulators can create the same simulation environment when comparing various algorithms. However, different clustering algorithms are usually simulated based on different assumptions, including traffic scenarios, channel models, and vehicle mobility models, and the criteria for evaluation are different. In this case, it is hard to compare different algorithms in a fair method. The rest of this section analyzes the performance evaluation methods of the existing clustering algorithms.

6.1 Performance metrics

According to the observation of existing clustering algorithms, there is a lack of fair comparison among these algorithms. Most clustering schemes aim to increase cluster stability. However, many of them did not explain the term "cluster stability" and the corresponding performance metrics. In this section, the most mentioned performance metrics are summarized in Table 7 and are classified into two categories: cluster performance and network performance. Each performance metric is given a short description and is assigned with an identifier from Arabic numbers 1 to 16.

6.1.1 Cluster performance metrics

Cluster performance metrics describe the performance of clustering algorithms and reflect the stability of backbone nodes in the network. It can be described from the aspects of macroscopic and microscopic cluster performance, as proposed in MBDC [19]. The macroscopic cluster performance describes the overall cluster performance, including CH lifetime, CM lifetime, number of clusters, cluster size, and cluster efficiency. In general, a good clustering algorithm prefers a higher CH lifetime, higher

Table 7 Existing performance metrics

Domain		ID	Performance metric	Description
Cluster performance	Macroscopic performance	1	CH/cluster lifetime	Avrg. time duration from a vehicle becoming a CH to giving up its state
		2	CM lifetime	Avrg. time duration from a vehicle becoming a CM to giving up its state
		3	No. of clusters/CH	Avrg. no. of clusters being formed during the simulation period
		4	Cluster size	Avrg. no. of vehicles in a single cluster
		5	Cluster efficiency	The percentage of vehicles participating in clustering process
	Microscopic performance	6	CH change rate	Avrg. no. of CH changes per unit time
		7	Cluster change rate	Avrg. no. of cluster changes per vehicle in a unit time
		8	State change rate	Avrg. no. of state changes per vehicle in a unit time
		9	Disconnection ratio	Avrg. percentage of disconnected vehicles
		10	CM reconnection ratio	Avrg. percentage of vehicles that re-cluster within a given time
		11	No. vehicles per hop	Avrg. no. of vehicles per hop distance
Network performance		12	Delivery ratio, success ratio	The percentage of vehicles that successfully receive the packets
		13	Collision ratio, packet loss ratio	The percentage of collision packets during packet transmission
		14	End-to-End delay from source to the destination	Avrg. latency of data packets transmitted
		15	Throughput over a communication channel	The rate of successful message delivery
		16	Overhead packets to the total no. of data packets	The ratio of the total no. of control

CM lifetime, fewer constructed clusters, large cluster size, and higher cluster efficiency. However, only the macroscopic cluster performance metrics cannot describe the details of communication links among vehicles in the network. The microscopic cluster performance metrics, including CH change rate, cluster change rate, state change rate, disconnection ratio, CM reconnection ratio, and the number of vehicles per hop, can describe more precisely the cluster stability. From this aspect, a stable cluster should have a lower CH change rate, lower cluster change rate, lower state change rate, smaller disconnection ratio, and higher CM reconnection ratio. Generally, a stable link connection requires the vehicle to maintain its current state for a longer duration. Therefore, a disconnected vehicle is expected to build a new link connection as soon as possible to avoid unnecessary packet loss.

6.1.2 Network performance metrics

Another class of performance metric is network performance, describing the overall network performance, including packet delivery ratio, packet loss rate, collision ratio, end-to-end delay, throughput, and overhead. The network performance metrics are used to evaluate the context-based clustering schemes, such as information dissemination,

