Advanced Diffusion of Classified Data in Vehicular Sensor Networks

Nadia Haddadou[1], Abderrezak Rachedi[1], and Yacine Ghamri-Doudane[1, 2]
[1]Université Paris-Est - Gaspard Monge Computer Science Laboratory (LIGM - UMR 8049),
75420 Champs sur Marne, France
[2]ENSIIE, 1 Square de la résistance, 91025 Evry Cedex, France

Abstract—**In this paper, we propose a newly distributed protocol called ADCD to manage information harvesting, and distribution in Vehicular Sensor Networks (VSN). The concept of ADCD is based on the characterization of sensed information (i.e. its importance, location and time of collection) and the diffusion of this information accordingly. Furthermore, ADCD uses an adaptive broadcasting strategy to avoid overwhelming users with messages for which they have no interests. Thanks to this adaptive broadcasting strategy, ADCD limits the generated overhead avoiding network congestions as well as long latency to deliver the harvested information, which are the main limitations of other existing protocols. Moreover, it is designed to be flexible regarding the use of roadside units or not, which is not the case in other schemes in the literature. To reach its objectives, ADCD operations are divided into three steps: (i) classification of data and the identification of their target area of diffusion, (ii) data-centric election of the set of broadcasters to avoid broadcasting redundancy, and (iii) iterative process for data dispatching in a targeted area. Performance evaluation shows that the ADCD protocol allows for mitigating the information redundancy and its delivery with an adequate latency while making the reception of interesting data for the drivers (related to their location) more adapted. Moreover, the ADCD protocol reduces the overhead by 90% compared to the classical broadcast and an adapted version of MobEyes [4]. The ADCD overhead is kept stable whatever the vehicular density.**

Keywords-Vehicular Sensor Network; Data Dissemination; Performance Evaluation

I. INTRODUCTION

Vehicular sensor networks have been widely investigated during the last few years as they offer applications for road safety and the driver's/passenger's comfort. They are considered as the most effective and cheapest way to avoid congestion on the roads leading to minimizing the consumption of fuel, and the time spent on the road. They also have fewer operational limitations (memory, processing, energy…) than basic Wireless Sensor Networks (WSN). However, the large amount of generated information and the frequent topology changes (density, topology, neighboring nodes…) make the existing solutions in the WSN field ineffective for this new kind of networks.

The main issues in VSN reside in the data harvesting and its dissemination in such a large-scale network, characterized by the frequent topology changes and network partitioning. In this paper we will focus on data harvesting and distribution based on targeting the concerned nodes, i.e. the nodes that should receive the information and can be interested in its content according to its geo-location and the harvested data specificities i.e. importance, location and time of collection.

Our main motivation in this work comes from the fact that the congestion and the redundancy induced in order to ensure the reception of relevant messages by the concerned nodes, is an important drawback of previous works. Thus, we propose a new protocol for Advanced Diffusion of Classified Data (ADCD). ADCD targets the receptors to avoid both redundancy and network congestion. This is performed by inserting intelligence in the diffusion process. Therefore, ADCD differentiates the sensed data according to their relevance and period of validity, in order to better predict its importance for the other nodes in the network. Once the first step achieved, the collector node customizes the diffusion by electing a number of broadcasters from its neighborhood according to the importance of the message, while reducing the redundancy by binding the election process to different criteria (node density, node positions). Finally, the third step consists in verifying whether the limits of the message broadcast, in terms of targeted broadcasting area and content validity in time, had been reached or not. By doing so, we show in this paper that the overhead/redundancy induced by ADCD is lower than the one induced by other existing schemes while the reception probability by concerned nodes is maintained high.

The rest of this paper is organized as follows: Section II presents and summarizes the main related works. Section III focuses on the description of the proposed Advanced Diffusion Protocol of Classified Data (ADCD) protocol. Then, Section IV presents the performance evaluation of ADCD and compares it to two other existing protocols i.e. classical broadcast and an adapted version of the MobEyes [4] protocols. Finally, Section V draws our conclusions and suggests future work.

II. RELATED WORK

Many studies have recently been made in the field of vehicular sensor networks. Among the main topics addressed by the research community in this domain, we find: data harvesting and data diffusion. Below, we explain the limitation of the related literature. These limitations had motivated our current work.

