An Efficient Destination-Based Data Management Policy for Vehicular Networks

Mildred Madai Caballeros Morales, Rim Haw, Jun Lee, Choong Seon Hong
Computer Engineering Department
Kyung Hee University
South Korea
madai@networking.khu.ac.kr, rhaw@khu.ac.kr, junlee@networking.khu.ac.kr, cshong@khu.ac.kr

Abstract—**The Vehicular Ad hoc Network (VANET) takes the advantage of the relative mobility of the vehicles to opportunistically share dynamic information when they meet, aiming to avoid accidents and traffic jams, get local information of the nearby places, enjoy entertainment applications among others. However, the shared information must be handled properly, saving bandwidth and tackling the unnecessary duplication of data, preventing the scalability problem and congestion. The current solutions of these problems incorporate the concept of Relevance to share the information in a clever way but they do not exploit the advantage of knowing the destination of the vehicles and context-aware services to follow behavior patterns. Therefore, we propose an efficient destination-based data management policy, which follows the vehicular behavior and context to determine the relevance or importance of the information using a publisher/subscriber model and cluster-based dissemination. In order to demonstrate the feasibility of the proposal we implemented sets of simulation in which the results reported an outstanding performance in terms of congestion and delivery ratio.**

Keywords-VANETs; data management; publish/subscribe; relevance; match function; context-aware service.

I. Introduction

Nowadays, the Intelligent Transportation Systems (ITS) have become more popular; thus, the amount of available information in Vehicular Ad hoc Networks (VANETs) increases exponentially and it is necessary to manage diverse information dynamically. This information such as road warning is a key factor to avoid accidents and traffic jams. However, the information cannot be shared with all the vehicles due to the scalability problem which generates congestion and unnecessary duplication of data in large scale networks, increasing the delivery delay and decreasing the delivery rate of critical information (e.g. accident warning). In order to overcome these problems, it has to be decided between the cost of having information about certain event and the cost of not having knowledge of the event. The cost of having information probably causes unnecessary duplication of data and congestion. On the other hand, the cost of not having information could increase the likelihood of accidents or traffic jams. The current solutions address this enigma through different techniques. *Connected Dominating Set* protocol (CDS) [1] measures the costs based on the vehicular density on the road. Similarly, *Direction-aware Function Driven Feedback-augmented Store & Forward Diffusion* (DFD-FSFD) [2] uses propagation functions associated with several targets, considering the density of the road, distance to the target and direction-awareness. However, both solutions do not take into account the importance and the type of information like variables to take the decision. These variables have been considered by Sormani et al. (2006) [3] through a publisher/subscribers (pub/sub) scheme, Kosch et al. (2006) [4] who proposed a relevance function to prioritize the messages based on the interests of the vehicles, and Delta et al. (2009) [5] through a match function. This function evaluates the importance of the event and the direction of the vehicle compared with the location of the target event. But, these solutions only compare the context-aware based on the direction of the vehicles; they do not leverage the destination of the vehicles to determine the relevance of the messages. Therefore, we propose an efficient data management policy. This measures the relevance of the messages taking into account the distance between the destination of the vehicles and the target location of the event and the type of messages to implement the classification of events. The policy implements a pub/sub model, a context-aware service and a match function over a cluster base dissemination algorithm to reduce the number of unnecessary duplication of data, the overhead in large scale networks, the delivery delay of critic data, and to improve the delivery ratio. The simulation results showed that our proposal significantly decreased the congestion and increased the delivery ratio. The rest of the paper is structured as follows. Section II briefly introduces related work. In section III, we explain the scenario and assumptions of the proposal. In section IV, the data management policy is described. Section V shows the performance evaluation of the proposal. Finally, the conclusions of our work are presented in section VI.