Table 8 Comparison of urban simulation scenarios

Algorithm	Network topology	Traffic simulator; network simulator	Transmission range (m)	No. of vehicles (density)	Vehicle velocity (km/h)
SPC [60]	Erlangen, Germany	SUMO; N/A	300	80, 120, 160 veh; default: 120 veh	20–50, default: 42
C-DRIVE [29]	3 * 3 km, 7 inters	N/A; NCTUns	350–200	5–40 veh/lane	30.6–50.4
TC-MAC-1 [38]	2 inters, Norfolk,VA	N/A; NS3	150–300	20% trucks, 80% sedans; 60 veh/lane/km; (interval: 10 m)	Max: 40, 80, 120
CA-ICA [102]	grid, 25 inters, 2 * 2 km, 2 dir, 2 lanes	VanetMobiSim; NS3	100	randomly	Max: 50
C-HetVNETs [104]	8.5 km, 1 eNB	N/A; N/A	N/A	100–400 veh/km (uniformly)	120
CAC [75]	1 * 1 km, between inter: 350 m	N/A; N/A	100–500	300 veh (randomly) (interval: 15 m)	10–100
CBMAC [25, 105]	City map Ulm in Germany	N/A; N/A	500	Day Avrg: 5 veh/km, Max: 20 veh/km; Rush Avrg: 8 veh/km, Max: 40 veh/km	N/A
MDMAC [66]	Washinton D.C., 1.087 * 0.942 km (TIGER)	VanetMobiSim; JiST/SWANS	70	500	Limit: 39.6–111.6; Avrg: 28.8
ALCA [81]	2 * 0.2 km (TIGER), with inters	MobiSim; N/A	200	400	30–50 miles/s
K-hop [16]	1 * 1 km,	N/A; NS2	120	100	36–126
MI-VANET [76]	southern Beijing, with inters, 1.7 * 1 km	VanetMobiSim; N/A	Bus: 400; Car: 150	100–250, 20% buses	0–108
DBC [37]	Washinton urban area, 2 * 3 km (TIGER),	VanetMobiSim; JiST/SWANS	N/A	100, 200, 300, 400, 500 veh	Limit: 39.6–111.6; Avrg: 28.8
AMACAD [41]	1.5 * 1.5 km, blocks of 100 m	N/A; Java developer 10G	100–300	50 veh, 0–5 veh/100 m^2	18–72
VMaSC-LTE [59]	5 km, 2 lanes, 2 dir	SUMO; NS3	200	100 veh	36–126
UOFC [63]	grid, 2 lanes, 4 lanes	iTETRIS; Octave (fuzzy logic)	250	N/A	Max: 30, 50, 70
LTE4V2X [74]	8 km^2	VanetMobiSim	300	100–300 veh	90–145
TC-OTP [91]	Washington D.C.	N/A; N/A	N/A	N/A	N/A
FCDOC [97]	New York, Roma, OpenStreetMap	SUMO; OMNET++	N/A	NY: 70, 96, 110 veh/km^2 RM: 70, 80, 87 veh/km^2	50
MCA-VANET [70]	7.960 * 10.575 km, City Kirchberg	SUMO; NS3, Ovnis	100, 200, 300	395 veh	N/A
MDDC [43]	5 km, 2 lanes/road, 10 roads, 5,10 inters	N/A; C programme	250, 500	10–100 veh (interval: 4 m)	Max: 40, 60, 80 (Min: 10)
HCA [17]	1.3 * 3.2 km	SUMO; OMNET++	200	10–100 veh (4 types)	Max: 18–144

routing, and traffic prediction. An efficient clustering algorithm is inclined to perform a higher packet delivery ratio, lower packet loss rate, lower collision rate, short end-to-end delay, larger throughput, and smaller overhead.

Discussion The performance metrics of each clustering algorithms are summarized in Tables 3, 4, and 5. Clustering algorithms are listed according to their context. Therefore, it is easier to observe which kind of performance metrics should be evaluated according to their context. The metrics are represented by the ID numbers defined in Table 7. It can be summarized that pure clustering algorithms can be evaluated only by the clustering performance since there is no application-related information in the network. On