978-1-4244-9538-2/11/$26.00 ©2011 IEEE

A. *Data Harvesting*

Data harvesting is considered as the first step of most of VSN applications. The aim of data harvesting in VSNs is to facilitate the two following steps in the VSN traditional life-cycle: i.e. data search and data dissemination.

The main data harvesting approach proposed in the literature is the DataTaxis [5] algorithm. This algorithm is inspired by the behavior of insects during their search for food. In Data Taxis, this behavior is thus adapted to vehicles allowing to cover a large area and to harvest the maximum amount of data effectively. This is done by placing a large number of agents in concentrated information areas. This is performed by the use of stigmergy, a communication mechanism used by insects such as ants. The redundancy of collected data is decreased. This idea is difficult to adapt to insure safety applications with time constraints, due to the important mobility of VSNs. Indeed, in the case of safety applications, it is difficult to forecast where the events will happen. So, it is difficult to predict where the data harvesting agents need to be placed unless a large number of agents is used. In the latter case the data harvesting scheme will be very complex to implement.

B. *Data Diffusion*

Defining a data sharing strategy depends on the type of application that uses this data as well as the required quality of service by this application. Among the existing works, two methods are frequently used. The first one consists in an immediate sending of the harvested data to support real-time applications. The drawback of such a method is the high redundancy level that is induced and the risk of network congestion. The second method performs first data processing, where vehicles only exchange summaries at regular periods of time [6]. The goal of this method is to reduce the amount of exchanged information on the network to avoid any congestion. However, the spreading of real-time applications is impossible and an incomplete data diffusion is probable [4]. We position ourselves in the middle of these two ideas, the pre-processing made in ADCD avoids sending messages repeatedly and limits the rebroadcasting using different concepts to ensure good latency and reception ratio.

To perform dissemination that takes into account the degree of importance of each individual data, a classification between different kinds of messages, according to their quality of service requirements in terms of delay and throughput, is established. The first class concerns emergency messages, like accidents notifications. These messages are short and need to be transmitted with a high propagation velocity to insure a real time service [1], [2]. For other situations, particularly the less urgent ones, a warning message is used to catch the driver's attention. These messages concern for instance, Driver Assistance Applications. The last type is used to develop collaborative driving; where drivers share information about the density and the average speed of vehicles within the road allowing to reduce the risk of traffic jams. Warning messages are less important than emergency ones but a higher rate is required for them. However, the question of how to introduce this classification in the dissemination phase remains vague. That is why ADCD protocol classifies the information before their diffusion in the network and specifies the corresponding diffusion strategy to each one within the broadcasted message. This is done so as to guarantee that the diffused messages will be received by a maximum number of concerned nodes.

Another futuristic VSN application of interest is content-sharing, as proposed in [7]. The technology used for this is peer-to-peer technology which remains innovative for this kind of network. Its content will include delay-sensitive and delay-tolerant data. These two concepts are also taken into account in our proposed protocol ADCD through the time "Mode" parameter. In addition, content sharing applications among vehicles need an efficient distribution mechanism.

The approach proposed in [3] differs from previous ones which use classical criteria for data classification. The authors develop a generic network architecture to support futuristic VSN applications to focus on space, time and user's interest in relation to information during its distribution. The paper assumes different kinds of applications, such as safety applications, location-based services, city-wide alerts and interactive services with their spatial, temporal and interest scope which define the area and time of spreading, followed by the type of targeted receiving vehicles. Setting up these applications supposes the use of roadside units which limits their spread. Furthermore, there is no specific distribution protocol proposed in this paper. Although our protocol, ADCD, shares the same philosophy about data classification, it proposes a newly adapted strategy for data dissemination which takes full advantage of data classification.

Another very interesting approach for data diffusion in VSNs is the Smart Mobs for Urban Monitoring with a Vehicular Sensor Network, MobEyes [4]. MobEyes ensures proactive urban monitoring by taking advantage of vehicle mobility. The exchanged messages can include 2 to 10 minutes of summaries regarding the captured data. This method targets a specific surveillance application where data are harvested, by police officers. A bloom filter is applied to retrieve the missing information tightly linked to avoiding redundancy. However, MobEyes is proposed for a specific application and its generalization is not straightforward. That is why, we propose a new protocol ADCD with more parameters which make it more general and thus it can be applied to different applications.

C. *Discussion*

Future application constraints for VSNs are tightened, with more requirements for network performance. As explained above, existing solutions cannot always meet the needs. A new view has to be established to respond to the remaining questions such as the dilemma of receiving information or not according to its pertinence and the delivery time required for its reception while avoiding an overhead.

