

# A Novel Stable Clustering Approach based on Gaussian Distribution and Relative Velocity in VANETs

Mohammed Saad Talib

Faculty of Information and Communication Technology  
Universiti Teknikal Malaysia Melaka  
College of Administration and Economics  
University of Babylon, Babel – Iraq

Aslinda Hassan, Burairah Hussin, Z.A. Abas

Center for Advanced Computing Technology  
Faculty of Information and Communication Technology  
Universiti Teknikal Malaysia Melaka  
Melaka- Malaysia

Zaniab Saad Talib

Faculty of Science, University of Kerbala  
Kerbala – Iraq

Zainab Sabah Rasoul

Faculty of Manufacturing Engineering  
Universiti Teknikal Malaysia Melaka  
Melaka- Malaysia

**Abstract**—Vehicles in Vehicular Ad-hoc Networks (VANETs) are characterized by their high dynamic mobility (velocity). Changing in VANET topology is happened frequently which caused continuous network communication failures. Clustering is one of the solutions applied to reduce the VANET topology changes. Stable clusters are required and Indispensable to control, improve and analyze VANET. In this paper, we introduce a new analytical VANET's clustering approach. This approach aims to enhance the network stability. The new proposed grouping process in this study depends on the vehicles velocities mean and standard deviation. The principle of the normal (Gaussian) distribution is utilized and emerged with the relative velocity to propose two clustering levels. The staying duration of vehicles in a cluster is also calculated and used as an indication. The first level represents a very high stabile cluster. To form this cluster, only the vehicles having velocities within the range of mean  $\pm$  standard deviation, collected in one cluster (i.e. only 68% of the vehicles allowed to compose this cluster). The cluster head is selected from the vehicles having velocities close to the average cluster velocity. The second level is to create a stable cluster by grouping about 95% of the vehicles. Only the vehicles having velocities within the range of mean  $\pm$  2 standard deviation are collected in one cluster. This type of clustering is less stable than the first one. The analytical analysis shows that the stability and the staying duration of vehicles in the first clustering approach are better than their values in the second clustering approach.

**Keywords**—VANET; clustering; cluster stability; Gaussian; relative velocity; staying duration

## I. INTRODUCTION

Vehicular ad hoc networks (VANETs) represent a crucial view of the Intelligent Transportation System (ITS). Increasing the daily traffic represents a great challenge for the citizen of all urban and rural places. The street traffic safety and management principles continuously affect the individual's life. Focusing must be made on street safety to

improve its traffic efficiency. Efficient street traffic safety and management can be achieved with the huge advancement in Information Technology (IT) by developing reliable communication among vehicles [1]. A vast development in manufacturing, wireless communication and intelligent technologies helps in equipping modern vehicles with onboard unit (OBU) wireless communication to enable vehicles communicate with their neighbors through a Vehicle-to-Vehicle (V2V) and with Road Side Units (RSUs) as a Vehicle-to-Infrastructure (V2I) or Infrastructure-to-Vehicle (I2V) communication style [2]. VANET is characterized by its high moving speed on streets. The vehicles high mobility is significantly affecting the connectivity, throughput and the process of exchanging safety or non-safety traffic information among vehicles [3], [4] VANET aims to prevent collision, accidents and to offer comfortable for passengers by disseminating emergency information in advance [5].

Dynamic and dense network topology characteristics in VANETs cause problems, such as congestion, rerouting and the hidden terminal problem. Clustering techniques have been proposed to improve the communication in VANETs by grouping nodes in geographical vicinity together to enhance communication efficiency. It can also be used for frequency reuse [6], to reduce data congestion and to support QoS requirements. By grouping the vehicles into sets of similar mobility, the relative mobility between communicating neighbor nodes can be reduced [7]. Clustering structures for VANETs are well studied and discussed in the related literatures. Various clustering approaches have been proposed in VANETs [8], [4], [9]. Grouping of vehicles in clusters may deliver certain level of routing scalability, improving resource sharing and developing communication efficiency [10].