Table 9 Comparison of highway simulation scenarios

Algorithm	Network topology	Traffic simulator; network simulator	Transmission range (m)	No. of vehicles (density)	Vehicle velocity (km/h)
E-SP-CL [62]	2 dir, 3 lanes/dir	SUMO; Custom	130	8–15, 5–9, 3–6, 2–5 veh/km/lane. 25% trucks	80–120
SP-CL [64]	2 km, 1 dir, 5 lanes	N/A; N/A	80, 125	8–15, 5–9, 3–6, 2–5, veh/km/lane	80–160
PPC [24]	10 km, 2/4 lanes	CORSIM; NS2	250	100 veh	60, 100
UF [36]	N/A	N/A; Traffic simulation 3.0	150	N/A	40–140
TBC [65]	15 km, 5 lanes/dir	N/A; C++	r: 150–300; R: 800–1000	400 (13–21 veh/km/lane)	Avrg: 70, 90, 110; dev: 21, 27, 33
FLBA [42]	8 km, 1 dir, 4 lanes	MOVE, SUMO; NS2	200	50–400 veh/km	80–120
MCTC	10 km, 2/4 lanes	N/A; MATLAB	N/A	250 (25 veh/km, 15 veh/km)	Left lane: 100–130; right lane: 80–110
LTE4V2X-3 [74]	3 lanes/dir, 8 km^2	VanetMobiSim; NS3	300	100–300	90–145
VWCA [39]	2 bands, 3 lanes/band, 2.5 * 0.05 km	N/A; MATLAB	Static: 300; dynamic: 100–1000	10–350 veh	70–120
C-RACCA [22]	N/A	N/A; NS2	150	N/A	54–162
FQGwS [100]	Multi-lane	VanetMobiSim; NS2, MATLAB	250	15–40	N/A
DMMAC [94]	8 km, 1 dir, 4 lanes	SUMO, MOVE; NS2	r: 300, TR: 2.5 * 300	0.05–0.4 veh/m	Uniformly 80–120
CASCADE [86]	100 km	ASH; SWANS	300	500 veh	Max: 108
DCSO [103]	5 km, 3 lanes	SUMO; OMNeT++, Veins	N/A	60 veh	Avrg: 60
ALM [21]	1 km, 2 dir, 2 lanes	SUMO; SIDE/SMURPH	N/A	N/A	LMS: 36, 54, 72, 90, 108
VPC [77]	10 km, 2 lanes	N/A; NS2	250	100–400 veh	Max: 72, 144
NDBC [108]	Real map	TIGER; SWANS++	0–250	50–250 veh	N/A
APROVE [23]	3 * 3 km, 1 dir, 3 lanes, rectangular looped	MOVE, SUMO; NS2	250	100 veh	40, 80, 120, 140
PassCAR [31]	5 km, 3 lanes	MOVE, SUMO; N/A	250	150–350 veh	80, 100, 120
CB-TIG [80]	3 km, 1 dir, 2 lanes	SUMO; OMNeT++, Veins	100–1000	N/A	60–120
TC-MAC [71, 95]	N/A	N/A; NS3	300	195 (5, 12, 21, 50 veh/lane)	Max: 104
MCMF [72]	50 km, 2 dir, 3 lanes/dir	N/A; N/A	500	1000 veh/dir	80–200
RMAC [78]	2 km, 1 dir, 4 lanes	Freeway mobility Generator; NS2	250	25, 50, 75 veh; 12 scenarios;	79.2–129.6
DBA-MAC [96]	8 km, 1 dir, 3 lanes	N/A; NS2	250	200, 400, 600 veh	72–108
CCP [27] [96]	3 lanes/dir, circular loop, 2 km	N/A; N/A	N/A	12, 24, 40 veh/km/lane	72–180
PBC-TT [88, 89]	10 km	N/A; NS2, Tossim	50, 100, 250, 500	50, 100, 150, 200 veh	90–126
SCB-INIA [87]	5 km, 3 lanes	N/A; JiST/SWANS	300	0–500	N/A
SBCA [85]	4 lanes,1 dir	N/A; NS2	300	50, 100, 150 veh	90–126
QuickSilver [109]	3 * 0.2 km	ONE simulator; MATLAB	N/A	60, 100 veh	Avrg: 54
CFT [68]	11 km	N/A; N/A	250–600 (per 50)	5–10 veh/km	60–120
MBDC [19]	15 km, 2 dir, 2 lanes/dir	MOVE, SUMO; NS2	300	100 veh, 50 veh/dir	Max: 36–144
UFC [18]	10 km, 2 dir, 2 lanes/dir	MOVE, SUMO; NS2	200	200 veh, 100 veh/dir	54, 72, 90, 108, 126

the contrary, the application-related clustering algorithms should be evaluated from the aspects of both clustering performance and network performance. In summary, the performance evaluation metrics should strictly depend on the context of clustering algorithms, and these metrics should be clearly defined before simulation evaluations.

6.2 Simulation scenarios

In the existing clustering algorithms, there is a lack of fair comparison between different algorithms since various traffic scenarios are created in different algorithms. The work of UFC [18] and FCC [110] have been proposed to compare clustering algorithms in a protocol-independent or a simulator-independent manner but still have not solved this issue completely. In this section, a detailed comparison of traffic scenarios is presented, including the comparison of the simulator, network topology, transmission range, vehicle density, and velocity. Tables 8 and 9 summarize the traffic scenarios for highway and urban city respectively.

According to Table 9, the road topology is always designed as a straight, multi-lane, 2-direction highway road. In Table 9, 31 clustering algorithms designed for highway scenarios have been summarized; however, 15 of them have not specified the traffic simulator, 4 of them have not addressed the network simulator, and 7 of them only used MATLAB or C++ for network simulation, which is considered as unreliable.

In Table 8, the simulation parameters of 20 clustering algorithms for urban scenarios are summarized. Nine of these algorithms apply real city map for performance evaluation, obtained from TIGER files or OpenStreetMap (OSM). And the transmission range varies from 50 to 300 m, which is a standard value of IEEE 802.11 Physical layer protocol. According to the theory of traffic flow in [111], the transmission range is a function of the local density of vehicles, which is determined by vehicle movement and speed.

7 Conclusion

According to our research, there is a lack of a real testbed for the fair evaluation of different clustering algorithms. As we know, most of the clustering algorithms for VANETs are implemented on the simulator and are evaluated under several assumptions. A real testbed will enable the evaluation of clustering algorithms in small-scale real traffic scenarios. Moreover, as an effective approach to supporting various vehicle applications, clustering by machine learning techniques may improve clustering efficiency and accuracy.