Our concept fits the identified constraints thanks to the data characterization and the corresponding enhanced diffusion.

III. ADVANCED DIFFUSION OF CLASSIFIED DATA

In this section, we describe our proposed protocol called Advanced Diffusion of Classified Data (ADCD). Its aim is to empower safety applications with vehicular sensor networks. For more robustness, the protocol is indifferent to the existence of roadside units. It merely supposes that a minimum number of vehicles are equipped with sensors and, means of communication, accepting to collaborate with other vehicles for safety information distribution. This protocol allows for regular harvesting along with smart information sharing between vehicles to avoid any risk of congestion or starvation in isolated areas.

We consider that each vehicle has to receive all emergency and local information messages available within a certain perimeter and under a corresponding time of validity. This is done with the aim of keeping it continuously aware of current traffic conditions without inappropriately flooding other areas with information of non-interest to them. Our approach is generic enough to make it possible to do that but also to distribute information on the entire network is possible if needed.

ADCD is based on three main parts: (i) harvesting and data classification, (ii) election of broadcaster, and data sharing, and (iii) iterative rebroadcast with corresponding scope.

A. Data Harvesting

A vehicle equipped with different types of sensors can collect several kinds of data. To illustrate our approach, we take the following message/data classes of interest: accident, traffic jam, landslide, risk of slipping, car crash, roadwork, number of vehicle, failure of a traffic light, road density. We consider that each piece of information depends on the region where it was collected. Thus, its diffusion is only useful in its surroundings during a fixed period of time to avoid the transmission of old information.

In order to carry out this concept, we characterize information with two parameters: *class* and *mode*. A class represents the importance level of information; it is used to define the broadcasting area within the VSN. The *mode* is a value in a scale representing the period of validity of the data.

ADCD defines an interval $[\partial_{min}, \partial_{max}]$ for the classes and modes. The most urgent messages and the longest in validity are characterized as class ∂_{max}. A piece of information is represented by C*xy*, where *x* represents the class and *y* the mode.

The vehicles concerned by particular information are those belonging to the targeted broadcast area and for which we advocate interest in receiving this information. In order to target these vehicles during the transmission, we attribute a diffusion perimeter, as a square centered upon with the coordinates of the collected data. The length of each square side corresponds to the class of data. "Table. 1" gives an example of how ADCD can be used for a traffic safety application. In this example five classes are defined. For each class, a square is associated for each one defining its targeted diffusion area ($200*200m^2$, $300*300m^2$, $400*400m^2$, $600*600m^2$ and $800*800m^2$ to the first, second, third, fourth and fifth class, respectively). Hence, in this example, a road density message should not exceed the perimeter of $200*200$ m^2.

TABLE I. INFORMATION CHARACTERISTICS OF SHARED DATA

Information	Class	Mode
Accident	5	3
traffic jam	4	4
Landslide	3	5
risk of slipping	3	4
car crash	3	2
Roadwork	3	5
failure of a traffic light	2	3
density' s road	1	2
number's vehicles	1	1

The same principle was followed to define the period's validity corresponding to the mode of the data.

B. Data Sharing

Each vehicle cooperates in the network by sending its collected data to other vehicles. To avoid redundancy, the information is shared only if it meets the following conditions:

- The vehicle has recently collected the data.
- The vehicle can send the same data again only if the previous message regarding it has reached its time validity and the information is always valid.
- None of the vehicle's neighbors already distributed this information (i.e. a message from a neighbor containing the information to share hadn't been received earlier).

If the conditions are met, the vehicle waits for a random time (backoff), bounded by the importance of the information. This mechanism aims at avoiding simultaneous sending leading to redundancy and resource wastage.

Our concept of *class* and *mode* allows broadcasting harvested information by the vehicles according to their coordinates and date of collection. Each message will contain both values in its header.

To share information among vehicles in VSNs a basic and straightforward approach is to use classical broadcast (i.e. flooding). This will ensure the reception of the message by all members but inevitably causes what we call a broadcast storm [10]. We decide to proceed differently. First, we choose a single hop broadcast; the message is received by all the direct neighbors of the source. Then, the continuation of the rebroadcasting procedure depends on the class and mode associated to the message i.e. the data.