Clustering performance is commonly evaluated by cluster stability. Clustering forms must ensure good stability with less overhead, longer Cluster Head (CH) and Cluster Member

(CM) lifetime duration, fewer numbers of changes in vehicles states and minimum CH changes [11], [12]. Most of the VANETs proposed clustering approaches are based on velocity, position, lanes, vehicles density, movement pattern and degree of connectivity [10], [9]. Ensuring stability represents the essential challenge for clustering procedures in dynamic environments. The well-organized clustering procedures must have the ability to maintain the existing cluster formation and reducing its overhead [13], [11], [14].

In this work, we present a new approach to form vehicles clusters. This approach aims to increase the network stability and to decrease its topology dynamics. The principles of the normal (Gaussian) distribution are utilized for the first time in accepting or rejecting the vehicles in the process of cluster formation. Relative velocity and staying duration are also emerged in this study as crucial decision variables. The velocities method and standard deviation are also represents the corner stone in this suggested clustering approach.

## II. CLUSTERING PROCESS

Clustering is the process of grouping the vehicles into different sets. These sets are called clusters. Every cluster must have one vehicle that plays the role of leader or CH. CH is responsible for the communication with the CMs and with other neighbor clusters. The clustering process is used to reduce the routing overhead, suitable use of the network bandwidth and enhance the message delivery. In order to make the clustering processes more efficient, different clustering algorithms are suggested and designed. These algorithms are developed for the cluster formation and cluster maintenances. In cluster formation step, the algorithm is used to make a cluster, elect the CH, assign members, gateway and assist communication. In maintenance step, the algorithm is used to improve the links and to describe the leaving vehicles, the joining vehicles and the merging clusters [15], [11].

## III. NORMAL DISTRIBUTION

The Normal distribution is also known as Gaussian distribution is usually formed as a bell- curve. It represents an organization of a data set in which most values are groups in the central of the sequence and the others are distributed symmetrically toward both ends. Normal distribution are widely used to approximated and model several actual random phenomena in a well manner. There is a very robust association between the size of a sample n and the possibility to which a sampling distribution tends to the Gaussian shape. Several sampling distributions created on big n can be approximated by the Gaussian distribution even though the population distribution itself is definitely not Gaussian. Gaussian is a continuous time probability distribution to represent all x values from  $-\infty$  to  $+\infty$ . So each possible interval of actual quantities has a probability other than zero. The Gaussian (Normal) distributed is usually denoted by  $N(\mu, \sigma^2)$  (where  $\mu$  is the data mean and  $\sigma^2$  is its variance). Equation 1 represents the probability density function ( $f(x)$ ) of the normal distribution. Mode, median and method have equal values in Gaussian distribution. Literatures show the percentage value of the data falls within any value of standard deviations from the mean. Their limits and percentages were empirically

calculated and listed in tables. Fig. 1 shows some of these percentage values under the normal curve [16], [17].

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (1)$$

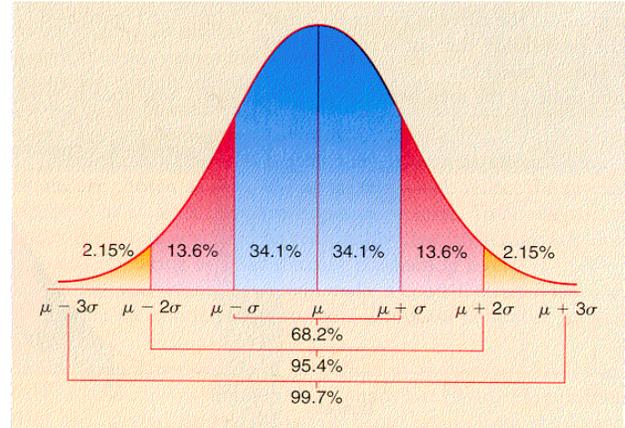


Fig. 1. Percentage values under the normal curve.