Since clustering algorithms require high cooperation among vehicles, vehicles are required to share their personal information with neighbors. In this case, how to balance collaboration and privacy is still an open issue. Moreover, not all of the sharing information are convinced enough; therefore, how to detect malicious vehicles is very important during the cooperation process.

When considering vehicles' privacy and cooperation security, the performance of the clustering algorithms could be affected. For example, the detection of a malicious vehicle may cause information transmission latency. Therefore, how to balance the clustering performance and clustering security is still an open challenge.

With the development of wireless access technologies, such as LTE and 5G, many hybrid clustering architectures are designed. All of these algorithms assumed that vehicles could have two interfaces, one for V2V communications (such as IEEE 802.11p interface) and another one for V2I communications. However, the feasibility of this design is still an open issue.

A framework of clustering algorithms is needed for different applications. The cooperation model can be changed according to different applications, such as safety applications and non-safety applications. Based on the specific requirements of new ITS applications, the framework of the clustering algorithm can enable various modes of vehicle cooperation.

References

1. Bali RS, Kumar N, Rodrigues JJPC (2014) Clustering in vehicular ad hoc networks: taxonomy, challenges and solutions. Vehicular Communications 1(3):134–152
2. Bali RS, Kumar N, Rodrigues JJPC (2014) Clustering in vehicular ad hoc networks: taxonomy, challenges and solutions. Vehicular Communications 1(3):134–152
3. Zhang Y, Moustafa H (2009) Vehicular networks: techniques, standards, and applications, 2nd edn. Auerbach Publications, Boca Raton
4. ETSI TR 102 638 V1.1.1, Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Definitions
5. Chatterjee M, Das SK, Turgut D (2000) An on-demand weighted clustering algorithm (wca) for ad hoc networks. In: Global Telecommunications Conference, 2000. GLOBECOM'00. IEEE, vol 3. IEEE, pp 1697–1701
6. Tolba FD, Magoni D, Lorenz P (2007) Connectivity, energy and mobility driven clustering algorithm for mobile ad hoc networks. In: Global telecommunications conference, 2007. GLOBECOM'07. IEEE. IEEE, pp 2786–2790
7. Vodopivec S, Bešter J, Kos A (2012) A survey on clustering algorithms for vehicular ad-hoc networks. In: 2012 35th International conference on telecommunications and signal processing (TSP), pp 52–56
8. Cooper C, Franklin D, Ros M, Safaei F, Abolhasan M (Firstquarter 2017) A comparative survey of $VANET$ clustering techniques. IEEE Communications Surveys Tutorials 19(1):657–681
9. Lin CR, Gerla M (1997) Adaptive clustering for mobile wireless networks. IEEE Journal on Selected Areas in Communications 15(7):1265–1275