II. Related Work

The current studies have managed data following either the mobility and road-state criteria, the context of the messages criteria or both. CDS [1] and DFD-FSFD [2] follow the criteria of mobility and road-state. CDS determines the number of duplication of information by selecting a path in the roads on which the information is sent, this selection is based on the densities of the paths. DFD-FSFD is based on a propagation function that encodes the destination areas of each message as well as the roads that it should follow to reach such areas. However, these solutions pay little attention in the delay tolerance of the messages because they do not differentiate

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by MEST (No. 010-0027645). Dr. CS Hong is corresponding author.
The simulation data were obtained from Wireless and Sensor networks Lab (WnSN). Shanghai Jiao Tong University.

between critical and normal messages. On the other hand, Hybrid publish/subscribe [3] and [4] consider the different context of the messages and they disseminate information mainly to subscribers without affecting non-interested vehicles to allow for a benefit-oriented message dissemination scheme, ensuring that the messages are propagated to where they are needed as efficiently as possible. But they do not have an accurate mobility pattern to disseminate the messages in a clever and efficient way, following the mobility pattern. *Vehicular Event Sharing Peer-to-peer Architecture* (VESPA) [5] follows the two criteria taking into account the relevance of the data to the drivers and the distance between the estimation of the future location (direction) of the vehicles and the position of the event, based on the concept of Encounter Probability. However, VESPA has tended to focus on the direction, rather than the destination of the vehicle and in certain cases the direction is not accurate enough.

III. SCENARIO

We considered an urban scenario with an Intelligent Transportation System (ITS) as shown in Fig. 1, in which the vehicles are interested in information to avoid accidents (road warnings), reduce congestion, update local maps, and provide entertainment (web browsing, email, and media streaming). All the vehicles employ Global Position System (GPS) or Navigation System (NS). The vehicles are organized by clusters, where the cluster head is responsible of disseminating the information to the cluster members. The clusters are formed considering the destination of the vehicles [6].

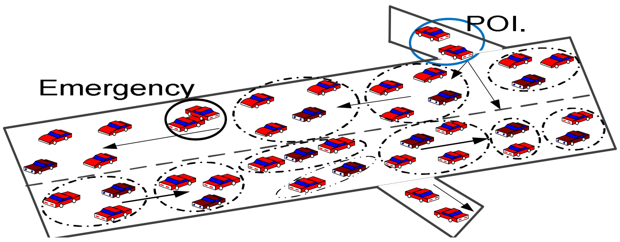

Figure 1. Urban Scenario

The vehicles rebroadcast the messages to reach the target location or Point of Interest (POI) sending the messages to vehicles that are not interested causing the scalability problem, duplication of data and congestion. In order to address these problems, we propose an efficient policy to manage the data according to the relevance of each message based on the destination of the vehicle.

A. *Assumptions*

In the network there are three entities: mobile nodes (MN), cluster head (CH), and base station (BS). The destination of each vehicle is known. This assumption is realistic in case of real-world systems because the GPS or NS allows drivers to enter their travel destination. There is historical information of the traffic and density of each street; these data is gathered every hour and at the end of the day the data is sent to the local server. The communication is content-based. The routing is geographic-based. A pub/sub strategy is utilized, where the CH can be both information providers (publishers) and information consumers (subscribers), and the MN are subscribers.

IV. DATA MANAGEMENT POLICY

The proper management of data is a key factor to overcome the scalability and duplication of data challenges within VANETs. The addition of vehicles to dense networks with poor data management policies, where the vehicles gossips data to all nearby neighbors, has a direct impact on the performance and increases the unnecessary duplication of data generating congestion, bottleneck and packet loss. Therefore, the data management policies have to deal with limited bandwidth considering the characteristics of the networks, such as the type of information that is shared, the density of the network and the mobility behavior of the vehicles. We propose a data management policy, which considers the characteristics of the networks, the mobility behavior, the dynamism and the relevance of the data through a pub/sub model and a match function.

A. *Publisher/Subscriber Model*

This model is characterized by interaction within distributed environments with asynchronous and many-to-many communications [7]. The model has two principal entities; the subscribers and the publishers. Subscribers consume specific type of information according to their interests and the publishers provide certain type of information, additionally a broker entity is required to disseminate the information to all the consumers. The vehicles can be publishers, subscribers and brokers. The explanation of the pub/sub model applied to ITS applications within VANETs is divided into the ITS application modules and context-aware service framework.

