A Two-level Cooperative Clustering Scheme for Vehicular Communications

José Ángel León Calvo and Rudolf Mathar
Institute for Theoretical Information Technology
RWTH Aachen University, 52074 Aachen, Germany
Email: {leon, mathar}@ti.rwth-aachen.de

***Abstract*—A two-level cooperative network scheme for vehicular communications is proposed in this paper. The network architecture consists on a set of deployed RoadSide Units (RSU) based on the expected density of nodes in the network whereas these nodes are organized as a vehicular ad-hoc network. The position of the RSUs is obtained using the k-means algorithm along with the gap statistic to obtain the optimal number of base stations. The nodes are clustered using spectral clustering based on the geographical position and dynamics of each node, subject to their predictable and highly correlated behavior with the environment. The head-cluster is chosen using concepts of coalitional games in order to extend the stability of the cluster. Additionally, using the beacons sent by the head-cluster to the RSU, a prediction in the dynamic behavior of the clustered nodes is achieved. The RSUs are interconnected using LTE links to provide a cooperative scheme, granting an optimal selection of the head-cluster, and prolonging its lifetime. Finally, the proposed two-level network scheme along with the clustering prediction method are analyzed and compared with the commonly used clustering techniques in a real scenario. The simulation results show the positive impact of the cooperative scheme developed predicting the movement of the clusters.**

Keywords*—VANETs, clustering, cooperative networks, vehicular communications

I. Introduction

Vehicular Ad hoc Networks (VANETs) are increasingly gaining importance currently as a way to improve safety and reduce latency in transportations systems, specially in urban environments where traffic jams and accidents are more likely to occur. Recently, the 802.11p/WAVE protocol has been approved by the IEEE [1], providing a viable technology for vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications. The main characteristic of this protocol is that it is specifically designed for vehicular communications granting low-delay and wide range communications. However, the main obstacle faced while obtaining a low-delay and optimal network is the fragmentation and fast variation of the VANET formation. Moreover, due to the surrounding environment, i.e., buildings and elements from the urban architecture, it is difficult to obtain a fully connected network deriving in disconnected vehicles which are isolated from the rest of the system. Nevertheless, a common way to overcome these problems is to allow the nodes of the network to form clusters. The usage of clustering techniques reduces the number of individual entities in the system and improves the network coordination. Therefore, it is important to smartly deploy the network elements to obtain a full coverage for all the network elements and at the same time allow a fast dissemination of the messages. The issue of clustering has been approached in different ways in the literature, but with the common goal of obtaining the most stable cluster. In [2], a method is developed from a density point of view, using the density in the graph connections of the nodes as a metric. In our two-level network, our aim is not use the density of the nodes to cluster the nodes but to obtain an optimal location of the RSUs in terms of the number of cars connected to them. Using the geographical location of the nodes during a time interval, the optimal location of the RSUs, which cover all the nodes in the network and allocate the resources based on the density, can be obtained.

Another way to approach the clustering problem is to use the mobility of the nodes for this matter. Most of the algorithms attempt to reduce the number of reconfigurations of the head-cluster which leads to increase the stability of the cluster. In [3], the used technique is to select the node which is expected to last longer as the head-cluster based on the lanes of the road. Using the exact knowledge of the road lanes, the nodes broadcast their location and obtain the optimal head-cluster. Additionally in [4], a beacon-based clustering algorithm was developed. This algorithm uses the general scheme implemented by 802.11p/WAVE of broadcasting every message to all the elements in the cluster. Thereafter the cluster is re-organized using the information contained in the beacons.

Our two-level network scheme presents a network architecture which has not been extensively studied so far, since the majority of studies use an LTE/802.11p architecture without additional infrastructure [5]–[8]. These studies use a node as a head-cluster where all the information is gathered for short links and use the LTE technology to provide the long-range communications. The main difference between those studies and our two-level scheme is that we do not store all the information sent by the nodes in a single node, rather in the RSUs, which provides a higher intelligence and cooperative scheme to the network. The RSUs are able to predict the dynamic profile of the clustered elements and provide an optimal head-cluster selection.