10. Yu JY, Chong PHJ (First 2005) A survey of clustering schemes for mobile ad hoc networks. IEEE Communications Surveys Tutorials 7(1):32–48
11. Basagni S (1999) Distributed clustering for ad hoc networks. In: Fourth international symposium on parallel architectures, algorithms, and networks, 1999. (I-SPAN '99) Proceedings, pp 310–315
12. Basu P, Khan N, Little TDC (2001) A mobility based metric for clustering in mobile ad hoc networks. In: 2001 international conference on distributed computing systems workshop. IEEE, pp 413–418
13. Chatterjee M, Das SK, Turgut D (2002) Wca: a weighted clustering algorithm for mobile ad hoc networks. Clust Comput 5(2):193–204
14. Chen G, Nocetti FG, Gonzalez JS, Stojmenovic I (2002) Connectivity based k-hop clustering in wireless networks. In: Proceedings of the 35th Annual Hawaii international conference on system sciences, 2002. HICSS, pp 2450–2459
15. Kwon TJ, Gerla M, Varma VK, Barton M, Hsing TR (2003) Efficient flooding with passive clustering-an overhead-free selective forward mechanism for ad hoc/sensor networks. Proc IEEE 91(8):1210–1220
16. Zhang Z, Boukerche A, Pazzi R (2011) A novel multi-hop clustering scheme for vehicular ad-hoc networks. In: Proceedings of the 9th ACM international symposium on mobility management and wireless access, MobiWac '11. ACM, New York, pp 19–26
17. Dror E, Avin C, Lotker Z (2011) Fast randomized algorithm for hierarchical clustering in vehicular ad-hoc networks. In: 2011 The 10th IFIP Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net), pp 1–8
18. Ren M, Zhang J, Khoukhi L, Labiod H, Vèque V (2017) A unified framework of clustering approach in vehicular ad hoc networks. IEEE Trans Intell Transp Sys PP(99):1–14
19. Ren M, Khoukhi L, Labiod H, Zhang J, Vèque V (2017) A mobility-based scheme for dynamic clustering in vehicular ad-hoc networks (vanets). Vehicular Communications 9:233–241
20. Hassanabadi B, Shea C, Zhang L, Valaee S (2014) Clustering in vehicular ad hoc networks using affinity propagation. Ad Hoc Networks 13, Part B:535–548
21. Souza E, Nikolaidis I, Gburzynski P (2010) A new aggregate local mobility clustering algorithm for *vanet*s. In: 2010 IEEE international conference on communications (ICC), pp 1–5
22. Taleb T, Benslimane A, Letaief KB (2010) Toward an effective risk-conscious and collaborative vehicular collision avoidance system. IEEE Trans Veh Technol 59(3):1474–1486
23. Shea C, Hassanabadi B, Valaee S (2009) Mobility-based clustering in *vanet*s using affinity propagation. In: GLOBECOM 2009 - 2009 IEEE Global telecommunications conference, pp 1–6
24. Wang Z, Liu L, Zhou MengChu, Ansari N (2008) A position-based clustering technique for ad hoc intervehicle communication. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 38(2):201–208
25. Gunter Y, Wiegel B, Grossmann HP (2007) Cluster-based medium access scheme for VANETs. In: 2007 IEEE intelligent transportation systems conference, pp 343–348
26. Su H, Zhang X (2007) Clustering-based multichannel mac protocols for qos provisionings over vehicular ad hoc networks. IEEE Trans Veh Technol 56(6):3309–3323
27. Zhang X, Su H, Chen HH (2006) Cluster-based multi-channel communications protocols in vehicle ad hoc networks. IEEE Wirel Commun 13(5):44–51
28. Maslekar N, Boussedjra M, Mouzna J, Houda L (2009) Direction based clustering algorithm for data dissemination in vehicular networks. In: Vehicular Networking Conference (VNC), 2009 IEEE, pp 1–6
29. Maslekar N, Mouzna J, Labiod H, Devisetty M, Pai M (2011) Modified c-drive: clustering based on direction in vehicular environment. In: Intelligent Vehicles Symposium (IV), 2011 IEEE, pp 845–850
30. Kayis O, Acarman T (2007) Clustering formation for inter-vehicle communication. In: Intelligent transportation systems conference, 2007. ITSC 2007. IEEE, pp 636–641
31. Wang S-S, Lin Y-S (2013) Passcar: a passive clustering aided routing protocol for vehicular ad hoc networks. Comput Commun 36(2):170–179
32. Maslekar N, Boussedjra M, Mouzna J, Labiod H (2011) A stable clustering algorithm for efficiency applications in *vanet*s. In: 2011 7th International wireless communications and mobile computing conference (IWCMC), pp 1188–1193
33. Véque V, Kaisser F, Johnen C, Busson A (2013) CONVOY: A new cluster-based routing protocol for vehicular networks, pp 91–129
34. Varaiya P (1993) Smart cars on smart roads: problems of control. IEEE Trans Autom Control 38(2):195–207
35. Amoozadeh M, Deng H, Chuah C-N, Zhang HM, Ghosal D (2015) Platoon management with cooperative adaptive cruise control enabled by *vanet*. Vehicular Communications 2(2):110–123
36. Fan P, Haran J, Dillenburg J, Nelson PC (2005) Ad-Hoc, mobile, and wireless networks: 4th international conference, ADHOC-NOW 2005, Cancun, Mexico, October 6-8, 2005. Proceedings, chapter Cluster-Based Framework in Vehicular Ad-Hoc Networks, pp 32–42. Springer: Berlin
37. Kuklinski S, Wolny G (2009) Density based clustering algorithm for vanets. In: 5th international conference on testbeds and research infrastructures for the development of networks communities and workshops, 2009. TridentCom 2009, pp 1–6
38. Mohammad SA, Michele CW (2010) Using traffic flow for cluster formation in vehicular ad-hoc networks. In: IEEE local computer network conference, pp 631–636
39. Daeinabi A, Rahbar AGP, Khademzadeh A (2011) VwcaL An efficient clustering algorithm in vehicular ad hoc networks. J Netw Comput Appl 34(1):207–222
40. Wang YX, Bao FS (2007) An entropy-based weighted clustering algorithm and its optimization for ad hoc networks. In: Third IEEE international conference on wireless and mobile computing, networking and communications, 2007. WiMOB 2007, pp 56–56
41. Morales MMC, Hong CS, Bang YC (2011) An adaptable mobility-aware clustering algorithm in vehicular networks. In: 2011 13th Asia-Pacific network operations and management symposium (APNOMS), pp 1–6
42. Hafeez KA, Zhao L, Liao Z, Ma BNW (2012) A fuzzy-logic-based cluster head selection algorithm in *vanet*s. In: 2012 IEEE international conference on communications (ICC), pp 203–207
43. Kakkasageri MS, Manvi SS (2012) Multiagent driven dynamic clustering of vehicles in *vanet*s. J Netw Comput Appl 35(6):1771–1780
44. Frey BJ, Dueck D (2007) Clustering by passing messages between data points. Science 315(5814):972–976
45. Yu JY, Chong PHJ (First 2005) A survey of clustering schemes for mobile ad hoc networks. IEEE Communications Surveys Tutorials 7(1):32–48
46. Dror E, Avin C, Lotker Z (2013) Fast randomized algorithm for 2-hops clustering in vehicular ad-hoc networks. Ad Hoc Networks 11(7):2002–2015. Theory, Algorithms and Applications of Wireless Networked RoboticsRecent Advances in Vehicular Communications and Networking