The following equations [16] are used to estimate the sample mean ( $\bar{X}$ ) and the sample variance ( $s^2$ ) or the standard deviation (s).

$$\mu = \bar{x} = \sum_{i=1}^n \frac{x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (2)$$

$$\sigma^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n - 1} \quad (3)$$

and;

$$\sigma = \sqrt{\sigma^2}$$

In this study,  $\mu$  is used to represent the sample mean ( $\bar{x}$ ) and  $\sigma$  is used to represent the sample standard deviation (s).

## IV. VEHICLES RELATIVE VELOCITY

The relative velocity between any two moving vehicles can be calculated as the absolute difference between these two vehicles velocities as shown in (4) [18].

$$Relative - velocity = |v_j - v_i| \quad (4)$$

Where  $v_j$  represent the velocity of the vehicle j and  $v_i$  is the velocity of the vehicle i. The impact of the relative vehicles velocities will result in two events affecting the cluster structure. These events are some vehicles leaving the cluster and/or some vehicles entering the cluster [19]. Relative speed level among vehicles can helps in keeping the vehicles in a cluster and reducing the probability of moving out of that cluster [20], [13], tried to create new reliable cluster. They depended on the velocity difference among vehicles, vehicles direction and vehicles location in their clustering formation process. Their approach generates two groups of vehicles whose relative velocity is less and greater than each vehicle [21].

All the related literatures aim to create relatively stable cluster in VANET although the CMs and CHs are high velocity vehicles in order to extend the clusters lifetime. When the relative velocity between CMs and their CH is small then the cluster will be more stable. The vehicle having relatively small velocity difference represents the optimal choice to play the role of the CH [22]. In VANETs, the connection duration between any two vehicles depends on their relative velocity. The connectivity duration between any two vehicles is a function of the communication range and their relative velocity. The relative velocity distribution can be used to estimate the connectivity duration. Velocities Gaussian distribution models can be used to model the connectivity duration. The connectivity duration ( $t_c$ ) between any two vehicles estimated using (5) [23].

$$t_c = \frac{2R}{\Delta v} \quad (5)$$

Where  $R$  is the vehicle communication range and  $\Delta v$  is the relative velocity between the two vehicles.

Merging the velocity and location of the adjacent vehicles represents one of the best approaches to measure the link duration. Most of the suggested cluster merging approaches are motivated either when the distance between two adjacent CHs is less than a specific threshold [13]. Or when the CHs stay associated for a time duration larger than a predetermined value [24].

The applied performance metrics to evaluate the cluster stability are the CM and CH duration, CH change rate, clustering overhead, and number of vehicles that needs to make a decision about the next state. The average relative speed is preferred to be low which means it has less moveable compared with their adjacent. It can be elected as a CH. To maximize the connection probability, the relative velocity of the vehicles must be minimized. When the relative vehicle velocities are increasing, the probability of breaking the connection will increase [25]. Therefore, the relative velocity between two vehicles can vary from 0 to 260 km/h as indicated in certain highways zones [26].

### V. CLUSTERING MANAGEMENT

In this paper, the proposed model is heavily based on the clustering process. The instantaneous velocity of each vehicle represents the pivot stone in this model. The moving vehicles in the same direction are considered in this study. Each vehicle must have a unique ID number and equipped with an on-board-unit (OBU) to be able to deal with the IEEE802.11 as a Dedicated Short Range Communications (DSRC) system. The vehicles are also equipped with a Global Positioning System (GPS) device to provide information about real-time vehicle speed, direction, and land attributes. The vehicles can also proactively and periodically collect real-time traffic information.

The cluster formation process must be proposed with the purposes of reducing the number of cluster heads in the network, maximizing the duration of CH and CMs to provide certain stability level and reducing the overhead [27].