Moreover, in [6] the selection of the head-cluster has solely been based on geographical properties, i.e., the proximity to the rest of vehicles, which in some cases cannot be the optimal choice for the head-cluster. A delay efficient network is introduced in [7], where the long-range links are based on LTE and the short-range ones between the vehicles are implemented using 802.11p/WAVE protocol. The manner for clustering the different network nodes in this study is similar to the one introduced in [6], where the vehicles with the strongest path, i.e., the maximum signal strength, were selected as head-cluster to disseminate the intra-cluster messages. A multi-hop

scheme is described in [8] obtaining high connectivity between the network elements but at the same time using a large part of the network bandwidth for the message dissemination.
In our approach, the head-cluster is chosen using the similarity parameters from all the nodes, and selecting the ones that provide a better overall performance using a coalitional game-theory approach. Moreover, the information from the head-cluster is sent only to the RSUs, decreasing the network load. This paper is organized as follows: the network architecture and the two-level clustering techniques are presented in Section II. The simulation setup and used scenario are developed in Section III along with the results and their comparison with other common approaches. Finally, the conclusion is drawn in Section IV.

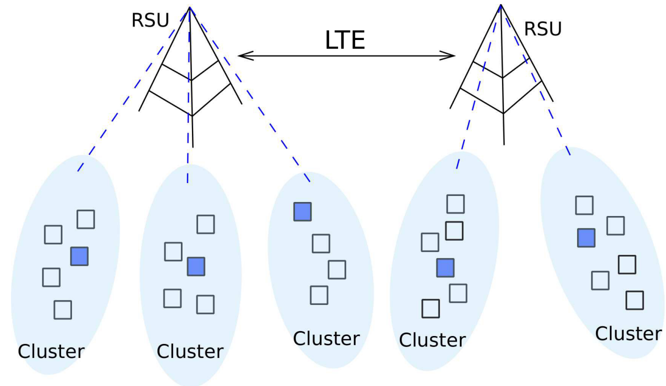

Fig. 1. Network Architecture Overview

II. System Model and Clustering Strategies

A. Two-level Network Architecture

The proposed system model is depicted in Fig. 1. The network architecture is divided into two layers: nodes (vehicles) and RSUs which are inter-connected. The nodes of the network form an ad-hoc structure which varies its form rapidly due to the high mobility of its elements. The nodes can communicate with each other as well as with the RSU using the 802.11p/WAVE technology. Usually, the communication range goes from 0-100 meters due to the high frequency range used in this technology. The beacons sent by each node consists of the triplet

$$m_i(t) := (p_i(t), v_i(t), h_i(t)),$$

where $p_i(t)$ is the geographical position, $v_i(t)$ is the speed and $h_i(t)$ the direction of the node, which is sent directly to the RSUs. Due to the use of 802.11p/WAVE protocol, typically the messages are broadcasted to all the elements of the network, which can cause problems in dense networks [9], [10]. However, in our approach the nodes only need to transmit their messages to the closest RSU, obtaining a less saturated network, i.e., having a cluster formed by $n_i, i \in \{1, ..., l\}$ nodes, the number of broadcasted messages in a typical 802.11p network is $l(l-1)$ per cluster. Nevertheless, in our proposed model, each node transmits solely to the RSU giving a total of $l+1$ messages per cluster, which are k messages from the nodes to the RSU and one more message sent by the RSU to the chosen head-cluster. The cooperation between the interconnected RSUs makes it possible to reduce the number of messages per cluster.
The second layer of the network is created by interconnected RSUs which collect the beacon messages sent by the nodes. These RSUs work both as relays of information and as broadcasters. Moreover, the RSUs have an information center which acts as sink node of the sent messages by the nodes, obtaining the information from the triplets and selecting the head-cluster. The methods and function of the information center will be explained in detail in section II-C along with the head-cluster selection. The RSUs are positioned in highly dense places in order to have a higher visibility. The placement of the RSUs is explained in section II-B.