47. Togou MA, Hafid A, Khoukhi L (2016) Scrp: stable cds-based routing protocol for urban vehicular ad hoc networks. IEEE Trans Intell Transp Sys PP(99):1–10
48. Colorni A, Dorigo M, Maniezzo V, et al. (1991) Distributed optimization by ant colonies. In: Proceedings of the first European conference on artificial life, vol 142, pp 134-142. Paris, France
49. Sahoo RR, Panda R, Behera DK, Naskar MK (2012) A trust based clustering with ant colony routing in $vanet$. In: 2012 Third international conference on computing communication networking technologies (ICCCNT), pp 1–8
50. Fathian M, Shiran GR, Jafarian-Moghaddam AR (2015) Two new clustering algorithms for vehicular ad-hoc network based on ant colony system. Wirel Pers Commun 83(1):473–491
51. Chettibi S (2020) Combination of hf set and mcdm for stable clustering in vanets. IET Intelligent Transport Systems 14(5):190–195
52. Alsarhan A, Kilani Y, Al-Dubai A, Zomaya AY, Hussain A (2020) Novel fuzzy and game theory based clustering and decision making for vanets. IEEE Trans Veh Technol 69(2):1568–1581
53. Kenney JB (2011) Dedicated short-range communications (dsrc) standards in the United States. Proc IEEE 99(7):1162–1182
54. Fathian M, Jafarian-Moghaddam AR (2015) New clustering algorithms for vehicular ad-hoc network in a highway communication environment. Wirel Netw 21(8):2765–2780
55. Li W, Tizghadam A, Leon-Garcia A (2012) Robust clustering for connected vehicles using local network criticality. In: 2012 IEEE international conference on communications (ICC), pp 7157–7161
56. Ucar S, Ergen SC, Ozkasap O (2013) Vmasc: Vehicular multi-hop algorithm for stable clustering in vehicular ad hoc networks. In: 2013 IEEE wireless communications and networking conference (WCNC), pp 2381–2386
57. Krajzewicz D, Erdmann J, Behrisch M, Bieker L (2012) Recent development and applications of sumo-simulation of urban mobility. Int J Adv Sys Measurements 5(3&4)
58. Chen Y, Fang M, Shi S, Guo W, Zheng X (2015) Distributed multi-hop clustering algorithm for $vanet$s based on neighborhood follow. EURASIP J Wirel Commun Netw 2015(1):1–12
59. Ucar S, Coleri Ergen S, Ozkasap O (2015) Multi-hop cluster based ieee802.11p and lte hybrid architecture for $vanet$ safety message dissemination. IEEE Trans Veh Technol PP(99):1–1
60. Maglaras LA, Katsaros D (2016) Social clustering of vehicles based on semi-Markov processes. IEEE Trans Veh Technol 65(1):318–332
61. Ren M, Zhang J, Khoukhi L, Labiod H, Vèque V. (2017) A study of the impact of merging schemes on cluster stability in $vanet$s. In: 2017 IEEE 28th IEEE international symposium on personal, indoor and mobile radio communications - PIMRC2017
62. Maglaras LA, Katsaros D (2013) Enhanced spring clustering in vanets with obstruction considerations. In: 2013 IEEE 77th vehicular technology conference (VTC Spring), pp 1–6
63. Tal I, Muntean GM (2013) User-oriented fuzzy logic-based clustering scheme for vehicular ad-hoc networks. In: 2013 IEEE 77th vehicular technology conference (VTC Spring), pp 1–5
64. Maglaras LA, Katsaros D (2012) Distributed clustering in vehicular networks. In: 2012 IEEE 8th international conference on wireless and mobile computing, networking and communications (WiMob), pp 593–599
65. Rawashdeh ZY, Mahmud SM (2012) A novel algorithm to form stable clusters in vehicular ad hoc networks on highways. EURASIP J Wirel Commun Netw 2012(1):1–13