The source vehicle is also in charge of the election of nodes among its neighbors to rebroadcast the message. The list of elected vehicles is thus inserted in the message, so that only these vehicles are authorized to rebroadcast the message. This allows to avoid broadcasting redundancy at the reception.

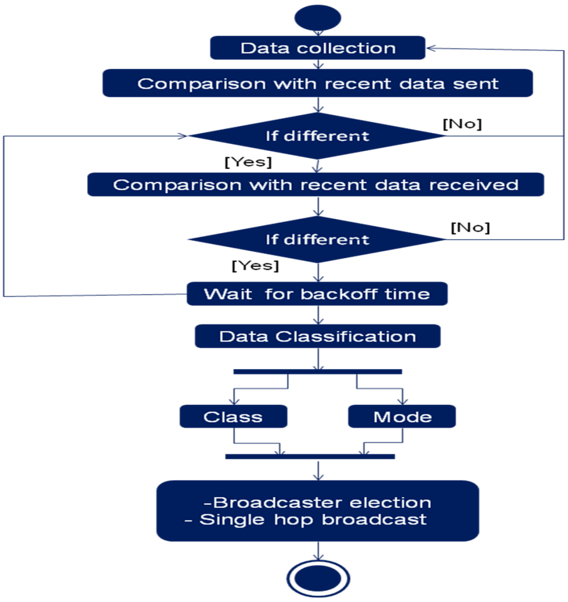

Figure 1. Sending diagram

The number of broadcaster nodes to elect depends on the information class. For the example depicted in “Table. 1”, we can have for instance three configurations. One is for class 1 where a simple single hop broadcast is sufficient while respecting the 200*200 m² perimeter. The second, for classes 2 and 3, where the source selects three broadcaster nodes to cover the larger area covered (300*300 m² and 400*400m², respectively). This can be implemented by choosing the node with the largest number of neighbors, and then we choose two other nodes with a wide coverage located at almost 120 ° and -120 ° from the current source node. Similarly, the same method can be applied to classes 4 and 5 where four broadcaster nodes and an angle of 90° are used.

“Fig. 2” represents an example of using the election algorithm for class 3 of depicted application example, in which the number of broadcaster is fixed at three. As we can see below, the first elected node is the neighbor with the highest density. Then, the two other elected nodes are chosen according to their rotation angle relative to the previous elect and their density.

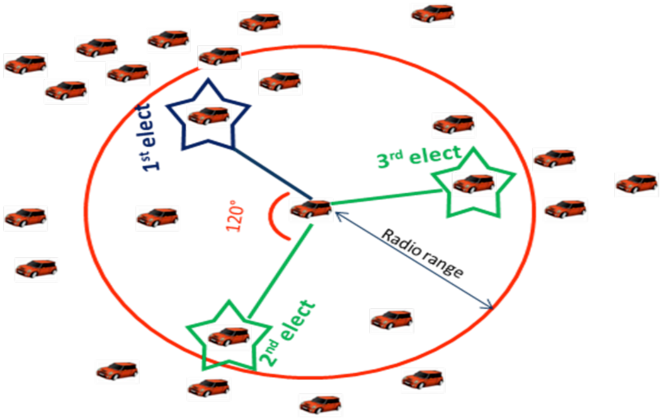

Figure 2. Example of election process for three broadcasters to elect

The election algorithm has, as an input, the number of broadcaster nodes we would like to have, and as an output, their coordinates. It is depicted in “Fig. 3”.

C. *Iterative Rebroadcasting*

At the reception of a message, the vehicle checks the validity of the information in time and space. This is performed according to the *mode* and *class* associated to the data in the message. If the information crossed the frontiers of its targeted area or the time validity has passed, the receiving node deletes it.

Otherwise, the receiving node consumes the information and verifies if it is elected as a broadcaster for it or not. If so, it broadcasts the message while safeguarding its characteristics, such as: class, mode, collection date and collection location coordinates.

```
Input: elect_number;
□=360° / elect_number;
I=0; Y=1;
Elect_table[0]=Look_for_the_node_with_highest_density
( the_list_of_neighbors_sender);

Delineate_areas_with_angle_according_to(□,coordinates_
                    of_Elect_table[0]);

While ( Y < elect_number ){
Elect_table[Y]=Look_for_the_node_with_highest_densit
y ( the_list_of_area(y));
        Y++;                    }
 Output: Elect_table;
```

