City Research Online

City, University of London Institutional Repository

Location Aware Data Aggregation for Efficient Message Dissemination in Vehicular Ad Hoc Networks

Milos Milojevic, *Student Member, IEEE*, Veselin Rakocevic, *Member, IEEE*

Abstract—Intelligent location-aware data aggregation mechanism for real-time observation, estimation and efficient dissemination of any kind of traffic information in vehicular ad-hoc networks (VANETs) is presented in this paper. The mechanism introduces location awareness algorithm, enabling spatiotemporal database indexing and providing location context of the messages without the use of advanced positioning systems like satellite navigation and digital maps. Intelligent passive clustering and adaptive broadcasting are used to minimize the number of messages exchanged, packet collisions and network load. The incoming messages are fused by Kalman filter allowing the description of the traffic related information as a system characterized by as many variables as needed, depending on the application design. The scheme allows the comparison of aggregates and single observations which enables their merging and better overall accuracy. Old information in aggregates is removed by real-time database refreshing thus leaving only newer relevant information for driver to make real-time decisions in traffic. The mechanism is generic and can be used for any kind of VANET information. It is evaluated by extensive simulations to show the efficiency and accuracy.

Index Terms—cooperation, dissemination, intelligent data aggregation, VANETs.

I. INTRODUCTION

Vehicular ad-hoc networks (VANETs) can be used to provide drivers with real-time data about some traffic states, which drivers can use to adjust their routes. Communicating potentially large quantities of data in VANETs can be quite challenging, especially in the case of high node density, when the broadcast storm problem can occur [1]. In such situations data aggregation can increase communication efficiency, by optimizing data gathering, processing or dissemination. Still, aggregation should not compromise the accuracy of the disseminated information.

Numerous data aggregation techniques have been proposed for application in wireless sensor networks (WSNs) [2][3]. These are not suitable for VANETs due to differences between the two types of networks. The WSNs are power constrained thus most protocols are designed with this limitation in mind. The nodes in WSNs are usually hierarchically organized with limited mobility and with the sink node responsible for data collection, processing and dissemination. VANETs are not power constrained and every node is practically a sink, their mobility is significantly higher than in WSNs and they move on constrained road network. The aggregation schemes for WSNs are usually based on certain structures like cluster, a tree or chain, which might be difficult or even impossible to build and maintain in VANETs. This is due to the high mobility of the nodes in VANETs, which make the formation of network structures challenging. Finally, the applications in VANETs require real-time information, whereas in WSNs they do not. The body of the research work on VANET data aggregation is not as large as for WSNs and only a limited number of proposals for VANETs can be found [4]-[13].

In this paper we present an intelligent, location-aware data aggregation mechanism for real-time observation, estimation and efficient dissemination of traffic information in VANETs. The main novelty of the mechanism is its significant reduction of the communication overhead of a fully distributed VANET while providing accurate location awareness. Unlike the existing data aggregation mechanisms, the proposed mechanism can be deployed in any type of vehicle with VANET communication capability, even without systems like navigation, digital maps or additional information from the roadside units or the local traffic authorities. It does not require any knowledge about street segmentation and identification of segments, thus makes the database maintenance significantly less computationally complex. Moreover, the scheme provides flexible segmentation of streets in order to provide data aggregation structure while the vehicle is moving. The motivation of our work comes from the need for efficient and scalable distributed protocols for the distribution of neighborhood information using VANETs. The main aim of our solution is to use data aggregation algorithm to increase scalability without compromising the accuracy of the communicated network information. Location awareness is achieved by using a simple direction parameter to create spatiotemporal database indexing for storing and sending messages in the network. Additionally, this enables the comparison of aggregates and single observations which contributes to better accuracy. The database is being constantly refreshed, which solves the problem of old information in aggregates, thus providing only fresh information. The mechanism is generic and messages can contain anything from traffic congestion information to accident location or free parking space information. To

achieve reliable estimation from a large number of observations and overcome the problem of noisy observations, the messages are fused using Kalman filter. Additionally, the mechanism gives the possibility of describing the traffic related information as a system characterized by as many variables as needed, depending on the application design. Communication efficiency is achieved by intelligent passive clustering and adaptive broadcasting based approach where individual vehicles intelligently decide if and when the message should be broadcasted to other vehicles. In this paper, the mechanism's performance is evaluated by extensive simulations to show the communication efficiency and accuracy.

II. RELATED WORK

The location information enables the vehicles to assign an observation to a certain location when sending or receiving the messages. Data aggregation should provide the spatiotemporal understanding and distribution of some traffic phenomenon to the driver. Existing data aggregation mechanisms for VANETs use fixed road segments for aggregation [6][7][9][10][11] or fixed areas such as city blocks containing more streets [4]. Using fixed road segments can be inconvenient because each segment has unique identification and vehicles have to maintain the database about all of them, and their number can be extremely large in a city. The problem with using fixed areas like city blocks as an aggregation structure is that areas contain several streets within. Thus, such aggregates are not particularly precise because they refer to large area and are then less relevant to the driver. One of the biggest problems with the existing data aggregation mechanisms that use both segments and areas as an aggregation structure is their dependence on location information obtained from external positioning systems like satellite navigation, GPS or digital maps. These positioning systems are still not widely available in most vehicles [14], and thus would limit the use of the data aggregation mechanisms. Additionally, GPS is often unreliable in urban environments [15] and requires complementing localization techniques. Therefore, there is a clear need for data aggregation mechanisms which can enable effective communication while providing spatiotemporal awareness of the environment without the use of positioning systems like GPS and digital maps.

The mechanisms presented in [4]-[8] use simple periodic broadcasting to disseminate data, which is proven to be causing broadcast storm and scalability issues in certain cases [1][16]. The broadcast storm problem is so severe that the performance of the 802.11p MAC layer was examined several times and found having limited performance especially in dense networks [17]. To reduce the broadcast storm and improve scalability in VANETs the common approach is the use of adaptive broadcasting [1][18][19][20] or clustering [21]. The former approach adapts the broadcast frequency according to certain criteria. However, to the best of our knowledge only authors of [9][10] used adaptive broadcasting within data aggregation mechanism, although for different

motivations. In [12] data aggregation is achieved by restricting forwarders based on the position of vehicle's neighbors. In terms of messages and aggregates, one of the challenges is the existence of the old information in aggregates. Another issue with aggregates is that they sometimes cannot be compared and merged to generic observations.