B. Node Clustering and RSU Placement

In this section the scheme used for placing the RSUs is analyzed. The RSUs are located according to the expected density of nodes using real-traffic data. Applying the k-means algorithm [11], the RSUs are preferentially placed where the highest density of nodes appear while, at the same time, the number of RSUs is proportional to the number of vehicles. For the k-means method only the position $p_i(t)$ of the nodes is used. Each of the nodes is iteratively assigned to its closest cluster centroid which will be the optimal position of the RSU. As shown in Fig. 2 the density of nodes has a shape influenced by the environment.

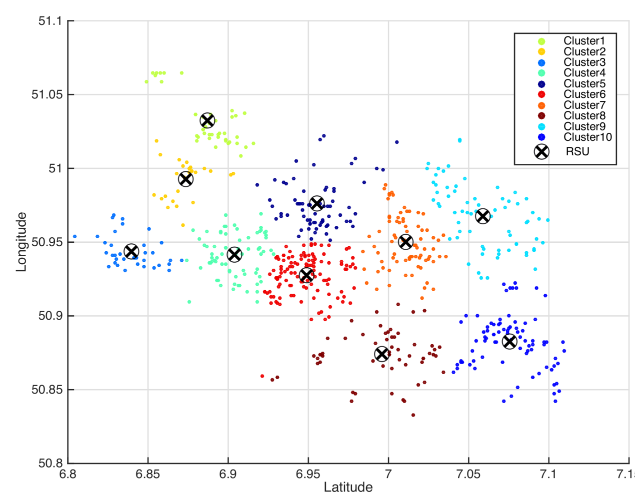

Fig. 2. RSU deployment based on K-means algorithm

However, it is known that the k-means method needs the number of clusters as input. Since the problem we are approaching can be denoted as unsupervised learning, the structure of the data does not provide the optimal number of clusters. It looks reasonable to choose the highest number of possible clusters since it will give us the smallest estimation error. However, it has been observed that after a certain number of clusters L the error tends to be flattened not obtaining any estimation gain by increasing the number of clusters. Therefore, we have used the gap statisticin [12] to obtain the optimal number of clusters, which yields the best trade-off between estimation error and number of clusters. The gap statistic works well with non-separated data and in this case for the density clustering the two features selected from each node are latitude and longitude.

However, since we plan to deploy an optimal vehicular network for every time instant and not only for a single instant t, we have to calculate all centroids, i.e., the RSU position each time instant and later apply an additional k-means method on them to obtain the final positions. Hence, the final deployment is obtained using a double k-means approach which consists on iteratively applying the k-means algorithm for all time instants and later apply an additional k-means method to obtain the final positions.

C. Spectral Clustering for Nodes

In this section, the set of nodes $N = \{n_1, \ldots, n_l\} \in \mathbb{N}^l$ that are assigned to the same RSU are clustered using a sub-spacial spectral clustering [13], [14]. This clustering method uses a metric based on the connection of all the nodes with a positive similarity using a full-connected graph form. For this method, we apply the Gaussian similarity function to obtain the affinity matrix $S \in \mathbb{R}^{l \times l}$:

$$S_{i,j} = exp(-\frac{\|p_i - p_j\|^2}{2\sigma^2}) \ \forall i \neq j \text{ and } S_{i,i} = 0, \quad (1)$$

where the parameter σ controls the similarity threshold between the neighbor nodes, and p_i, p_j are the position of the nodes forming the cluster. The spectral clustering works well for different dataset structure even when it does not have a convex form. Using the top eigenvectors from the affinity matrix, the spectral clustering algorithm obtains great results in a reduce amount of time due to its simple structure. However, since we attempt to cluster the nodes in terms of a higher dimensional data, i.e., position, direction and speed, in our approach a subspace transformation needs to be performed. It is noteworthy that since the dataset that are compared for the similarity metric are related to different subspaces, the sense of distance becomes meaningless. Therefore, the dataset will be transformed using the analogy of an intensity matrix as follows:

$$v_i(t) \in \mathbb{R} \longrightarrow \tilde{v}_i(t) \in [0, 1] \quad (2)$$
$$p_i(t) \in \mathbb{R}^2 \longrightarrow \tilde{p}_i(t) \in [0, 1]^2 \quad (3)$$
$$h_i(t) \in \mathbb{R} \longrightarrow \tilde{h}_i(t) \in [0, 1] \quad (4)$$

After this transformation, all the values in the dataset are comparable and hence, the similarity graph and the metric associated to it is meaningful. For simplicity the different dataset are weighted equally.