66. Wolny G (2008) Modified $dmac$ clustering algorithm for $vanet$s. In: 2008 Third international conference on systems and networks communications, pp 268–273
67. Allouche Y, Segal M (2015) Cluster-based beaconing process for $vanet$. Veh Commun 2(2):80–94
68. Luo Q, Li C, Ye Q, Luan TH, Zhu L, Han X (2017) CFT: A cluster-based file transfer scheme for highway. ArXiv e-prints
69. Chiti F, Fantacci R, Rigazzi G (2014) A mobility driven joint clustering and relay selection for IEEE 802.11p/WAVE vehicular networks. In: 2014 IEEE international conference on communications (ICC), pp 348–353
70. Vodopivec S, Bešter J., Kos A (2014) A multihoming clustering algorithm for vehicular ad hoc networks. Int J Distributed Sensor Netw 10(3):07085
71. Almalag MS, Olariu S, Weigle MC, El-Tawab S (2013) Peer-to-peer file sharing in $vanet$s using $tc - mac$. In: 2013 IEEE international conference on pervasive computing and communications workshops (PERCOM workshops), pp 84–89
72. Venkataraman H, Delcelier R, Muntean Gabriel-Miro (2013) A moving cluster architecture and an intelligent resource reuse protocol for vehicular networks. Wirel Netw 19(8):1881–1900
73. Fan P (2007) Improving broadcasting performance by clustering with stability for inter-vehicle communication. In: 2007 IEEE 65th vehicular technology conference - VTC2007-Spring, pp 2491–2495
74. Rémy G, Senouci SM, Jan F, Gourhant Y (2012) Lte4v2x - collection, dissemination and multi-hop forwarding. In: 2012 IEEE international conference on communications (ICC), pp 120–125
75. Cheng ST, Horng GJ, Chou CL (2011) Using cellular automata to form car society in vehicular ad hoc networks. IEEE Trans Intell Transp Syst 12(4):1374–1384
76. Luo J, Gu X, Zhao T, Yan W (2010) $mi - vanet$: a new mobile infrastructure based vanet architecture for urban environment. In: 2010 IEEE 72nd vehicular technology conference - fall, pp 1–5
77. Wang S-S, Lin Y-S (2010) Performance evaluation of passive clustering based techniques for inter-vehicle communications. In: The 19th annual wireless and optical communications conference (WOCC 2010), pp 1–5
78. Goonewardene RT, Ali FH, Stipidis E (2009) Robust mobility adaptive clustering scheme with support for geographic routing for vehicular ad hoc networks. IET Intell Transp Syst 3(2):148–158
79. Santos RA, Edwards RM, Edwards A (2004) Cluster-based location routing algorithm for vehicle to vehicle communication. In: Proceedings. 2004 IEEE radio and wireless conference (IEEE Cat. No.04TH8746), pp 39–42
80. Arkian HR, Atani RE, Pourkhalili A, Kamali S (2014) Cluster-based traffic information generalization in vehicular ad-hoc networks. Veh Commun 1(4):197–207
81. Kumar N, Chilamkurti N, Park JH (2013) Alca: agent learning—based clustering algorithm in vehicular ad hoc networks. Personal Ubiquitous Comput 17(8):1683–1692
82. Gazdar T, Benslimane A, Belghith A (2011) Secure clustering scheme based keys management in $vanet$s. In: 2011 IEEE 73rd vehicular technology conference (VTC Spring), pp 1–5
83. Oubabas S, Aoudjit R, Rodrigues JJPC, Talbi S (2018) Secure and stable vehicular ad hoc network clustering algorithm based on hybrid mobility similarities and trust management scheme. Veh Commun 13:128–138
84. Vegni AM, Stramacci A, Natalizio E (2012) Srb: A selective reliable broadcast protocol for safety applications in $vanet$s. In: 2012 international conference on selected topics in mobile and wireless networking (iCOST), pp 89–94