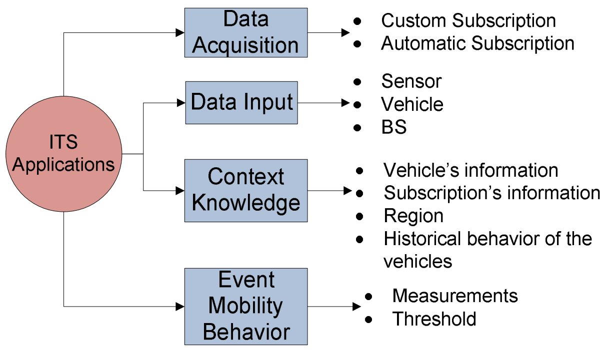

Figure 2. ITS Application Modules

ITS Application Modules

These applications take into account the information included in the Data Acquisition, Data Input, Context-Knowledge and Event Mobility modules as Fig. 2 shows.

a) Data Acquisition: It describes how the drivers or vehicles request the services. Vehicle may act as a subscriber, by expressing its interests in a certain set of messages. The subscriptions are divided into automatic and custom subscriptions [3], [5] [1]. Custom subscriptions are specific

[1] Events relevant to the vehicles (automatic subscriptions) and relevant for the drivers (custom subscriptions).

interests of the drivers, such as, information on nearby restaurants, gas stations, parking lots or hotels. Automatic subscriptions are automatically calculated depending on the current location. For example, the vehicle will subscribe to accident or traffic warning service that affect the current route concerning the location of the vehicles.

b) Data Input: This module includes the sources of information such as sensors, vehicles or BS.

c) Context-Knowledge: It includes all the information that comprises the context, which becomes an essential resource to improve the performance of the communications. We include in this module the information of vehicles e.g, id, location, speed and destination (relative and final destination) [2]; subscriptions e.g., id, type and description; region [3] e.g, id, location and road maps; and the historical behavior of the vehicles e.g, statistics of paths and density in the roads which helps to predict and recognize patterns.

d) Event Mobility Behavior: This module controls the mobility behavior through measurements. These measurements determine the thresholds to respond in real time according to the current status of the network.

Context-Aware Service

The meta-model of ContextUML [8] is used to formalize the design and development of ITS context-aware service in conjunction with the pub/sub strategy within the vehicular networks. ITS context-aware service adapts the behavior of the user based on the context information corresponding to a specific location. Fig. 3 illustrates how to select the messages that will be published by the publisher, such as the automatic or custom subscriptions according to the current location of the subscribed vehicles. The ITS context-aware service consists of three key elements which are described below.

a) Context: The Position, Custom Service, Event, Service Position and Event Position constitute the context of the service. The atomic context (stereotype itsuml.Atomic Context) represents the simple entities (i.e. Position, Custom Service and Event). The "Position" comprises the latitude and longitude coordinates of the vehicles mapped into (x, y) points on the road of a specific region; the "Custom Service" is the information of the services requested by the user (drivers) such as information of restaurants, gas stations, hotels, available parking, advertisements of entertainment or sales; finally, the "Event" is the information related to the accidents, traffic flow, road warning (ice, oil, water, low bridges). On the other hand, the "Service Position" and "Event Position" are composite context (stereotype itsuml.compositeContext), requiring the aggregation of two atomic context, for instance "Service Position" are the customs services available in a specific region, because it is constituted by "service" and "position" context, similarly "Event Position" includes the context of "event" and "position", defining the road warnings when an event on a specific road appears. The Event Position context addresses all the automatic subscriptions while the Service Position addresses the custom subscriptions.

b) Context-Aware Objects: The context-aware objects are basically service components. The main service in ITS context-aware is "Vehicle's Subscriptions" (stereotype itsuml.service), which has to select the publications to be sent to the subscribed vehicles through the operation of "Match Subscription". The input and output messages of the Match Subscription are represented by the stereotype itsuml.message. The input messages are the list of custom subscriptions for each vehicle and the location (x, y) because the context and automatic subscription are geographically dependent. The output messages are the set of publications that match with the preferences of each vehicle (subscriptions), and then these publications are published or sent to the vehicle.

c) Context Awareness Mechanism: To provide the ITS context-Aware service we specify two mechanisms The "Service Position Binding" (stereotype itsum.contxBindig) connecting the "Service Position" and the input message (Input) of "Match Subscription", and the "Event Position Trigger" with the stereotype itsuml.contextTriggering, implementing the MatchTargetLocation function and data filtering between "Event Position" and "Vehicle's Location", when a specific event appears.

B. Match Function

The match function is the core of the match subscription service and is performed when a message arrives. The explanation of this function comprises of the Data Structure and the Process.