Figure 3. Election algorithm

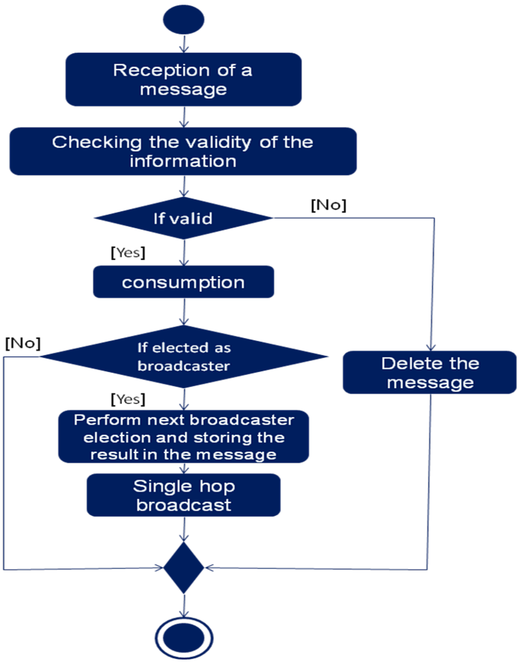

Figure 4. Message consumption and rebroadcasting

IV. PERFORMANCE EVALUATION

In order to evaluate the performance of our protocol, ADCD, and compare it to other existing protocols, we used NS2 [9]. The parameter values for simulation are given in “Table. 2”.

Using this evaluation scenario, ADCD is compared to two other diffusion algorithms. The first one is the classical broadcast where each receiving node performs one rebroadcast for message. The second one is an adapted diffusion of MobEyes, where a node transmits a message each 12 seconds containing a maximum of 5 received or harvested summaries. To increase the effectiveness of its sharing information, we set the k-hop to 3. We consider a message as redundant if the 5 summaries contained in it had already been received by the node, or if they exceed their interest perimeter or time validity.

For clarity, we measure the performance on a simple application, with different messages depicted in “Table. 1”; their *class* and *mode* belong to the interval [1, 5].

In our simulation, a new message (i.e. harvested data) is generated each 5 seconds and can be harvested (i.e. sensed) by at least one node. The events happen with certain probabilities according to their class, the probabilities of harvesting are 0.60, 0.20, 0.10, 0.05, 0.05 to the first, second, third, fourth and fifth class, respectively.

We select two performance metrics for our evaluation:

- The percentage of useful reception according to the target diffusion area.
- The overhead induced by each scheme.
- The sensitivity of these two metrics regarding the vehicular density is also analyzed.

A. *Effectiveness of the target diffusion*

A message loses its relevance after a period of time which differs according to its content. In “Fig. 5”, we represent the percentage of received messages by the concerned nodes as a function of time.

TABLE II. PARAMETERS VALUES FOR THE SIMULATION

Number of nodes	100, 200, 300	Mobility Model	SUMO [8]
Mac layers protocols	IEEE 802.11	traffic environment	urban
Bandwidth	11 Mbps	Area size	9000*9000m²
Propagation model	Two ray ground reflexion	Maximum speed	13.9 m/s
Transmission range	250 m	Maximum simulation time	300 seconds