Related research work in the area of data aggregation in VANETs does, however, present a number of individual solutions which address some of the technical challenges described above. For example, in [4] the authors presented hierarchical probabilistic data aggregation scheme for VANETs. Here, the aggregates are duplicate-insensitive and are based on square areas of different sizes: small, medium and large. A structure-free data aggregation scheme based on fuzzy reasoning, enabling the vehicles to reach the aggregation decision based on the application specific set of criteria is presented in [5]. Authors of [6] present a cluster-based data aggregation approach where they use compression to provide aggregation without losing accuracy. In [10][11] authors insert the delay before forwarding the message to enable the aggregates, which are based on street segments, to meet at a certain point. In the [9] dissemination scheme uses aggregation per street segment and adapts the broadcast interval based on the type of the event observed.

Finally, one of the biggest challenges as recognized by [22] is lack of generic proposals which could be used for more than one application types. For example, in SOTIS [7], each vehicle analyses traffic conditions based on the messages received from other vehicles which are aggregated per road section. In Traffic View [8], a congestion detection system is presented, including two proposed node-centric data aggregation mechanisms. Additionally, apart from being application specific, most schemes are scenario specific as well, thus considering only one type of scenario, for example highway traffic [5][6][7][10][11]. There is a clear need for a more universal data aggregation solution which can be deployed without restrictions on vehicle onboard systems, on any type of application and in any kind of scenario.

III. LOCATION-AWARE DATA AGGREGATION

Our data aggregation mechanism is conceptually based on a modified generic VANET data aggregation *Architecture* model as proposed in [21]. There, data aggregation in VANETs is defined as a process containing four modules, each performing a specific function. These modules are: *Decision*, the *World Model*, *Dissemination* and *Fusion*. The *Decision* module decides if and how the data obtained and received is being aggregated. The *World Model* presents vehicle's knowledge about the network and environment while the *Dissemination* module specifies if and how the dissemination is being done. The *Fusion* module specifies the way the information is being fused with other information. We modified this model by introducing the new Location Awareness algorithm (*LA*) module which provides location information to the whole data aggregation mechanism. Position of the new *LA* module in the architecture can be seen in Fig. 1.

A. Location Awareness (LA)

We assume that each vehicle has access to basic information such as speed, traversed distance and direction, and does not have access to any external positioning system. As in all VANET applications we also assume that vehicles have onboard unit (OBU) with communication interface, memory and processing units. Additionally, we assume that vehicles are able to obtain generic observation M from on-board sensors, depending on the application type. Theoretically, M can be the number of free parking spaces, the congestion estimate, the level of noise, or any other measurement provided that vehicle has the sensor to obtain it.

Let θ be the angle representing the current movement direction of the vehicle, having values in the following range:

$$-\pi \leq \theta \leq \pi \tag{1}$$

We also assume that vehicles can obtain the value of θ at any time, from a simple device like compass, and that drivers are able to recognize the direction of the roads in front of them. Vehicles can use the values of θ to divide their route into street sections, using the change in θ as an indicator of the street section change. According to our LA algorithm, every time angle θ changes, the new street section on the route will be detected by the vehicle, and thus the route S of the vehicle is defined as a sequence of consecutive street sections s_i and each of them characterized with direction θ_i:

$$S = \sum_i s_i\left(\theta_i\right) \tag{2}$$

Theoretically, angle θ should be constant per street section, but this would give infinite number of unique angles θ and thus infinite number of street sections. To prevent this, we define a finite offset value $\Delta\theta$ and define the street section to be unique as long as the following equation for two successive angle values θ_1 and θ_2 is fulfilled:

$$\left|\theta_2 - \theta_1\right| \leq \Delta\theta \tag{3}$$

When the angle difference from the equation (3) becomes greater than $\Delta\theta$ the vehicle detects that street section has changed. Each time the street section is changed the counter of street sections on the route i is incremented. This way the algorithm enables the vehicles to approximate their routes with a finite number of street sections, as shown in Fig. 2. Based on this methodology, the vehicle counts how many street sections it traversed as it moves, and what were the values of θ, i and measurement M during that time. The pseudocode of the LA module is shown in Table I. These parameters enable the vehicles to map certain values of observation M to street sections they traversed. The same parameters are later included in messages which are sent to other vehicles.

Additionally we introduce the knowledge depth parameter K, representing the size of the aggregates that are sent to other vehicles. K also defines one dimension of the database size of the vehicle. According to K, the vehicle sends the aggregate containing observations from last K street sections that it traversed, in the following format:
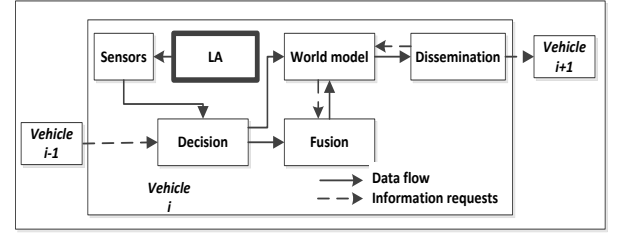

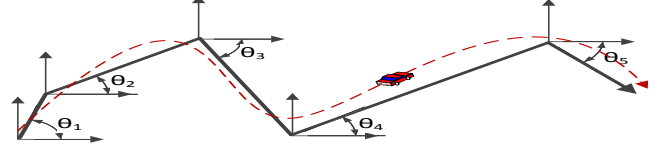Fig. 1. Modified generic data aggregation architecture model

Fig. 2. Route approximation process by consecutive street sections.