Data Structure

The data structure of the messages includes the following information:

a) Publication ID: It is the identification of the message.

b) Home location: Is the location (x, y) and region id, where the message was created or the event occured.

c) Creation Time (CT): It is the date and time when the message was created.

d) Publication Type (PT): This item classifies the messages. The publication type consists of three elements:

- Category: The classification of the messages is based on the nature of the event emphasizing two categories: automatic and custom messages.
- Priority: The priority is a value to rank the urgency and importance of each message. The priority takes values between 1 to n, where n is the lowest priority and 1 is the highest priority.
- Path: It indicates the maximum number of the paths in which the information should be sent. The value of path is 0 when the information is sent to all the possible paths.

[2] Relative destination is the nearest destination according to the current region and final destination is the final destination which is registered in the GPS or NS, both in terms of (x, y) coordinates.

[3] Region is a part of the city compounded by streets and avenues.

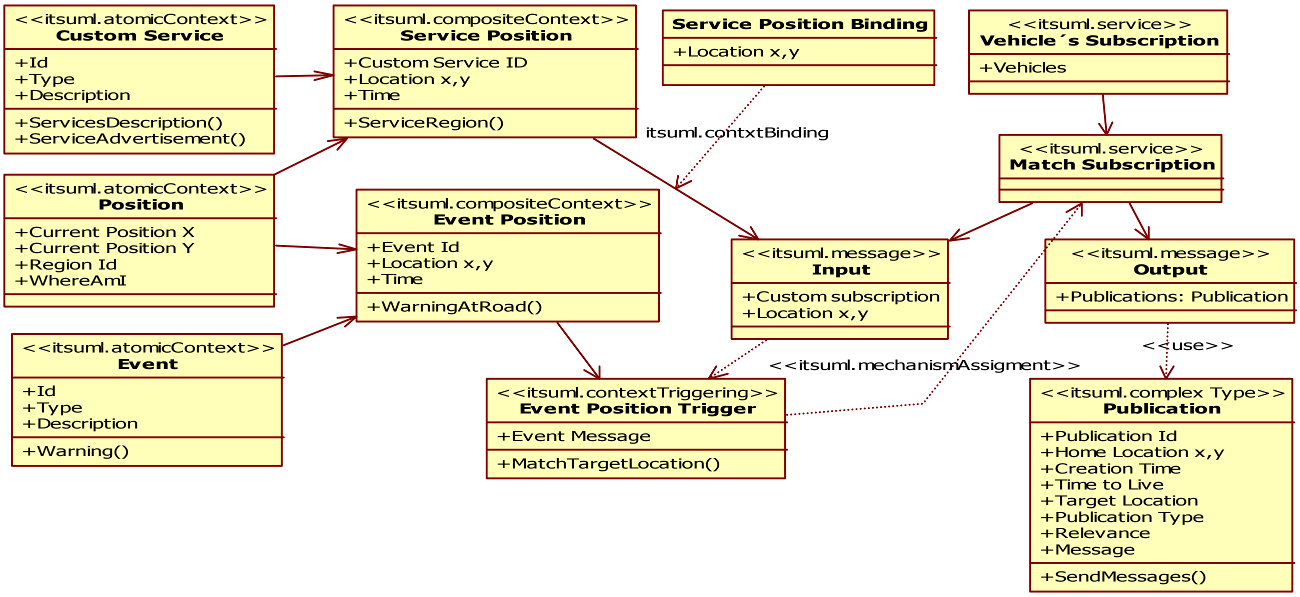

Figure 3. ITS Context-Aware Service

- Type id: It identifies the different types of events such as accident warning, restaurant advertisement, etc.

e) Time to Live (TTL): Is the threshold of aging for each message, the TTL value depends on the type of publication because the TTL duration of each type of event is different, for example, a message informing about parking lot availability is a message of short duration; on the other hand, the duration of a low bridge warning is longer.

f) Target Location (TL): The locations are managed by mapping the latitude and longitude coordinates of the vehicles to points (x, y) on the road. The target area to send the information is a zone (zone of relevance ZOR) or point of interest (POI). The messages can have a single record (x, y) or an array of records according to the numbers of locations where the message should be sent. The target location describes the center point (x, y) of the region which has a radius R.

g) Re-diffussion: It is the number of times that the message has been sent previously. [5][4]

h) Message: Is the detailed description of the message.