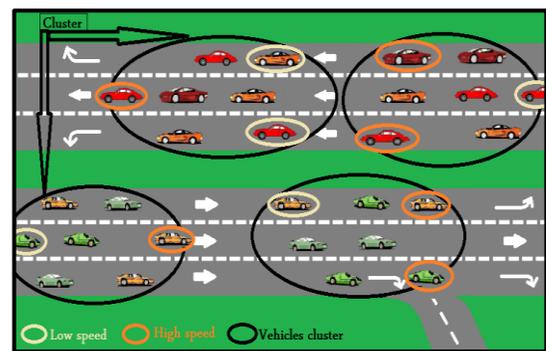
Most literatures stated that the vehicles velocities are normally distributed [28]. So the fitted normal curve can be applied and used to estimate the required parameters. These parameters are the numbers of vehicles, velocities median, mean, mode and standard deviation.

### VI. PROPOSED SOLUTIONS

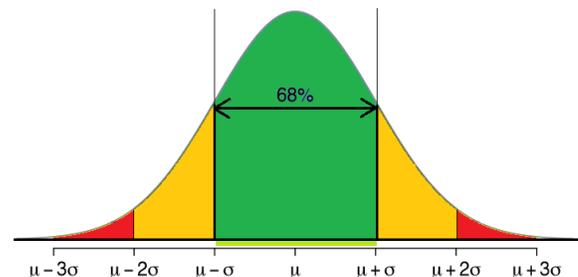
The suggested approach in this analytical study is to design new reliable stable vehicles clusters. Merging the fundamentals of the normal probability distribution function, vehicles mean velocity; velocities standard deviation, relative velocity and vehicles transmission range can be utilized in creating a novel real time clustering approach.

Due to the communication facilities, the number of vehicles in certain street segment can be calculated and controlled. Their velocities and locations are changeable with time. The velocities mean ( $\mu$ ) and standard deviation ( $\sigma$ ) can be calculated easily using equations (2 and 3). The following percentages and numbers can be developed:

- Area under the curve from ( $\mu - \sigma$  to  $\mu + \sigma$ ) represents 68.2% of the total area. We can use this value to select about (68%) of the vehicles to compose a very high stability cluster and exclude about 32% of the vehicles. These excepted vehicles represent the vehicle affecting the clusters maintain. Some of these vehicles will leaving the cluster after certain time due to their very high or very low velocity. Such cluster will have long life time cluster. All the vehicles in this cluster are within a very close level of velocity. The relative velocity among them is very small. Fig. 2 shows in (a) a simple representation of the vehicles grouping and in (b) the suggested area under the normal curve with its percentage.



(a) A simple vehicles grouping representation.



(b) The suggested area under the normal curve and its percentage.

Fig. 2. A simple representation and the suggested area under the curve and its percentages.

- The area under the curve from  $(\mu - 2\sigma$  to  $\mu + 2\sigma)$  represents 95.4% of the total area as indicated in Fig. 1. Fig. 3 represents the limitations and the (95%) of the area under the normal curve. We can use this value to select about (95%) of the vehicles to create a stable cluster and exclude about 5% of the vehicles. These excluded vehicles represent the vehicle affecting the cluster maintains. Some of these vehicles will leaving the cluster after certain time due to their very high or very low velocity. Such cluster has less life time compared with the previous one. Most of the vehicles in this cluster are within a close level of velocities.

There is a tangible difference among the vehicles velocities in this cluster. The relative velocity in this cluster is greater than that in the first clustering approach.

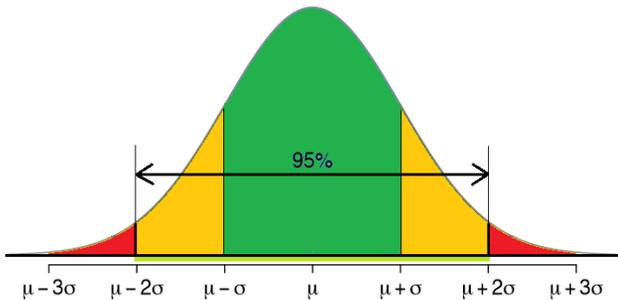


Fig. 3. The limitations and the area that represents 95% under the normal curve.