TABLE I. PSEUDOCODE OF LA MODULE

LA module pseudocode			
1	initialize (timer=0 , location=0, buffer =0, route=0)	5	route= route + location
2	location=get(current_direction)	6	location_buffer = location
3	if location ≠ location_buffer then	7	end
4	i=i+1	8	else (timer = timer+1);

$$\left|M\left(k\right),\ \theta\left(k\right),\ i\left(k\right)\right|, where\ k = 1,...,K \tag{4}$$

K determines the historical "depth" of the observed measurement M the receiving vehicle will obtain. Therefore when vehicle receives the message from another vehicle, it knows the values of observations of that vehicle on the previous K street sections. For example, in the case of $K=3$, vehicles in the network would send the values of observation M for the last three street sections on their route.

Each measurement taken by the vehicle is characterized by the number of street segment i that it was taken on, and by the value of the angle θ_r of that street segment. Upon obtaining the current and previous directions of the sending vehicle, the receiving vehicle stores the measurements in the database with reference to θ_r and i. Clearly, storing information based on real value of θ_r would be inefficient because theoretically it can have infinite number of values. Because of that, the mapping function m is introduced which stores received information based on θ_r and i, but in one of the finite number of predefined memory slots:

$$m:\theta_r \rightarrow \alpha_j\ \text{where}\ \ \alpha_j = \frac{360^o}{g} \tag{5}$$

$$g \in Z\ \text{and}\ \sum_j \alpha_j = 360^o \tag{6}$$

The memory slots α_j represent bands of angles θ_r is mapped into, and their number is determined by the granularity parameter g. This parameter shows the precision of the location awareness, and the larger the value of g, the location information is more precise, but at the cost of efficiency both

in terms of the memory space and the processing power and time. To show this in a practical example, we assume that $g=20$. This means that the size of the memory slots is 18°, and their distribution is as follows:

$$\alpha = 18^o$$
$$\alpha_i = \{\alpha_1,\ \alpha_2,\ \alpha_3,\ ...,\ \alpha_{18},\ \alpha_{19},\ \alpha_{20}\} = \quad (7)$$
$$= \{0^o - 17^o,\ 18^o - 35^o,\ 36^o - 53^o,\ ...,\ 342^o - 360^o\}$$

If we assume that angle of the incoming message to be $\theta_r =32^o$, the result of the mapping function would be:

$$m(\theta = 32^o) \rightarrow \alpha_2 = 18^o - 35^o \quad (8)$$

As described in the example above, the result of mapping the incoming data based on the angle θ_r would be in the second range of angles between 18° and 36°. By using the mapping function m the vehicle calculates the approximate direction of the street that incoming measurement originates from. In this case the approximate direction angle of the incoming measurements is the street with direction somewhere between 18° and 36°, while parameter i describes the 'historical' distance of the observation. By combining the distance and the direction calculated from incoming messages, the *LA* module creates spatiotemporal database indexing and provides the drivers with measurements in the streets sorted per their direction angle and per distance i. This way it enables the drivers to know about the approximate traffic states in the streets in their nearest surroundings. The measurement whose $i=3$ refers to the closest street section with θ_r in the vehicle's communication range. The measurements with $i=2$ and $i=1$ are one and two street sections further away, respectively.

Therefore, the street sections characterized with angle θ are used as the structure for aggregation of observations in the memory. The algorithm does not need fixed predefined structure or any location information like coordinates or street names in order to aggregate the observations – it performs the aggregation independently while vehicle is on the move. The detailed architecture of database in the vehicles is presented in the *World Model* section.

B. World Model

The *World Model* represents the database of the vehicle and provides spatiotemporal context to the stored information, thus enabling the driver or the application to use this information to adjust their traffic routes in real-time. In this section we further explain the architecture of the database and the concept of the aggregation process, while the process of merging the information is shown later in the *Fusion* section.

Fig. 3 can be used to approximately illustrate the structure of the database, and the process of storing the received observations. The two vehicles are shown, the red sending vehicle and the blue receiving vehicle together with their movement directions measured as shown in the Fig. 3. We assume that red vehicle traversed through street sections 1, 2 and 3, and that it observed measurement values of 7, 4 and 11, on those street sections respectively. Assuming the knowledge depth $K=3$, the red vehicle sends the observations for the last three street sections traversed, sections 1, 2 and 3.
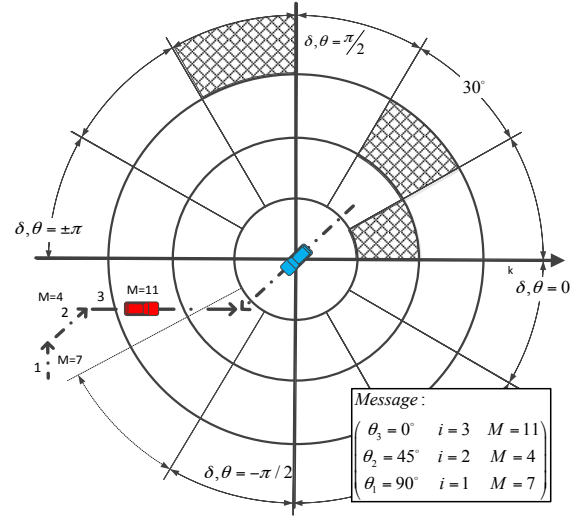

Fig. 3. Example scenario of location awareness.

Therefore the content of the message sent by the red vehicle will be as shown in Fig. 3. Additionally, for this example, the granularity of the aggregation is set to $g=12$, meaning that the vehicle's spatial awareness is divided into 12 equal ranges of angles of 30°, as shown in Fig. 3. According to those angular ranges, the memory slots are created in the database as shown in Table II, where rows refer to the street direction while columns refer to distance of the observation. Based on knowledge depth K and granularity parameter g the vehicle will have the database dimensioned K x g. Now vehicle stores the incoming observations based on the combination of θ_r and i and by using the mapping function m. The example in Fig. 3 shows the message containing observations from three street sections with angles θ_r having values of $\theta_{r3}=0^o$, $\theta_{r2}=45^o$ and $\theta_{r1}=90^o$. These angles are mapped with m function and θ_{r3}, θ_{r2}, and θ_{r1} are mapped in the following angle bands 0°-29°, 30°-59° and 90°-119° as shown in Fig. 3 and Table II. Therefore, by following the described procedure, the blue vehicle becomes aware that observation valued 11 will be in the first street section in the vicinity whose direction is $\theta r3=0^o$. The observation valued 4 will be in next street section further away whose direction is $\theta_{r2}=45^o$. Finally, the observed value 7 is two street sections away with direction $\theta_{r1}=90^o$. This way the blue vehicle becomes aware about the street sections that are in its immediate vicinity, but out of its communication range.