Process

The execution of the match function constitutes of the next steps.

a) Aging threshold: It compares the aging threshold of the message against the time that the message has traveled in the network. Δa quantifies this comparison through equation (1) where "t" is the current time.

The equation shows two cases; when the time that the message is traveling in the network $(t-CT)$ is less than or equal to the TTL, the value of Δa is between 0 to 1; otherwise the value of Δa is greater than the TTL value.

$$\Delta a \begin{cases} \frac{t-CT}{TTL} : (t-CT) \le TTL, \\ 1-CT, \text{otherwise.} \end{cases} \quad (1)$$

The equation (1) warrants the penalty of all messages that surpassed its TTL value.

b) Priority: It is the calculation of the priority of the message which is a value between 1 to n.

c) Distance threshold (Δd): The calculation of Δd requires two values; the distance from the target location of the message to the current location and to the relative destination of the vehicle.

- Distance to the current location ΔCL : It is the difference in distance between the current location of the vehicle (CL) and the target location (TL) in terms of (x, y) coordinates calculated through the equation (2).

$$\Delta CL = \sqrt{(CL_x - TL_x)^2 + (CL_y - TL_y)^2} \quad (2)$$

- Distance to the relative destination ΔD : It is the difference in distance between the relative destination of the vehicle (RD) and the target location (TL) in terms of (x, y) coordinates. This difference can be calculated with the information of maps as shown in equation (3).

[4] Similar to HopNumber in the section 3.1 Event Classification

$$\Delta D = \sqrt{\left(P_x - TL_x\right)^2 + (P_y - TL_y)^2}, P \in path \quad (3)$$

Where, P is the nearest point to the target location. This point is part of the path of the vehicle. However, the availability of maps in some remote places is limited, in this case ΔD is calculated with the relative destination.

$$\Delta D = \sqrt{\left(RD_x - TL_x\right)^2 + \left(RD_y - TL_y\right)^2} \quad (4)$$

The calculation of Δd emphasizes the relevance of the messages when the vehicles are within the radius R of the target zone of the message as presented in equation (5).

$$\Delta d \begin{cases} \frac{\Delta CL}{R} : \Delta CL \le R, \\ \Delta D, otherwise. \end{cases} \quad (5)$$

The value of Δd is calculated for each target location selecting the minimum Δd .

d) Match Subscription (MS): The match subscription is a binary value, 0 when no subscription matches with the message and 1 when at least one subscription mathes it with the message. If the message belongs to the automatic category, the match subscription is 1 if $\Delta CL \le$ R otherwise is 0. On the other hand if the message belongs to the custom category, the match subscription is 1 if at least one of the lists of subscriptions has the same type of message.

Finally, the relevance of the messages is calculated taking into account the ageing threshold, priority, distance threshold and the match subscription.

$$\mathrm{Re}\,levance = w_1\Delta a + w_2 Priority + w_3\Delta d + w_4 MS \quad (6)$$

In the equation (6) the weights w1, w2, w3 and w4 determine the importance of each variable.

The publisher selects which messages publish or delete according to the "relevance" value, to send to each vehicle only the important or interesting messages for it, at the most opportune time avoiding congestion and duplication of data. Low values of "relevance" have more importance. The effectiveness of managing data through the relevance and context-aware, which uses the destination of the vehicles as a key factor is evident in the dissemination algorithm.

V. Performance Evaluation

The evaluation of the proposal is performed through a set of simulations. These simulations were developed in Java JDeveloper due to the diversity and flexibility of the scenario makes the task of configuring and adapting difficult in a generic simulator. In order to perform these simulations, we included the information of 4000 taxis of Shanghai city obtained from the Wireless and Sensor Networks laboratory of Shanghai Jiao Tong University [1]. This data contains a collection of GPS information of real-time locations of taxis in the urban city of Shanghai. The GPS information comprises the Id of taxi, the longitude, latitude and angle of the current location of the taxi, speed of the taxi, and time that GPS record was sent. This information provides real traces of the taxis, fitting each of the GPS registers to maps of the roads, with help of the interpolation in cases when the data obtained by GPS did not provide a clear trajectory. Additionally we modified this data by adding the destination of the taxi's trajectories. The general parameters that we used during the simulation are listed in Table I.