D. Head-Cluster Selection and Dynamics Prediction

Once the nodes are clustered as described in II-C, the next step is to choose the head-cluster with the expected longest lifetime. In this section, a scheme to select the head-cluster based on coalitional games is developed. The main goal of a coalitional game is to obtain the maximum payoff from the collaboration of all the nodes in a coalition. A coalition inside a cluster can be defines as $S \subseteq N$, where $N = \{n_1, \ldots, n_l\}$ is the set of all nodes in the cluster. In coalitional games, the maximum output is always given by the payoff function obtained by the collaboration of all the nodes, creating a grand coalition.

In the present approach, the coalition S is formed by all the nodes connected to the same RSU obtained using the spectral clustering method. Since all the nodes forming the cluster have similar properties as mentioned in Section II-C, it is reasonable to assume that the nodes will have a common goal that can be maximized, obtaining an optimal solution for all the nodes.

The proposed coalitional game assumes that the RSUs are myopic, i.e., the head cluster is selected considering the actual status of the cluster. In each round of the algorithm, the RSU extracts the information sent in the beacons $m_i(t)$ by the nodes and selects the node which minimizes the payoff function. The payoff function is based on the similarity concept used to create the cluster and aims to obtain the head-cluster which describes the entire cluster precisely. Therefore, the payoff function for a time instant over all node n_i, n_j in a cluster is as follows:

$$f(t) = \{d(p_i(t), p_j(t)) + d(v_i(t), v_j(t)) + d(h_i(t), h_j(t))\} \quad (5)$$

where $d(\cdot, \cdot)$ is defined as the Euclidean distance among the different points. The goal is to obtain the minimal value for the payoff function in each time instant as

$$\min_{\forall i,j} f(t) \quad t \in \{t_0, \ldots, t_p\} \quad (6)$$

where t denotes the simulation time. Once the head-cluster which minimizes the payoff function is selected, it can be defined as:

$$m_{hc} := (\hat{p}(t), \hat{v}(t), \hat{h}(t)) \quad (7)$$

which also defines the entire cluster. Considering only the head-clusters of the network, it is possible to define the traffic as macroscopic, i.e., instead of individual cars which make the simulation infeasible, choosing the head-clusters to obtain the general profile of the traffic. The goal of using a macroscopic traffic simulation has two benefits. On the one hand, it is too complex to simulate and predict the dynamics of each individual car. On the other hand, the bandwidth use is reduced grouping the nodes and choosing a head-cluster.

In road networks, vehicles can be modeled as a fluid which help to anticipate traffic phenomena and predict the interaction between the network elements [15], [16]. The principal goal of this method is to obtain an accurate prediction of the geographical properties of the vehicles. Generally, the speed of each car $v_i(t)$ is obtained and recorded, however, in our approach we will denominate $\tilde{v}(p, t)$ which is associated to the entire cluster, being $p(n)$ the position in space covered by the head-cluster. In addition, the same concept will be used for the rest of properties, i.e., the direction and position. The principal parameters in traffic flow models are: velocity, density of traffic and flow. These parameters are addressed by the following equation

$$q(p, t) = \rho(p, t) v(p, t) \quad (8)$$

which is related to the conservation of cars and the relationship between the car velocity and traffic density.

Using the cooperative scheme due to the interconnected RSUs, the head-cluster selection can be modified using the dynamics prediction. Applying the knowledge of the head-cluster from different RSUs, it is possible to predict the next position of the cluster, and hence, choose a head-cluster which may not be the optimal for the time instant t, but it will create a more stable cluster in future time instants.