### VII. CASE EXAMPLE

In order to implement and apply this idea, the following example is suggested. The data sample in Table I represents a suggested example of 21 vehicles (ID) and their velocities in (km/hr.) in certain street.

TABLE I. SAMPLE OF 21 VEHICLES ID AND THEIR VELOCITIES

Vehicle ID	Velocity Km/hr.	Vehicle ID	Velocity Km/hr.	Vehicle ID	Velocity Km/hr.
1	70	8	95	15	70
2	90	9	100	16	60
3	60	10	90	17	120
4	75	11	80	18	100
5	100	12	75	19	95
6	120	13	110	20	85
7	110	14	90	21	90

According to equations (2 and 3):  
 The sample Mean  $\mu = 89.7619$   
 Standard deviation  $\sigma = 17.4983$

The first approach is to collect the vehicles within the percentage range  $(\mu - \sigma$  to  $\mu + \sigma)$  in one cluster. About 68% of the vehicles (13 vehicles out of 21) will be grouped to form this cluster. The following indications can be reached after analyzing the collected data with the Gaussian fundamentals:

Maximum velocity allowed = 107.2602 km/hr  
 Real data Maximum velocity = 100 km/hr  
 Minimum velocity allowed = 72.2636 km/hr  
 Real data Minimum velocity = 75 km/hr

Velocities Range (faster - slower) = 34.9966 km/hr  
 Real velocities Range = 25 km/hr  
 Number of vehicles in this cluster = 13 vehicles  
 Cluster head velocity = 90 km/hr  
 Maximum relative velocity = 100-75 = 25 km/hr

The relative velocity between the CH and the slowest vehicle in this cluster is 15 km/hr.

The relative velocity is ranged from 0 to 25 km/hr.

The expected duration for this cluster can be estimated using equation (5) after suggesting the transmission range to be 1000 m (1km).

$t_c = 2/15 = 0.1333$ hr. (This means that the slowest vehicle will stay about 8 min in this cluster).

$t_c = 2/10 = 0.2$  hr = 12 min ((this means that at least 11 vehicles (85%) will stay after 12 min in this cluster).

Table II shows the IDs and velocities for the vehicles forming this high stable cluster.

TABLE II. VEHICLES ID AND VELOCITY

ID	2	4	5	8	9	10	11	12	14	18	19	20	21
Velocity	90	75	100	95	100	90	80	75	90	100	95	85	90

The second approach is to collect the vehicles within the percentage range of  $(\mu - 2\sigma$  to  $\mu + 2\sigma)$  in a cluster. About 95% of the vehicles (20 vehicles) will be grouped to form this cluster. Practically, according to the velocities data range in Table I, 20 vehicles will be collected in one cluster.

Maximum velocity allowed in this cluster is = 142.2568 km/hr  
 Real Maximum velocity = 120 km/hr  
 Minimum velocity allowed is = 54.7653 km/hr  
 Real Minimum velocity = 60 km/hr  
 Velocities Range (faster - slower) = 60 km/hr  
 Real velocities Range = 60 km/hr  
 Number of vehicles in this cluster = 20 vehicles  
 Cluster head velocity = 90 km/hr

Maximum relative velocity = 90-60 = 30 km/hr (between the CH and the slowest vehicle in this cluster). The relative velocity is ranged from 0 to 30 km/hr.

The expected duration for this cluster can be estimated using (5) after suggesting the transmission range to be 1000 m (1km).

$t_c = 2/30 = 0.066$  hr. (this means that the slowest vehicle will stay about 4 min in this cluster).

$t_c = 2/10 = 0.2$  hr = 12 min ((this means that at least 11 vehicles will stay after 3.6 min in this cluster).

The same data in Table I represents the vehicles forming this stable cluster.