TABLE II.　　DATABASE STRUCTURE

Direction	Distance		
	i=1	i=2	i=3
θ_r =0°-29°			M=11
θ_r =30°-59°		M=4	
θ_r =…			
θ_r =90°-119°	M=7		
θ_r =…			
θ_r = -1°-30°			

Finally, in order to work properly, our mechanism refreshes the database by erasing the *World Model* each time vehicle changes the street section. This way, the content of the *World Model* is kept up to date and vehicle is provided only with real-time knowledge about observations. Additionally, it should be pointed out that our approach enables direct comparison between measurements and aggregates, because they both refer to same street section in approximately the same time period.

C. Decision

The decision module of the aggregation scheme is responsible for reaching the decision if and when observations should be aggregated or stored in the database. In our mechanism upon the reception of the message the vehicle extracts the measurements together with values of θ and i, and searches the *World Model* accordingly to aggregate the measurements with existing content of *World Model*. Therefore our mechanism assumes that all observations are always being aggregated in corresponding slots in the *World* Model because it is refreshed each time a vehicle enters a new street section and thus every observation received is highly relevant.

D. Fusion

Fusion enables the vehicles to store the incoming measurements into the database by merging them with existing values in the database. There are plenty of data fusion mechanisms and many of them have been widely used in wireless sensor networks. Extensive survey of such works can be found in [23]. We propose the Kalman filter [24] because of several reasons. The Kalman filter is a well-known state estimator that allows detailed description of measurements as a system, depending on the requirements of the application. This means that system can be described with as many variables as needed. These can include intensity, position, location, or something else. Additionally, Kalman filter copes well with noisy measurements and makes optimal estimations.

Prior to describing the system equations of the filter, we introduce the measurements that we use in the paper. As mentioned in Section II, the proposed mechanism can be used for any kind of VANET application, and for this paper, we use traffic congestion management as an example in which vehicle exchange measurements of traffic congestion based on one of our previous works [25]. There we presented the concept of quantification of traffic congestion based on the current values of a vehicle's speed and the speed trend over time. According to this concept each vehicle derives the value of the congestion level independently by monitoring the trend of its current speed. The trend is monitored based on pre-defined values of speed threshold V_t=6m/s and time intervals T={5s, 10s, 15s, 20s, 25s, 30s, 35s, 40s, 45s, 50s} which are used to quantify the level of congestion. The values of V_t and T can be adjusted (calibrated) to fit certain city environment, and aforementioned values are used here as an example scenario. The values of congestion levels are calculated according to the following:

$$\text{if } V_c > V_t \text{ and } T > 5s \text{ then } M = 1 + N, \tag{9}$$

$$\text{else } V_c \leq V_t \text{ and } T = \eta \cdot 5s, \eta = \{2, 3.., 10\} \text{ then } M = \eta + N \tag{10}$$

Here V_c is the current speed of the vehicle, while V_t is the threshold for activating the congestion detection process. M is the level of traffic congestion and can have values between 1 and 10, while T is the time period which refers to time trend of M. Thus, each vehicle obtains noisy measurements with additive white Gaussian noise N, of the traffic congestion, which are then communicated to other vehicles. Each vehicle feeds the Kalman filter with obtained and received measurements in order to fuse them into the *World Model*.

We model the congestion M as the system which is denoted as x and which evolves from the state x_{k-1} to x_k according to the following equations:

$$x_k = Ax_{k-1} + Bu_k + w_{k-1} \tag{11}$$

$$z_k = Hx_k + v_k \tag{12}$$

In equations (11) and (12) k denotes discrete time samples and according to equation (11) the real congestion value x_k is combination of previous value and a control signal u_k and process noise w_k. z_k is the measurement value obtained from other vehicles or on its own, while v_k is the measurement noise, and both v_k and w_k are additive white Gaussian noise with normal probability distribution. The process noise and measurement noise covariance are as follows, respectively:

$$p(w) \sim N(0, Q) \tag{13}$$

$$p(v) \sim N(0, R) \tag{14}$$

A, B and H are matrices or scalars, depending on the application and are linear scaling factors relating to change of state, control input and measurements, which we consider constant. Also, in our case we assume that there is no control signal u_k and that the congestion level M is a system described with only intensity, thus we use scalar Kalman filter. From equations (11) and (12), time update and measurement update equations are derived [24]. This process has been shown many times and we will skip it here and just present the final equations. The time update equations are:

$$x_k^- = Ax_{k-1} + Bu_k \tag{15}$$

$$P_k^- = AP_{k-1}A^T + Q \tag{16}$$

The filter predicts the value of congestion x_k^- which depends from estimated value from previous timestamp x_{k-1} and control signal, while the P_k is the predicted value of error covariance. Once the prediction is done when measurements arrive, the filter uses the measurement update equations:

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1} \tag{17}$$

$$x_k = x_k^- + K_k(z_k - Hx_k^-) \tag{18}$$

$$P_k = (I - K_k H)P_k^- \tag{19}$$

K_k is the Kalman gain. Based on previous equations

vehicles calculate the current (estimated) level of congestion as a function of predicted value, the incoming measurement and the Kalman gain. Therefore, the vehicle estimates the congestion level based on measurements obtained on its own and measurements received from other vehicles.