III. Simulation Results and Analysis

A. Scenario Setup

The vehicular mobility dataset introduced in this paper is mainly based on data made available by the TAPASCologne project [17], an initiative of the Institute of Transportation Systems at the German Aerospace Center (ITS-DLR). The available traffic-data records a two-hours traffic movement in Cologne, Germany. Since the data shows the traffic in rush hour, it presents a good dataset to model the maximum capacity of the network, showing the behavior of the two-level network in these cases. The area of the studied scenario is 20 squared kilometers and the simulation is implemented using SUMO [18] which works as a macroscopic traffic simulator.

B. Different Clustering Methods

First, we will analyze the different common methods to cluster ad-hoc network in MANET and VANETs.

1) Lowest-ID [19] consists of selecting as head-cluster the node with the lowest ID in range. The IDs are distributed to each node once it enters the network and they do not change. This clustering scheme works acceptably well in networks with a low dynamic profile.
2) Highest Degree [20] is based on the connectivity among the nodes. The algorithm creates a graph scheme to select the head-cluster using the broadcasted messages among the nodes and selects the node with the highest degree of connectivity.
3) Utility Function [21] uses a similar clustering approach as the method introduced in the present work. It takes the head-cluster which is closer in distance and speed to the rest of nodes in range, and selects it as head-cluster. However, it does not use the information from other cluster, through RSU connectivity, to optimize the selection.

C. Number of beacons

This parameter is of special importance due to the high number of nodes that can be in a cluster simultaneously [10]. The protocol 802.11p/WAVE broadcasts the messages to all the nodes in the same cluster, however, in our approach this is not needed since the infrastructure controls the status of all the nodes.

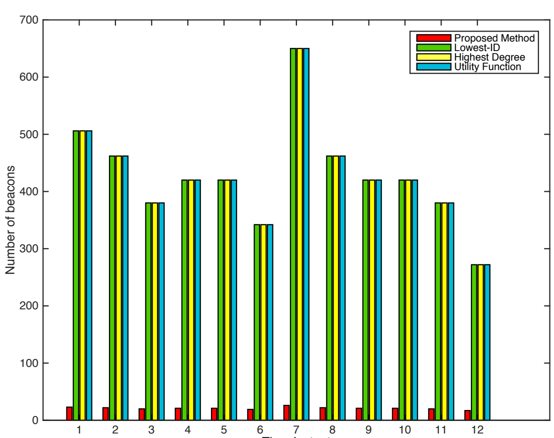

Fig. 3. Number of beacons in a cluster

As shown in Fig. 3, the number of beacons in the three common clustering techniques, Lowest-ID, Highest Degree and Utility Function is the same, being it $l(l-1)$ with l nodes in the cluster. In our proposed model, the number of beacons is much reduced providing a less congested network. Nonetheless, the main drawback of this approach is the requirement of installing an infrastructure which could be logistically difficult and economically expensive.

D. Head-cluster Stability

In this section, the results obtained from the head-cluster stability are analyzed.