### VIII. CONCLUSION

The dynamic features of VANET require new developed approaches to form stable clusters. In this paper, we presented a new VANET cluster formation analytical approach.

According to vehicles velocities, the average (mean) and standard deviation are estimated and utilized to establish the cluster velocity limits. These velocity limits are imported from the normal (Gaussian) distribution fundamentals using the mean and standard deviation. The upper limit is estimated by adding the (k) standard deviation to the mean and the lower velocity limit by subtracting (k) standard deviation from the mean (where  $k = 1, 2, \text{ and } 3$ ). So, one stability level can be selected to form a cluster out of these two possible clustering levels. These limits are used to create two stable clustering levels. Any vehicle in these clusters having velocity equal or close to the average value can be elected to play the role of the CH. Each vehicle having velocity within the velocity limits registered as CM. The staying duration and relative velocity are also used in this study to evaluate and analyze the cluster maintain step. The cluster stability will be increased with the decreasing in the suggested velocity limit. The resulted relative velocity is decreased with the decreasing in velocity limit while the staying duration is increased. Developing equations to estimate the vehicles leaving and joining probabilities are proposed for the future work.

#### ACKNOWLEDGMENT

This research is fully sponsored by Universiti Teknikal Malaysia Melaka (UTeM) Zamalah Scheme.