E. Dissemination

The final goal of the aggregation mechanism is to enable the vehicles in the network to have global knowledge about some traffic parameter by exchanging the minimum number of messages. General clustering mechanisms use the cluster-head selection process which induces additional communication overhead, thus the passive clustering seems more appropriate. In adaptive broadcasting the nodes change their broadcast frequency based on local measurements and some threshold. We want to reduce both broadcast frequency and the number of broadcasting nodes in the network and thus our approach is based on both passive clustering and adaptive broadcasting concepts. The nodes will determine if and when they should broadcast the information based on their observations and the received observations. Ideally, only a small percentage of nodes should decide to broadcast while majority should refrain.

The clustering is achieved by using the current direction angle θ obtained by *LA* algorithm, where all the vehicles with same θ belong to the same cluster. Therefore the streets are segmented in clusters with unique θ and every vehicle is aware of it. The passiveness of clustering, and eventual adaptation of broadcast interval is achieved by refraining the vehicles from broadcasting based on relationship between the vehicle's own estimations of the traffic state and estimations of other vehicles from the same cluster. With this approach, each broadcast decision of every vehicle has location context attached, resulting in reduced number of broadcasts per street section. The formal formulation of the broadcast criteria is:

$$if \ \left(M_0 * M_r\right) == \text{True} \ and \ \theta_o = \theta_r \rightarrow Broadcast \ decision = YES \quad (20)$$

In (20) the symbol * refers to a mathematical condition which describes relationship which depends on the type of application and its requirements. Since we use the application of traffic congestion management as a case study in this paper, the goal is to disseminate the information about the congestion level on the road, thus condition * will be "greater than". M_o and M_r are estimations obtained by the vehicle and received from other vehicles, respectively. Therefore the vehicle will broadcast the message if and only if its local estimation indicates higher level of congestion, than estimation based on the received measurements from the same cluster. The decision making process of *LA* mechanism together with the periodic broadcasting mechanism is described with pseudo-code in Table III.

IV. SIMULATION, ANALYSIS AND EVALUATION

The performance evaluation of the proposed data aggregation mechanism is based on a comprehensive simulation of a real city scenario and VANET based on 802.11p standard. For this purpose the Veins simulation framework [26] was used because it is designed specifically for VANETs, supports full 802.11p standard and enables real-time integration with the traffic mobility simulator. Veins is the framework based on OMNeT++ [27] network simulator and is bi-directionally coupled with SUMO [28], the traffic mobility simulator. The scenario used for the simulation is based on real map of the city of Erlangen in Germany, which is included in the Veins simulator. The map realistically represents urban environment and includes buildings which make radio propagation similar to a real scenario, and thus directly influences the dissemination of messages. The map of Erlangen is showed in Fig. 4. There are three simulation scenarios with different traffic densities used for evaluation of our mechanism including: low with 250 vehicles, medium with 500 and high with 1000 vehicles. Every vehicle has its own route, chosen from randomly generated set of routes, representing the real life situation in which vehicles have independent routes. The simulation setup parameters are shown in Table IV. The vehicles in the simulation used the proposed data aggregation mechanism with the knowledge depth parameter set to $K=3$ and granularity parameter set to $g=12$. The vehicles in the simulation measured the traffic congestion level as described in Section III.

To evaluate the performance of *LA* mechanism we compare the results with the results of the reference simulation. We performed two sets of simulations, one simulating our *LA* mechanism and the other simulating the reference mechanism "*PB*". In the *PB* mechanism the vehicles use location awareness method like in *LA* mechanism to create and maintain their world model and also use the same routes as the vehicles in *LA* mechanism. The only difference is that in PB mechanism the vehicles use periodic broadcasting and not dissemination module like the vehicles in *LA* mechanism. This provides very high level of data accuracy as messages are frequently exchanged. Additionally since most existing data aggregation schemes use periodic broadcasting we wanted to test how *LA* mechanism compares to it. We set the broadcast interval for *PB* to 10 seconds. In the *LA* mechanism the broadcast decision is made every 10 seconds, using the dissemination module of our mechanism as described in Fig. 4.

TABLE III.   PSEUDOCODE OF DISSEMINATION MECHANISMS

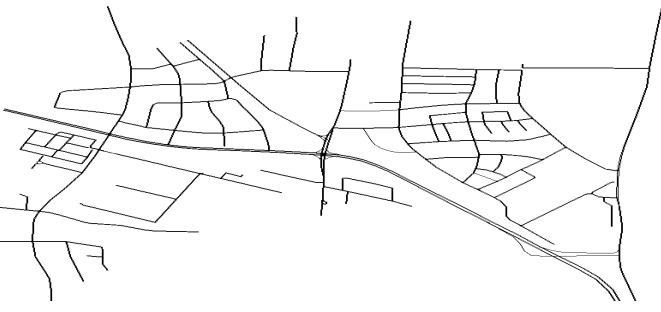|	LA mechanism	Periodic broadcasting
1	timer=0	timer=0
2	if (timer > broadcast_interval) then	if (timer > broadcast_interval) then
3	location=get(current_direction)	broadcast_decision=yes
4	M$_o$=get(own_measurement(location)	timer=0
5	M$_r$=get(received_measurement(location)	end
6	if (M$_o$*M$_r$)= =True)) then	else (timer=timer+1)
7	broadcast_decision=yes	
8	timer=0	
9	end	
10	end	
11	else (timer=timer+1);	

Fig. 4. Erlangen city map taken from Veins simulator.

TABLE IV.   SIMULATION PARAMETERS

Road Traffic Density	Low/medium/high
Number of vehicles	250/500/1000
Total city area	2.25 km x 2.25 km (~5km^2)
Communication standard	IEEE 802.11p
Tx power	10 mW
Rx sensitivity	-89 dBm
MAC Bit rate	18 Mb/s

The mechanism is evaluated from communication efficiency and accuracy point of view.