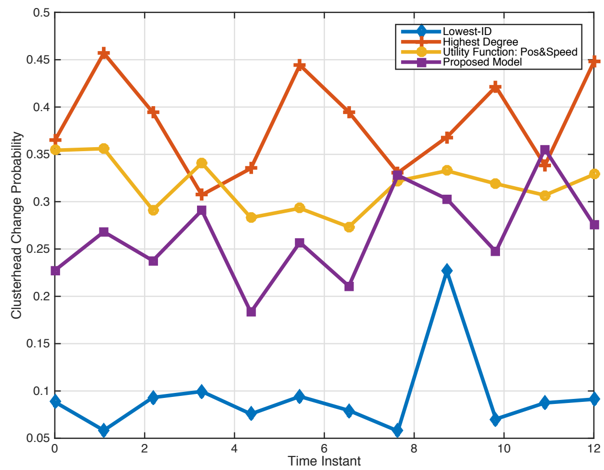

Fig. 4. Cluster Lifetime comparison

As shown in Fig. 4, the best approach regarding this parameter is Lowest-ID. However having an extremely low head-cluster variation is a drawback in VANETs, since it will not provide an optimal selection for the different changes in the network. Highest-Degree performs worse in this aspect, having a really volatile head-cluster. This frequent changes produce a high number of exchanged messages in the network, and as it was mentioned in previous section, the number of beacons to select a head-cluster is remarkably high, which potentially congests the network. The last clustering scheme studied is the Utility Function using the average speed and position of the cluster. It works fairly well obtaining a balance between the longevity of the cluster and the adaptation to the cluster changes.
Analyzing the proposed model in this paper, we can observe a better performance than the Utility Function, due to the expected prediction of the head-cluster based on the dynamics. Moreover, the proposed model adapts itself to the dynamic nature of VANETs while providing an acceptable head-cluster lifetime.

E. Cluster Dynamics Prediction

The proposed model uses the cooperation between the RSUs to predict the dynamics of the head-clusters in the future time instants.
As shown in Fig. 5, the mean error distance between the predicted head-cluster and the value obtained from real traffic data is around 9 meters. The first step in the simulation is the one with the biggest error. Nevertheless, the error after this point is stabilized at around 8-10 meters. In view of the

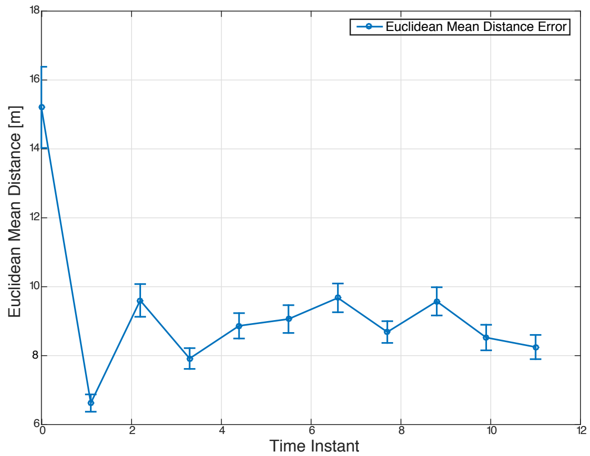

Fig. 5. Head-Cluster Dynamics Prediction

results obtained from predicting the head-clusters, it can be concluded that the idea of modeling the traffic data as a fluid works accurately, adding more knowledge to the network.

IV. Conclusion

In this paper, a two-level clustering scheme for vehicular communications was developed and analyzed. This model creates a hierarchical network where the RSUs are interconnected creating a cooperative scheme. The first level of the network is composed by the nodes (vehicles) creating an ad-hoc network, where the main characteristic is the high mobility of the nodes. The fast variation in the elements of the network motivates the application of clustering techniques.

The second level of the network is created by the RSUs which are deployed in terms of the expected density of nodes. Once the RSUs are deployed, the nodes of the network are grouped in clusters in terms of their similarity parameters using spectral clustering. This clustering technique fits perfectly the aforementioned problem because of the similar behavior of the nodes due to the typical constraints of the scenario in VANETs. The obtained results show an improvement in the average cluster lifetime. This improvement in the cluster stability is obtained using a smaller part of the spectrum due to the reduced number of beacons used in the two-level architecture. Moreover, the proposed architecture is capable of predicting with great accuracy the future positions of the head-clusters. It is noteworthy that the prediction scheme used in this paper is a simple one, foreseeing only one time instant in the future. However, according to our estimation a further improvement can be obtained by extending the cooperative scheme and the prediction time interval.

References

[1] IEEE, "Ieee standard for wireless access in vehicular environments (wave)–multi-channel operation," *IEEE Std 1609.4-2010 (Revision of IEEE Std 1609.4-2006)*, pp. 1–89, Feb 2011.

[2] S. Kuklinski and G. Wolny, "Density based clustering algorithm for vanets," in *Testbeds and Research Infrastructures for the Development of Networks Communities and Workshops, 2009. TridentCom 2009. 5th International Conference on*, April 2009, pp. 1–6.

[3] S. A. Mohammad and C. W. Michele, "Using traffic flow for cluster formation in vehicular ad-hoc networks," in *Local Computer Networks (LCN), 2010 IEEE 35th Conference on*, Oct 2010, pp. 631–636.