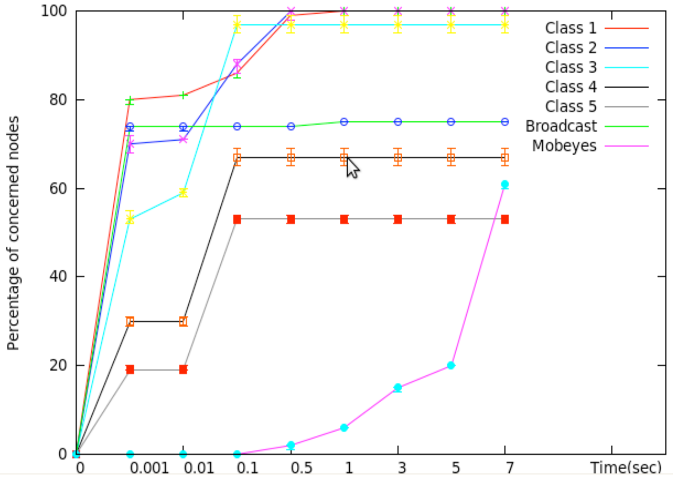

Figure 5. Reception rate according to the thresholds

We notice that ADCD quickly informs all concerned nodes compared to adapted MobEyes. In terms of promptness, ADCD gives better results than both concurrent protocols. In addition, it reaches 100% of concerned nodes in the cases of class 1, 2 and 3 in less than 0.5 s. However, in the cases of class 4 and 5 the concerned nodes receive more than 65% and 50% respectively, of messages after 0.1s. We also note that extending the broadcasting time doesn’t allow to improve this percentage. The explanation of this difference among the different classes is mainly due to the size of the targeted area. The latter is more important in the case of classes 4 and 5. A classical broadcast reaches a 75% reception ratio on average (i.e. no differentiation between classes), which corresponds to 77%, 76%, 75%, 52% and 40% for the first, second, third, fourth and fifth class, respectively. MobEyes, however, requires more time to distribute its messages due to its design features, which also leads to a low percentage of informed vehicles which corresponds to 61% (58%, 74%, 61%, 60%, and 35%). For reliability, we used confidence intervals estimated using standard deviation.

B. *Overhead*

Each message transmitted to a node that is not relevant to or after the expiry of its validity period is uninteresting and considered as overhead. “Fig. 6”, shows the wide gap between the overhead generated by the three algorithms.

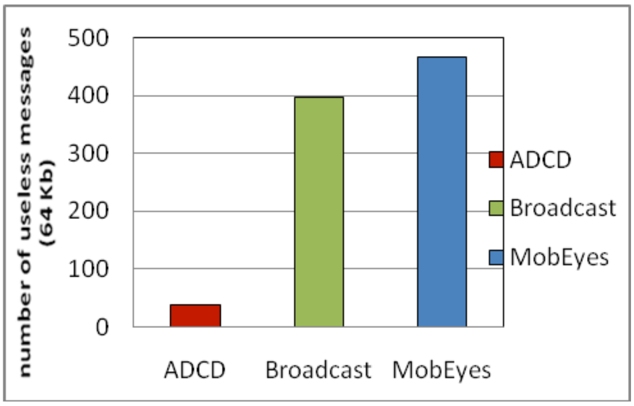

Figure 6. Overhead generated by 300 nodes

We note that ADCD provides less overhead compared to the broadcast and adapted MobEyes protocols. These results are mainly due to the target diffusion using the concept of class and period of validity according to the type of information. Moreover, ADCD allows for deleting useless messages where the validity is expired and for preventing their rebroadcast. In addition, the selection of the broadcaster nodes based on the election process allows to significantly reduce the redundancy.

C. *Vehicle density impact*

In order to evaluate the impact of the vehicle nodes density on the ADCD performance, we varied the number of nodes from 100 to 300. The obtained results are plotted in "Fig. 7", which illustrates the average percentage of concerned nodes' received messages according to the nodes' density. We remark that, even if the nodes density increases the percentage of concerned nodes remains stable greater than 90% in the case of ADCD. However, in the case of both protocols (Broadcast and adapted MobEyes) with 100 nodes, the percentage of concerned nodes cannot reach 50% and 40% in the case of Broadcast and MobEyes respectively. When the nodes' density increases the Broadcast and MobEyes' concerned nodes cannot exceed 75% and 60% respectively. These results confirm the efficiency of ADCD for the target diffusion even with a small density which can limit the possibilities of rebroadcast in a large area.