TABLE I. Simulation Parameters

Simulation Area	5000*5000 m
Simulation Time	86,400 s
Average Speed	9.75 m/s
Range Speed	0 m/s -30m/s
Vehicular Density	4000 taxies
Transmission Range TR	300 m
Packet size	1kb

Each taxi followed a trajectory, when it reached its destination was removed from the network and relocated to its initial location. This process was repeated during the simulation time. The taxis were grouped in clusters and we assumed that detailed map of the road are available. The metrics considers in our performance evaluation are: congestion (Messages received per hour) and delivery ratio (% of the delivered messages).

A. Simulation Results:

In order to evaluate the message delivery of the network we considered two scenarios, when the information is propagated to the interested vehicles and when is propagated to non-interested vehicles. The number of sent messages is distributed according to a Poisson distribution [9] as is indicated in Fig. 4, where 40% of the messages were automatic subscriptions and 60% custom subscriptions assuming that the life time of each message was 300 seconds and a new message was generated every 120 seconds. The distribution of the messages is directly related with the density of the vehicles during the day.

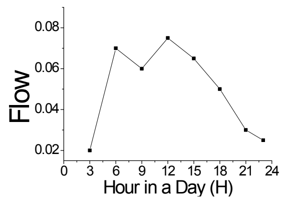

Figure 4. Distribution of Messages

Fig. 5 and Fig. 6 show the congestion and unnecessary duplication of data in terms of the average number of messages received by subscribers and non-subscribers, respectively, during the simulation time (24 hours). The congestion of our proposal called DBDM (Destination-Based Data Management Policy) was compared with the flooding strategy [5], CDS, Hybrid publish/subscribe and VESPA, revealing some interesting trends. The flooding strategy followed the tendency of the density. The congestion increases in rush hours because

[5] Flooding strategy sends the messages to all the neighbors, often causing congestion in the network

the messages being sent to all its neighbors, opposite to the behavior of CDS which maintained a stable number of messages during high vehicular density between hour 12 and 15 may be due to the efficiency of selecting the best path according to the density on the road. However, CDS does not distinguish between interested and non-interested vehicles, sending unnecessary messages to the vehicles. Hybrid publish/subscribe and VESPA had a similar trend. Both techniques consider the vehicles which are interested in the information. The small difference of tendencies is probably a consequence of the mobility and road state criteria implemented by VESPA. However, DBDM had less congestion than VESPA due to DBDM takes the advantage of the group mobility through the cluster-base dissemination. Additionally, DBDM considers the destination of the vehicles to improve the efficiency of the context-aware avoiding the congestion and unnecessary duplication of data. All the strategies sent sometimes unnecessary information to the vehicles because these vehicles work as carriers to send the information to the interested vehicles.