[4] E. Souza, I. Nikolaidis, and P. Gburzynski, "A new aggregate local mobility clustering algorithm for vanets," in *Communications (ICC), 2010 IEEE International Conference on*, May 2010, pp. 1–5.

[5] A. Vinel, "3gpp lte versus ieee 802.11p/wave: Which technology is able to support cooperative vehicular safety applications?" *IEEE Wireless Communications Letters*, vol. 1, no. 2, pp. 125–128, April 2012.

[6] R. Sivaraj, A. K. Gopalakrishna, M. G. Chandra, and P. Balamuralidhar, "Qos-enabled group communication in integrated vanet-lte heterogeneous wireless networks," in *2011 IEEE 7th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Oct 2011, pp. 17–24.

[7] R. Atat, E. Yaacoub, M. S. Alouini, and F. Filali, "Delay efficient cooperation in public safety vehicular networks using lte and ieee 802.11p," in *2012 IEEE Consumer Communications and Networking Conference (CCNC)*, Jan 2012, pp. 316–320.

[8] S. Ucar, S. C. Ergen, and O. Ozkasap, "Multihop-cluster-based ieee 802.11p and lte hybrid architecture for vanet safety message dissemination," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 4, pp. 2621–2636, April 2016.

[9] E. M. van Eenennaam, W. Klein Wolterink, G. Karagiannis, and G. J. Heijenk, "Exploring the solution space of beaconing in vanets," in *First IEEE Vehicular Networking Conference, VNC2009, Tokyo, Japan*. IEEE Communications Society, October 2009, pp. 261–268.

[10] A. Vinel, Y. Koucheryavy, S. Andreev, and D. Staehle, "Estimation of a successful beacon reception probability in vehicular ad-hoc networks," in *Proceedings of the 2009 International Conference on Wireless Communications and Mobile Computing: Connecting the World Wirelessly*, ser. IWCMC '09. New York, NY, USA: ACM, 2009, pp. 416–420. [Online]. Available: http://doi.acm.org/10.1145/1582379.1582470

[11] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. Berkeley, Calif.: University of California Press, 1967, pp. 281–297. [Online]. Available: http://projecteuclid.org/euclid.bsmsp/1200512992

[12] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 2, pp. 411–423, 2001. [Online]. Available: http://dx.doi.org/10.1111/1467-9868.00293

[13] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS*. MIT Press, 2001, pp. 849–856.

[14] G. L. Scott and H. C. Longuet-Higgins, "Feature grouping by'relocalisation'of eigenvectors of the proximity matrix." in *BMVC*. Citeseer, 1990, pp. 1–6.

[15] M. J. Lighthill and G. B. Whitham, "On kinematic waves. ii. a theory of traffic flow on long crowded roads," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 229, no. 1178, pp. 317–345, 1955. [Online]. Available: http://rspa.royalsocietypublishing.org/content/229/1178/317

[16] R. Haberman and R. W. Kolkka, "Mathematical models," 1977.

[17] TAPASCologne. Tapascologne project. [Online]. Available: http://sourceforge.net/apps/mediawiki/sumo/index.php?title=TAPASCologne.

[18] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent development and applications of SUMO - Simulation of Urban MObility," *International Journal On Advances in Systems and Measurements*, vol. 5, no. 3&4, pp. 128–138, December 2012.

[19] J. L. M.Jiang and Y. Tay. (1999) Cluster based routing protocol. [Online]. Available: https://tools.ietf.org/html/draft-ietf-manet-cbrp-spec-01

[20] S. Sivavakeesar and G. Pavlou, "Associativity-based stable cluster formation in mobile ad hoc networks," in *Second IEEE Consumer Communications and Networking Conference, 2005. CCNC. 2005*, Jan 2005, pp. 196–201.

[21] P. Fan, J. G. Haran, J. Dillenburg, and P. C. Nelson, *Cluster-Based Framework in Vehicular Ad-Hoc Networks*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 32–42. [Online]. Available: http://dx.doi.org/10.1007/11561354_5