85. Ahizoune A, Hafid A (2012) A new stability based clustering algorithm (*sbca*) for *vanet*s. In: 37th Annual IEEE conference on local computer networks - workshops, pp 843–847
86. Ibrahim K, Weigle MC (2008) *cascade*: cluster-based accurate syntactic compression of aggregated data in *vanet*s. In: 2008 IEEE Globecom workshops, pp 1–10
87. Dietzel S, Peter A, Kargl F (2015) Secure cluster-based in-network information aggregation for vehicular networks. In: 2015 IEEE 81st vehicular technology conference (VTC Spring), pp 1–5
88. Khakpour S, Pazzi RW, El-Khatib K (2014) A prediction based clustering algorithm for target tracking in vehicular ad-hoc networks. In: Proceedings of the Fourth ACM international symposium on development and analysis of intelligent vehicular networks and applications, DIVANet '14. ACM, New York, pp 39–46
89. Khakpour S, Pazzi RW, El-Khatib K (2017) Using clustering for target tracking in vehicular ad hoc networks. Vehicular Communications 9:83–96
90. Zhang L, El-Sayed H (2012) A novel cluster-based protocol for topology discovery in vehicular ad hoc network. Procedia Computer Science 10:525–534
91. Zhang B, Xing K, Cheng X, Huang L, Bie R (2012) Traffic clustering and online traffic prediction in vehicle networks: a social in uence perspective. In: 2012 Proceedings IEEE INFOCOM, pp 495–503
92. Alsuhli GH, Khattab A, Fahmy YA, Massoud Y (2019) Enhanced urban clustering in vanets using online machine learning. In: 2019 IEEE International conference on vehicular electronics and safety (ICVES), pp 1–6
93. Yang F, Tang Y, Huang L (2014) A multi-channel cooperative clustering-based *mac* protocol for *vanet*s. In: 2014 wireless telecommunications symposium, pp 1–5
94. Hafeez KA, Zhao L, Mark JW, Shen X, Niu Z (2013) Distributed multichannel and mobility-aware cluster-based mac protocol for vehicular ad hoc networks. IEEE Trans Veh Technol 62(8):3886–3902
95. Almalag MS, Olariu S, Weigle MC (2012) *tdma* cluster-based mac for *vanet*s ($tc - mac$). In: 2012 IEEE international symposium on a world of wireless, mobile and multimedia networks (WoWMoM), pp 1–6
96. Bononi L, di Felice M (2007) A cross layered *mac* and clustering scheme for efficient broadcast in *vanet*s. In: 2007 IEEE international conference on mobile Adhoc and sensor systems, pp 1–8
97. Salvo P, Turcanu I, Cuomo F, Baiocchi A, Rubin I (2016) *lte* floating car data application off-loading via *vanet* driven clustering formation. In: 2016 12th Annual conference on wireless on-demand network systems and services (WONS), pp 1–8
98. Ide C, Kurtz F, Wietfeld C (2013) Cluster-based vehicular data collection for efficient lte machine-type communication. In: 2013 IEEE 78th vehicular technology conference (VTC Fall), pp 1–5
99. Rémy G, Senouci SM, Jan F, Gourhant Y (2011) Lte4v2x - impact of high mobility in highway scenarios. In: Global information infrastructure symposium - GIIS 2011, pp 1–7
100. collab=G. e. m. Zhioua, N. Tabbane, H. Labiod, S. T. (2015) A fuzzy multi-metric qos-balancing gateway selection algorithm in a clustered vanet to lte advanced hybrid cellular network. IEEE Trans Veh Technol 64(2):804–817
101. Benslimane A, Taleb T, Sivaraj R (2011) Dynamic clustering-based adaptive mobile gateway management in integrated vanet - 3g heterogeneous wireless networks. IEEE Journal on Selected Areas in Communications 29(3):559–570
102. Tung LC, Mena J, Gerla M, Sommer C (2013) A cluster based architecture for intersection collision avoidance using heterogeneous networks. In: 2013 12th Annual Mediterranean Ad Hoc Networking Workshop (MED-HOC-NET), pp 82–88
103. Garbiso J, Diaconescu A, Coupechoux M, Leroy B (2016) Dynamic cluster size optimization in hybrid cellular-vehicular networks. In: 2016 IEEE 19th International conference on intelligent transportation systems (ITSC), pp 557–563
104. Zheng Q, Zheng K, Sun L, Leung VCM (2015) Dynamic performance analysis of uplink transmission in cluster-based heterogeneous vehicular networks. IEEE Trans Veh Technol 64(12):5584–5595
105. Guenter Y, Wiegel B, Grossmann HP (2007) Medium access concept for $VANET$s based on clustering. In: 2007 IEEE 66th Vehicular technology conference, pp 2189–2193
106. Taleb T, Benslimane A (2010) Design guidelines for a network architecture integrating vanet with 3g amp; beyond networks. In: 2010 IEEE Global telecommunications conference GLOBECOM 2010, pp 1–5
107. Remy G, Senouci SM, Jan F, Gourhant Y (2011) $lte4v2x$: Lte for a centralized *vanet* organization. In: 2011 IEEE Global Telecommunications Conference - GLOBECOM 2011, pp 1–6
108. Fan P, Mohammadian A, Nelson PC, Haran J, Dillenburg J (2007) A novel direction-based clustering algorithm in vehicular ad hoc networks. In: Transportation research board 86th annual meeting, number 07-1673
109. Crepaldi R, Bakht M, Kravets R (2012) Quicksilver: Application-driven inter- and intra-cluster communication in vanets. In: Proceedings of the third ACM international workshop on mobile opportunistic networks, MobiOpp '12. ACM, New York, pp 69–76
110. Zhang J, Ren M, Labiod H, Khoukhi L (2017) Every dog has its day: a comparative study of clustering algorithms in vanets. In: 2017 IEEE symposium on computers and communications (ISCC), pp 383–389
111. Maerivoet S, De Moor B (2005) Traffic flow theory. ArXiv Physics e-prints