A. Efficiency

In order to evaluate how the proposed mechanism reflects on network load, Fig. 5 presents the overview of simulations showing average number of sent messages, received messages, the number of times vehicle went into back-off procedure and the number of lost packets. To compare the performance of *LA* mechanism with existing schemes, we implemented *DA2RF* [12] scheme, which performs aggregation by restricting forwarders based on the position of the vehicle's neighbors. Here the vehicle flags itself as non-forwarder jf there is a forwarder in front and behind. The results refer to average values per vehicle and they are normalized to *PB* simulation, thus the data from the chart shows percentages. The chart shows that each of the parameters in both *LA* and *DA2RF* case is significantly lower than the values of results in the *PB* case. Comparison of the results of *LA* and *DA2RF* shows significantly better performance of *LA* mechanism for all evaluated parameters. The reduced number of sent messages results in smaller number of received messages, less contention and eventually in smaller number of lost packets in the communication. According to Fig. 5, the number of messages sent and received in the *LA* mechanism decrease when the traffic density increases, thus the performance of the *LA* mechanism is better when the traffic density is higher.

To further examine the broadcast activity of individual vehicles in the network we calculate the broadcast frequency of every vehicle and show their histogram in Fig. 6. Broadcast frequency is calculated as a ratio of the number of sent messages and the time interval the vehicle spent in the simulation. Since the broadcast interval in the *PB* mechanism was set to 10 seconds, the majority of vehicles have broadcast
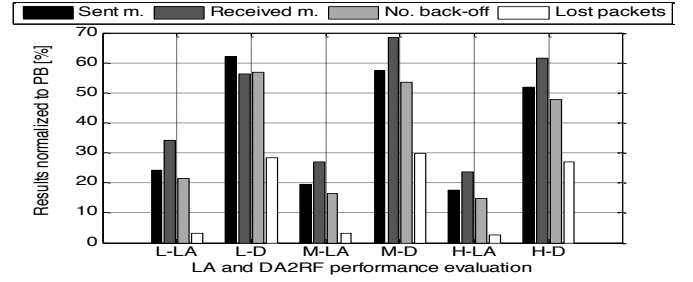

Fig. 5. Network load evaluation of LA and DA2RF (D) schemes showing the average values per vehicle normalized to results of PB mechanism in three traffic density scenarios: low (L), medium (M) and high (H).

frequency 0.1 as shown in Fig. 6. On the other hand, the broadcast frequency for the *LA* mechanism for the majority of vehicles is significantly lower, whereas the small portion of vehicles has broadcast frequency 0.05, or higher. This shows the clustering effect of our scheme, where only limited number of vehicles have high frequency, whilst the most of them broadcast significantly less.

Apart from the overall analysis of broadcast activity of the vehicles, it is important to understand how the broadcast activity in both simulations is spatially distributed in terms of aggregation structures such as street segments and city blocks. To spatially examine the broadcast activity we introduce the measure of *Spatial Communication* (**SC**) and define it as:

$$SC = \frac{Number \ of \ sent \ messages}{Number \ of \ traversed \ street \ segments} \quad (21)$$

SC represents the average number of sent messages per segment per node and shows the average communication activity per street segment. Since this is the average value, ideally *SC* should be as low as possible because it means that the majority of the nodes have less communication activity. We recorded the *SC* values for the *LA* scenario and the *PB* scenario that use the *LA* module for localization. Fig. 7 shows
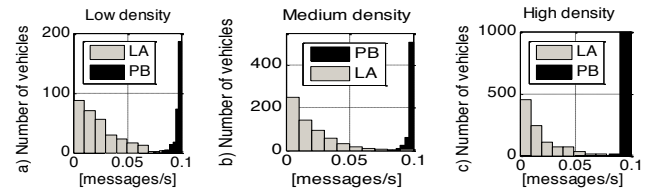

Fig. 6. Distribution of broadcast frequencies in LA and PB mechanism for three traffic densities: a) low, b) medium, c) high
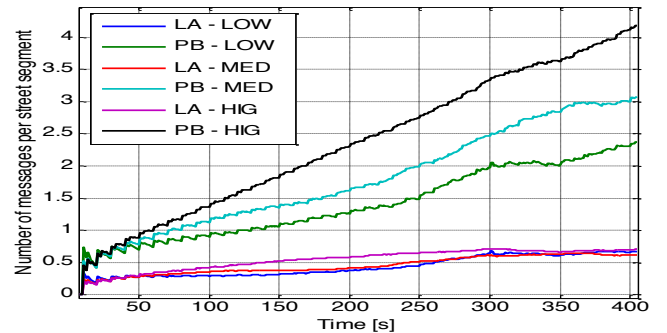

Fig. 7. Average number of sent messages per segment for LA and PB mechanism in traffic with low, medium and high density.

the *LA* mechanism achieving the smallest number of broadcasts per segment and always less than one for all three traffic densities. This means that average vehicle will refrain from broadcasting if using the LA mechanism, while in case of the *PB* this number is at least four times higher.

B. Accuracy

While it reduces the network load, data aggregation mechanism should not compromise the accuracy of the disseminated information. To test the accuracy, the values of vehicles' *World Model (WM)* in the case of *LA* and *PB* mechanisms are compared. Since the *PB* mechanism is chosen as a reference for the "best" way to disseminate the message we introduce the error *E* as a metric of accuracy as:

$$E(i) = WM_{PB}(i) - WM_{LA}(i)$$
$$i \in Z, \ i = 1...(g \cdot K)$$
(22)

Error *E(i)* is defined for each database slot individually as a difference between the same slots of the same vehicle recorded during two scenarios, *PB* and *LA*. All vehicles use the same routes in both simulations, thus their movement will be the same and only broadcasting activity will differ. In our example since $g=12$ and $K=3$, we totally have 36 slots per vehicle and $N=450$ vehicles. The average error per vehicle is defined as:

$$E_{AVG} = \frac{\sum_{i=1}^{N} \sum_{i=1}^{g \cdot K} E(i)}{g \cdot K \cdot N}$$
(23)