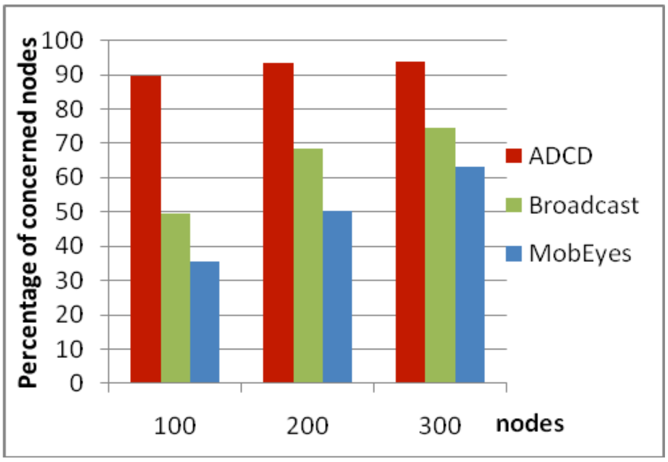

Figure 7. Percentage of reception with 100, 200 and 300 nodes

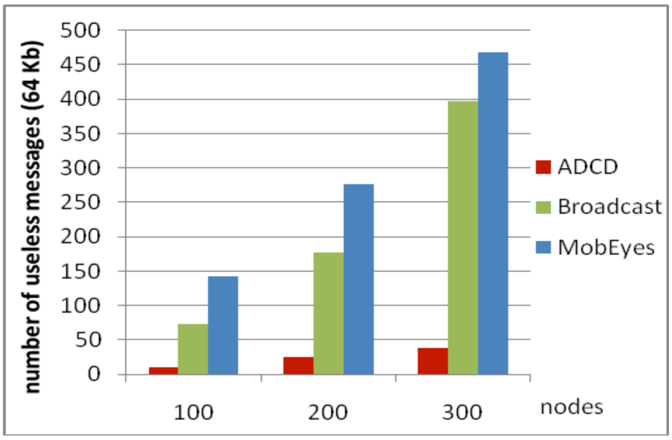

Figure 8. The overhead generated by 100, 200 and 300 nodes

"Fig. 8" shows the overhead according to the node density. We notice the linear increase of the ADCD overhead compared to the exponential of Broadcast and MobEyes. A large overhead causes the congestion of a network leading to the impossibility of injecting new messages. This can have very serious consequences on safety application for the VSNs. In our case, it was impossible to push simulations with larger density because of the overhead.

V. CONCLUSION AND FUTURE WORK

The main objective of our study was to develop a new protocol to ensure a tradeoff between the reception gain of data and the latency since it is harvested. This needs to be performed while avoiding redundancy and the implied network congestions. Our protocol called ADCD, for Advanced Diffusion of Classified Data, achieves this goal by characterizing the data to be diffused (distributed area range, validity period) and using an adapted broadcasting strategy. Our performance evaluation showed that ADCD allow for reducing the overhead by up to 90% compared to other existing protocols. This is performed while keeping the stability and reactivity of ADCD at an acceptable level for VSN applications. Our future work includes the adaptation of ADCD to different traffic environments (sub-urban, rural, highways …). Taking into account security concerns, such as confidentiality and message integrity, is another important field of investigation.

REFERENCES

[1] A. Benslimane,"Optimized dissemination of alarm messages in Vehicular Ad-Hoc Networks (VANET)," High Speed Networks and Multimedia Communication, 2004, Volume 3079/2004, pp. 655-666, HSNMC'04, Toulouse, France, July 2004.

[2] M. TorrentMoreno, D. Jiang and H. Hartenstein, "Broadcast reception rates and effects of priority access in 802.11Based Vehicular AdHoc Networks," In ACM International Workshop on Vehicular Ad hoc Networks, VANET '04, Philadelphia, PA,USA, October 2004.

[3] F. Bai, B. Krishnamachari, "Exploiting the wisdom of the crowd: localized, distributed information-centric VANETs," IEEE Communications Magazine, Volume 48, Issue 5, IEEE Press Piscataway, NJ, USA May 2010.

[4] U. Lee, E. Magistretti, B. Zhou, M. Gerla, P. Bellavista and A. Corradi, " MobEyes: smart mobs for urban monitoring with a Vehicular Sensor Network," IEEE Wireless Communications, Volume 13, Issue 5, pp. 52 – 57, Novembre 2006.

[5] U. Lee, E. Magistretti, M. Gerla, P. Bellavista, P. Li´o and K. Lee, "Bio-inspired multi-agent data harvesting in a proactive urban monitoring environment," Ad Hoc Networks Journal, Special Issue on Bio-Inspired Computing and Communication in Wireless Ad Hoc and Sensor Network, Volume 7, Issue 4, pp. 725-741, September. 2009

[6] U. Lee, E. Magistretti, B. Zhou, M. Gerla, P. Bellavista and A. Corradi," Efficient data harvesting in mobile sensor platforms," IEEE International Conference on Pervasive Computing and Communications, PERCOM'06, Pisa - Italy, March 2006.

[7] M. Gerla and L. Kleinrock, "Vehicular networks and the future of the mobile internet," Computer Networks, Volume 55, Issue 2, Pages 457-469, February 2011.

[8] SUMO, tp://sourceforge.net/apps/mediawiki/sumo/

[9] NS-2, http://www.isi.edu/nsnam/ns/

[10] S. Ni, Y. Tseng, Y. Chen and J. Sheu," The broadcast storm problem in a mobile ad hoc network," MobiCom 99.