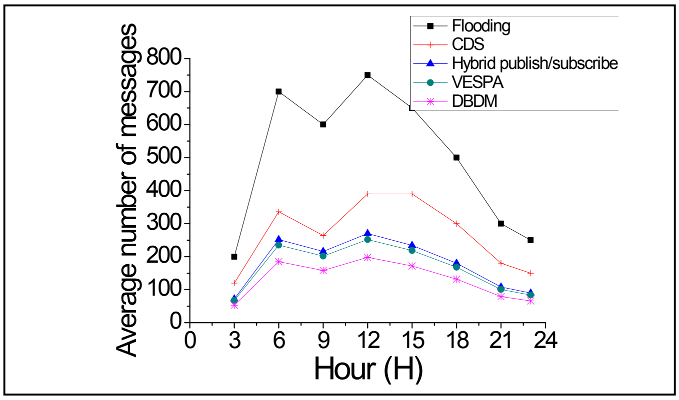

Figure 5. Average Number of Messages Received by Subscribers

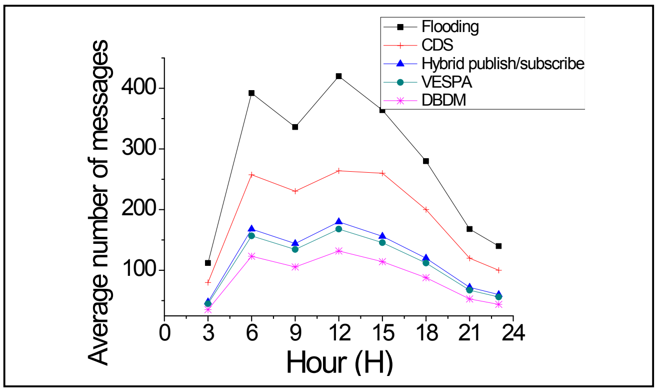

Figure 6. Average Number of Messages Received by Non-subscribers

The low overhead of DBDM described before could lead to low delivery rate. However, Fig. 7 demonstrates that DBDM performance in terms of delivery rate was about 94% in the worst case and about 99% in the great majority of cases. Additionally, DBDM performs well with low and high vehicular density due to the match function propagates the information to the right vehicles and the destination-based cluster algorithm exploits the group mobility.

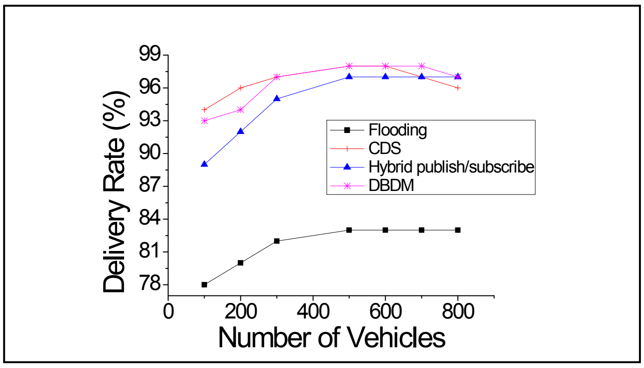

Figure 7. Delivery Ratio

VI. CONCLUSION

To scale properly VANET requires an effective management of mobile data to avoid congestions, unnecessary duplications of data and low delivery rate especially for critical information. Therefore, we proposed an efficient destination-based data management policy which considers the mobility and road-state criteria and the context of the messages criteria through a context-aware service based on the geographic location of the vehicles and the data. The policy implements a pub/sub strategy and a destination-based cluster algorithm to disseminate the data. The simulation results showed outstanding performance in terms of congestion and delivery ratio. The main contributions of the proposal are the adaptability to dynamic and diverse data, the consideration of information priority to disseminate data and the implementation of group mobility pattern decreasing the unnecessary data duplication.

REFERENCES

[1] K. Liu, M. Li, Y. Liu, M. Li, X.-Yang Li, and L. Guibas, "Chasing the hidden Connectivity in Urban Vehicular Networks," 2009.

[2] D. Sormani, G. Turconi, P. Costa, D. Frey, M. Migliavacca and L. Mottola, "Towards Lightweight Information Dissemination in Inter-Vehicular Networks," VANET'06, 2006.

[3] I. Leontiadis, P. Costa, C. Mascolo, "A Hybrid Approach for content-based publish/subscribe in vehicular networks," in Journal of Pervasive and Mobile Computing, 2009.

[4] T. Kosch, C. J. Adler, S. Eichler, C. Schroth and M. Strassberger, "The Scalability Problem of Vehicular Ad Hoc Networks and How to Solve It," IEEE Wireless Communication, 1536-1284. 2006.

[5] T. Delot, N. Cenerario, S. Ilarri, "Vehicular event sharing with a mobile peer-to-peer architecture VESPA," in Journal Transportantion Research Part C, 2009.

[6] M. Caballeros, C. S. Hong, "An Adaptable Mobility-Aware Clustering Algorithm in Vehicular Networks," The 13 Asia-Pacific Network Operations and Management Symposium, September 2011.

[7] J. Barcelo-Ordinas, C. Casetti, G. Karlsson, S. Sargento, "Information Mobility: a New Paradigm for Wireless Content Dissemination," in proceedings of 2nd European Wireless Technology Conference 2009.

[8] Q. Z. Sheng, B. Benatallah, "ContextUML: A UML-Based Modeling Language for Model-Driven Development of Context-Aware Web Services," in International Conference on Mobile Business (ICMB 2005), Sydney, Australia.

[9] J. Kim, V. Sridhara, S. Bohacek, "Realistic mobility simulation of urban mesh networks", in Journal Ad Hoc Networks, United States, 2009.