E_{AVG} is calculated and normalized to the maximum value of traffic congestion 10, thus the results shown in Fig. 8 refer to percentages, with maximum value of 9% for low density scenario. The error of the LA mechanism is the lowest in the case of high traffic density, and is the largest in low traffic density. This is due to network segmentation problem that occurs when the traffic density is low. To further examine the spatial distribution of errors, Fig. 9 shows the error breakdown per hop for the three hops in the scenario. Each hop refers to 12 database slots where hop 0 refers to the messages that come from vehicles that are within the communication range of the single vehicle. This reflects the streets closest to the vehicle. Hop 1 refers to the previous street section of vehicles within the communication range of a vehicle, while hop 2 refers to the street before the previous one. Errors per hop are defined as:

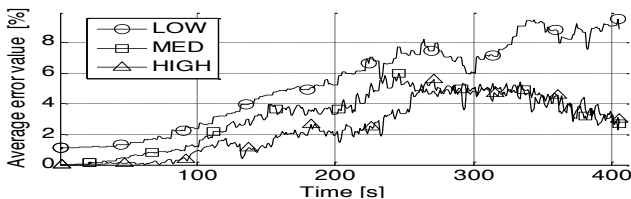$$E_{Hj} = \frac{\sum_{i=1}^{N} \sum_{i=1}^{g} E(i)}{g \cdot N}, \ j = 1,...,K$$
(24)

Fig. 8. Average error value between databases in LA and PB scenario.
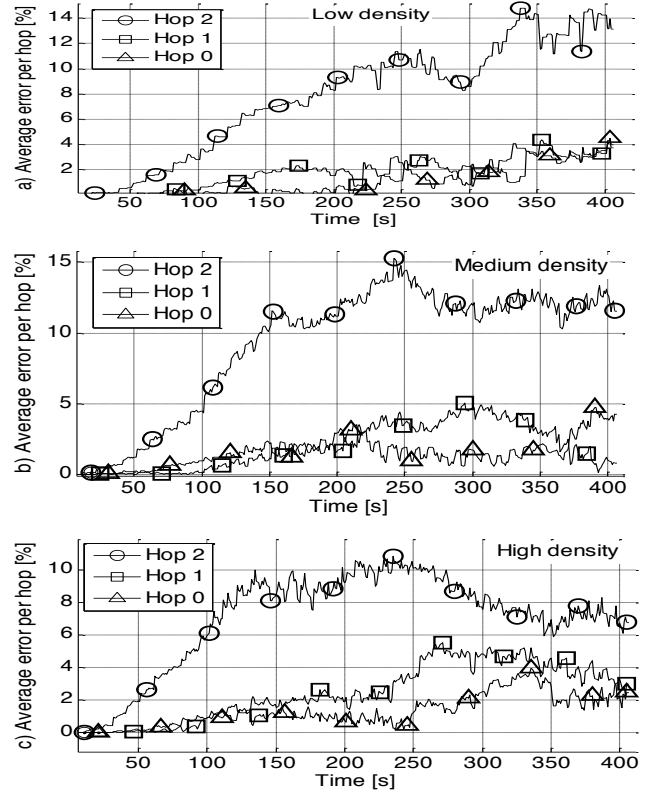

Fig. 9. Average error value LA and PB mechanism per hop for low (a), medium (b) and high (c) traffic densities.

The results of per-hop error analysis show the *LA* mechanism in low and medium traffic scenarios induces 15% error maximum for hop 2, while in the high density case it is 10%. For hop 0 and hop 1, the maximum error is around 5%. The results show that the average accuracy of the disseminated information is not significantly compromised even when the number of broadcasting vehicles and their frequency are reduced. Finally, other than average errors we present the individual errors for the whole *World Model* of single randomly chosen vehicle from simulation. Fig. 10 shows the real error values for all 36 database slots, where in majority of simulation the error was under 10%, in rare cases the error increased to 30% for certain slots and in couple of cases for short period of time the error increased up to 70%.
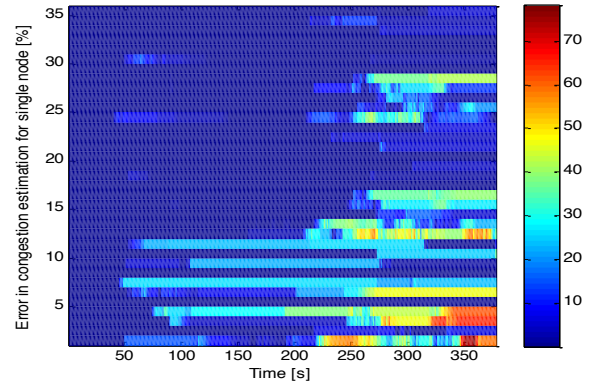

Fig. 10. Error for single vehicle in the simulation.

V. CONCLUSION

In this paper we introduced location aware data aggregation mechanism which can be used to efficiently disseminate messages in VANETs. The mechanism reduces the number of broadcasting nodes and their broadcasting frequency based on the difference of local observations obtained by each vehicle and observations received by surrounding vehicles about the same local area. The vehicles use the *LA* module to develop spatial understanding of the surroundings by creating spatiotemporal database and by assigning the location context to the messages they send. As a result of reducing the broadcasting activity of the nodes, the overall number of dropped packets, collisions and contentions is reduced as well. As a result of data aggregation the accuracy of information that is communicated is not compromised.

REFERENCES

[1] N. Wisitpongphan, O. K. Tonguz, J. S. Parikh, P. Mudalige, F. Bai, and V. Sadekar, "Broadcast storm mitigation techniques in vehicular ad hoc networks", IEEE Wireless Communications, Vol. 14 , Issue 6, 2007.

[2] E. Fasolo and M. Rossi,"In-network aggregation techniques for wireless sensor networks: a survey," IEEE Wireless Communications, Volume 14, Issue 2, April 2007.

[3] R. Rajagopalan and P. K. Varshney,"Data-aggregation techniques in sensor networks: a survey," IEEE Communications Survays and Tutorials, Volume 8, Issue 4, Fourth Quarter 2006.