#### REFERENCES

- [1] I. Hussain and C. Bingcai, "Cluster Formation and Cluster Head Selection approach for Vehicle Ad-Hoc Network (VANETs) using K-Means and Floyd-Warshall Technique," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 12, pp. 11–15, 2017.
- [2] I. T. Abdel-Halim and H. M. A. Fahmy, "Prediction-based protocols for vehicular Ad Hoc Networks: Survey and taxonomy," *Computer Networks*, vol. 130, Elsevier B.V., pp. 34–50, 2018.
- [3] K. Liu, M. Li, Y. Liu, M. Li, Z. Guo, and F. Hong, "Passive Diagnosis for Wireless Sensor Networks," 6th ACM Conf. Embed. Networked Sens. Syst., vol. 18, no. 4, pp. 1132–1144, 2010.
- [4] R. S. Bali, N. Kumar, and J. J. P. C. Rodrigues, "Clustering in vehicular ad hoc networks: Taxonomy, challenges and solutions," *Veh. Commun.*, vol. 1, no. 3, pp. 134–152, Jul. 2014.
- [5] C. Wu, S. Wang, D. Ben, X. Wang, and M. Guizani, "Dynamic cluster based price control and gateway management for VANETs," in *IEEE International Conference on Communications*, 2012, pp. 5026–5030.
- [6] M. Azizian, S. Cherkaoui, and A. S. Hafid, "A distributed D-hop cluster formation for VANET," *IEEE Wirel. Commun. Netw. Conf. WCNC*, vol. 2016–Sept, no. Wcnc, pp. 1–6, 2016.
- [7] C. Shea, B. Hassanabadi, and S. Valaee, "Mobility-based clustering in VANETs using affinity propagation," in *GLOBECOM - IEEE Global Telecommunications Conference*, 2009, pp. 1–6.
- [8] A. Ahizoune and A. Hafid, "A new stability based clustering algorithm (SBCA) for VANETs," in *Proceedings - Conference on Local Computer Networks, LCN*, 2012, pp. 843–847.
- [9] C. Cooper, D. Franklin, M. Ros, F. Safaei, and M. Abolhasan, "A Comparative Survey of VANET Clustering Techniques," *IEEE Commun. Surv. Tutorials*, vol. 19, no. 1, pp. 657–681, 2017.
- [10] M. Azizian, S. Cherkaoui, and A. S. Hafid, "DCEV: A distributed cluster formation for VANET based on end-to-end relative mobility," in *2016 International Wireless Communications and Mobile Computing Conference, IWCMC 2016*, 2016, pp. 287–291.
- [11] M. Ren, L. Khoukhi, H. Labiod, J. Zhang, and V. Vèque, "A mobility-based scheme for dynamic clustering in vehicular ad-hoc networks (VANETs)," *Veh. Commun.*, vol. 9, pp. 233–241, 2017.
- [12] E. Daknou, M. Thaalbi, and N. Tabbane, "Clustering enhancement for VANETs in highway scenarios," in *2015 5th International Conference on Communications and Networking, COMNET 2015 - Proceedings*, 2016, pp. 1–5.
- [13] Z. Y. Rawashdeh and S. Mahmud, "A novel algorithm to form stable clusters in vehicular ad hoc networks on highways," *EURASIP J. Wirel. Commun. Netw.*, vol. 2012, no. 1, p. 15, 2012.
- [14] S. Y. Kannekanti and G. S. P. Nunna, "An Efficient Clustering Scheme in Vehicular Ad-Hoc Networks," in *Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*, 2017 IEEE 8th Annual, 2017, pp. 282–287.
- [15] B. Sivakumar and K. Srilatha, "A Review on Clustering in VANET," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 4, no. 5, pp. 365–373, 2016.
- [16] R. E. Walpole, R. H. Myers, S. L. Myers, and K. Ye, *Probability and Statistics for Engineers and Scientists*, vol. 3rd, 2012.
- [17] P. Sahoo, *Probability and Mathematical Statistics*. 2013.
- [18] S.-C. Lo, Y.-J. Lin, and J.-S. Gao, "A Multi-Head Clustering Algorithm in Vehicular Ad Hoc Networks," *Int. J. Comput. Theory Eng.*, vol. 5, no. 2, pp. 242–247, 2013.
- [19] K. Abboud and W. Zhuang, "Stochastic modeling of single-hop cluster stability in vehicular ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 1, pp. 226–240, 2016.
- [20] Z. Y. Rawshdeh and S. M. Mahmud, "Toward strongly connected clustering structure in vehicular ad hoc networks," in *IEEE Vehicular Technology Conference*, 2009, pp. 0–4.
- [21] I. Kim and J. Song, "Classification of Clustering Schemes in VANETs," in *Proc. of the Intl. Conf. on Advances in Computing and Information Technology-- ACIT*, 2014, pp. 102–105.
- [22] R. Chai, B. Yang, L. Li, X. Sun, and Q. Chen, "Clustering-based data transmission algorithms for VANET," in *2013 International Conference on Wireless Communications and Signal Processing, WCSP 2013*, 2013, no. 61102063.
- [23] S. M. Abuelenin and A. Y. Abul-Magd, "Empirical study of traffic velocity distribution and its effect on VANETs connectivity," in *2014 International Conference on Connected Vehicles and Expo, ICCVE 2014 - Proceedings*, 2015, pp. 391–395.
- [24] B. Hassanabadi, C. Shea, L. Zhang, and S. Valaee, "Clustering in Vehicular Ad Hoc Networks using Affinity Propagation," *Ad Hoc Networks*, vol. 13, Part B, pp. 535–548, Feb. 2014.
- [25] S. Ucar, S. C. Ergen, and O. Ozkasap, "Multihop-Cluster-Based IEEE 802.11p and LTE Hybrid Architecture for VANET Safety Message Dissemination," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 2621–2636, 2016.
- [26] L. Rivovard, M. Berbineau, M. Wahl, D. Gruyer, and P. Sondi, "Using Real-World Car Traffic Dataset in Vehicular Ad Hoc Network Performance Evaluation," *IJACSA Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 12, pp. 390–398, 2016.
- [27] R. P. Surender Pal, "VANET BASED SIMULATION USING CLUSTER," *Int. J. Latest Res. Sci. Technol.*, vol. 6, no. 3, pp. 25–28, 2017.
- [28] Z. Khan and P. Fan, "A Novel Triple Cluster Based Routing Protocol (TCRP) for VANETs," *IEEE Veh. Technol. Conf.*, vol. 2016–July, 2016.