[4] C. Lochert, B. Scheuermann and M. Mauve,"Probabilistic aggregation for data dissemination in VANETs", Proceedings of the fourth ACM international workshop on Vehicular ad hoc networks VANET '07, USA 2007.

[5] S. Dietzel, E. Schoch, B. Bako, F. Kargl, "A fuzzy logic approach for structure-free aggregation in vehicular ad-hoc networks," Proceedings of the sixth ACM international workshop on VehiculAr InterNETworking, VANET '09.

[6] K. Ibrahim and M. C. Weigle, "CASCADE: cluster-based accurate syntactic compression of aggregated data in VANETs," IEEE GLOBECOM Workshops, November 2008.

[7] L. Wischhof, A. Ebner, H. Rohling, M. Lott and R. Halfmann," SOTIS - a self-organizing traffic information system," IEEE Vehicular Technology Conference 2003,VTC 2003-Spring.

[8] T. Nadeem, S. Dashtinezhad, C. Liao, and L. Iftode,"TrafficView: a scalable traffic monitoring system," IEEE International Conference on Mobile Data Management, 2004.

[9] L. Wischhof, A. Ebner, H.Rohling,"Information dissemination in self-organizing intervehicle networks," IEEE Transactions on Intelligent Transportation Systems, Vol. 6, No. 1, March 2005.

[10] B. Yu, C. Xu and M. Guo,"Adapting forwarding delay control for VANET data aggregation," IEEE Transactions on Parallel and Distributed Systems, Vol. 23, January 2012.

[11] B. Yu, J. Gong, C. Xu, "Catch-up: a data aggregation scheme for vanets," Proceedings of the fifth ACM international workshop on VehiculAr Inter-NETworking VANET '08, New York, NY, USA 2008.

[12] Y. Yuan, J. Luo, W. Yan, T. Zhao, "DA2RF: A data aggregation algorithm by restricting forwarders for VANETs," International Conference on Computing, Networking and Communications (ICNC), 2014

[13] S. Dietzel, J. Petit, F. Kargla and B. Scheuermann, "In-network aggregation for vehicular ad-hoc networks," IEEE Communications Surveys and Tutorials, Vol. 16, No. 4, 2014.

[14] W. Chen, W-L. Jin, A. Regan, "Broadcasting safety information in vehicular networks: issues and approaches," IEEE Network, Volume 24, Issue 1, January 2010.

[15] A. Boukerche, H. A. B. F. Oliveira and E. F. Nakamura, "Vehicular Ad Hoc Networks: A New Challenge for Localization-Based Systems", Computer Communications, Volume 31, Issue 12, July 2008.

[16] T. Kosch, C. J. Adler, S. Eichler, C. Schroth,"The scalability problem of vehicular ad hoc networks and how to solve it," IEEE Wireless Communications, October 2006, Issue 5, Vol. 13.

[17] K. Bilstrup, E. Uhlemann, E. G. Strom, U. Bilstrup, **"**On the ability of the 802.11p MAC method and STDMA to support real-time vehicle-to-vehicle communication," EURASIP Journal on Wireless Communications and Networking - Special issue on wireless access in vehicular environments, January 2009.

[18] T. Tielert, D. Jiang, H. Hartenstein, and L. Delgrossi, "Joint power/rate cogestion control optimizing packet reception in vehicle safety communications," in Proc. of ACM VANET, Taipei, Taiwan, 2013, pp. 51-60

[19] C.-L. Huang, Y. P. Fallah, R. Sengupta, and H. Krishnan, "Adaptive intervehicle communication control for cooperative safety systems," IEEE Network, vol. 24, no. 1, pp. 6-13, 2010.

[20] G. Bansal, J. B. Kenney, and C. E. Rohrs, "LIMERIC: A linear adaptive message rate algorithm for dsrc congestion control," IEEE Transactions on Vehicular Technology, vol. 62, no. 9, pp. 4182– 4197, Nov 2013.

[21] C. Shea, B. Hassanabadi and S. Valaee, "Mobility-based clustering in VANETs using affinity propagation," IEEE Global Telecommunications Conference (GLOBECOM), 2009.

[22] S. Dietzel, F. Kargl, G. Heijenk, F. Schaub, "Modeling in-network aggregation in VANETs," IEEE Communications Magazine, Vol. 49, Issue 11, November 2011.

[23] E. F. Nakamura, "Information fusion for wireless sensor networks: methods, models and classifications," ACM Computing Surveys, Vol. 39, No. 3, August 2007.

[24] R. E. Kalman, "A new approach to linear filtering and prediction problems," J. Basic Eng., Vol.82, No.1, March 1960.

[25] M. Milojevic and V. Rakocevic, "Distributed road traffic congestion quantification using cooperative VANETs," IEEE Mediterranean Ad Hoc Networking Workshop, Piran, June 2014.

[26] C. Sommer, R. German, and F. Dressler, "Bidirectionally coupled network and road traffic simulation for improved IVC analysis," IEEE Transactions on Mobile Computing, vol. 10 (1), pp. 3-15, 2011.

[27] A. Varga, "The OMNeT++ Discrete Event Simulation System,"Proc. European Simulation Multiconf. (ESM '01),June 2001.

[28] 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, 5 (3&4):128-138, December 2012

Milos Milojevic (St. M'09) received his BSc and MSc degrees in electrical engineering from the School of Electrical Engineering, University of Belgrade, Serbia, in 2009 and 2010. He is working towards his PhD degree in Electronic Engineering at City University London that he started in 2011. His research interests include vehicular ad-hoc networks, intelligent transport systems, data aggregation and message dissemination in vehicular ad-hoc networks.

Veselin Rakocevic (M'01) received his Ph.D. degree from Queen Mary, University of London, United Kingdom, in 2002, and his Dipl.Ing. degree from the University of Belgrade, Serbia, in 1998, both in Electronic Engineering. He currently works as Reader in Electronic Engineering at City University London, UK, where he has been since 2002. His main research interest is in the operation of multihop wireless networks, especially addressing the problems of optimal scheduling, rate control, and quality of service, with application in vehicular networks and wireless sensor networks.