

Modeling and Performance Evaluation of Advanced Diffusion with Classified Data in Vehicular Sensor Networks

Nadia Haddadou, Abderrezak Rachedi, Yacine Ghamri-Doudane

► **To cite this version:**

Nadia Haddadou, Abderrezak Rachedi, Yacine Ghamri-Doudane. Modeling and Performance Evaluation of Advanced Diffusion with Classified Data in Vehicular Sensor Networks. *Wireless Communications and Mobile Computing*, Hindawi Publishing Corporation, 2011, 11 (12), pp.1689-1701. 10.1002/wcm.1220 . hal-00680861

HAL Id: hal-00680861

<https://hal-upec-upem.archives-ouvertes.fr/hal-00680861>

Submitted on 7 Jun 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modeling and Performance Evaluation of Advanced Diffusion with Classified Data in Vehicular Sensor Networks

Nadia Haddadou*, Abderrezak Rachedi*, and Yacine Ghamri-Doudane*,†

*Université Paris-Est - Gaspard Monge Computer Science Laboratory (LIGM - UMR 8049)
75420 Champs sur Marne, France

† ENSIIE, 1 Square de la résistance, 91025 Evry Cedex, France

Abstract—In this paper, we propose a newly distributed protocol called Advanced Diffusion of Classified Data (ADCD) to manage information harvesting and distribution in vehicular sensor networks. ADCD aims at reducing the generated overhead, avoiding network congestions as well as long latency to deliver the harvested information. The concept of ADCD is based on the characterization of sensed information (i.e., based on 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 in which they have no interest. Also, we propose in this paper a new probabilistic model for ADCD based on Markov chain. This one aims to optimally tune the parameters of ADCD, such as the optimal number of broadcaster nodes. The analytical and simulation results based on different metrics, such as the overhead, the delivery ratio, the probability of a complete transmission, and the minimal number of hops, are presented. These results illustrate that ADCD allows 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 with the classical broadcast and an adapted version of MobEyes. The ADCD overhead is kept stable whatever the vehicular density.

Index Terms—Vehicular Sensor Network; Data characterization; election of broadcasters; Data Dissemination; Markovian model; Performance Evaluation

I. INTRODUCTION

Vehicular sensor networks (VSNs) 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, etc.) than basic wireless sensor networks. However, the large amount of generated information and the frequent topology changes (density, topology, neighboring nodes, etc.) make the existing solutions in the wireless sensor network field ineffective for this new kind of networks.

The main issues in VSNs reside in 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, that is, the nodes that should receive the information and can be interested in its content according to its geo-location, and on the harvested data specificities, that is, 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 called 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 is 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 and 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. 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. Last but not least, on the basis of our previous work [1], we propose a new analytical model based on Markov chains for ADCD. This model results the delivery ratio, the overhead, the probability of a complete transmission, and the minimal numbers of hops; and allows us to optimally select the ADCD parameters. The performance evaluation has been carried out, and the analytical and the simulation results are presented, showing the important improvements achieved by ADCD compared with other protocols from the literature.

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 ADCD protocol. Then, Section IV relates a theoretical part, studying the effects of variations in the number of broadcasters for

ADCD election process. The numerical analysis is given in Section V, followed by the performance evaluation of ADCD and its comparison with two other existing protocols, that is, a classical broadcast and an adapted version of the MobEyes [2] protocols. Finally, Section VII draws our conclusion and suggests some future research directions.

II. RELATED WORK

Many studies have recently been made in the field of VSNs. In [3] and [4], the authors presented a survey of architectures and communication protocols in the vehicular networks. The first study presents different architectures, proposed to efficiently connect the vehicles between them and to form the network. Thereafter, it lists some routing strategies and, at the end, links between their performance results and some factors such as the wireless access methods, the vehicle mobility, and the localization. The second study focuses more on the communication protocols; it details some broadcasting algorithms such as the diffusion, the probabilistic flooding, the parameterless reliable broadcasting [5], followed by some routing strategies such as epidemic routing and delay bounded routing [6]. Among the main topics addressed by the research community in this domain, we are interested in both the following themes: data harvesting and data diffusion. Later, we explain the limitation of the related literature relative to these two topics. These limitations had motivated our current work.

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: data search and data dissemination.

The main data harvesting approach proposed in the literature is the DataTaxis [7] algorithm. This algorithm is inspired by the behavior of insects during their search for food. In DataTaxis, this behavior is thus adapted to vehicles, allowing to cover a large area and to harvest the maximum amount of data effectively. This is carried out 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, because of 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 (QoS) by this application. Among the existing works, two methods are frequently used. The first one [8], [9], 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. Another kind of broadcasting is proposed in [5] on the basis of using acknowledged and connected dominating sets to fix a set of nodes regarded by the data according to their location beside the source; the performance results are good for the highway scenario but remain insufficient for the urban scenario. The second method performs first data processing, where vehicles only exchange summaries at regular time [10]. 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 [2]. We position ourselves in the middle of these two ideas; the preprocessing made in ADCD avoids sending messages repeatedly and limits the rebroadcasting of each individual data on the basis of its degree of importance. By doing so, we can ensure good latency and reception ratio.

For a dissemination that takes into account the degree of importance of each individual data to be performed, a classification between different kinds of messages, according to their QoS requirements in terms of delay and throughput, is established. The first class concerns emergency messages, such as accident notifications. These messages are short and need to be transmitted with a high propagation velocity to insure a real-time service [11], [12]. 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 reducing the risk of traffic jams. Warning messages are less important than emergency ones, but a higher rate is required for them. Although such classification of messages already exist, the question of how to introduce this classification in the dissemination phase remains vague in the literature. 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 carried out 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 [13]. The technology used for this is peer-to-peer technology, which remains innovative for this kind of networks. The content to be shared includes both, 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 [14] differs from previous ones that use classical criteria for data classification. The authors developed 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, citywide alerts, and interactive services with their spatial, temporal, and interest scope that 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 that 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 that takes full advantage of data classification.

With the same approach, the proposed protocol delay-bounded vehicular data gathering [6] aims at spreading information corresponding to a certain geographic area, and satisfying a specific delay bound, it uses fixed base stations that create the information and determine a target area for the data propagation during a fixed time interval. According to a delivery probability related to a distance that delay-bounded vehicular data gathering calculates, it chooses between two options: immediately forwarding the message to another node if it has a great probability of successfully relaying this message or waiting a moment and probably aggregating data from different sources; this second choice depends on the delay bound of the data and allows reducing the bandwidth consumption and waiting for a more adapted broadcaster. The concept of linking between space and time criteria for the data diffusion is the same as ADCD. However, more metrics are studied in this paper about this idea.

Another very interesting approach for data diffusion in VSNs is the smart mobs for urban monitoring with a VSN, MobEyes [2], whose performance analytical model is presented in [15] with more detailed simulation results. MobEyes ensures proactive urban monitoring by taking advantage of vehicle mobility. The exchanged messages can include 2-10 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 avoid redundancy. However, MobEyes is proposed for a specific delay-tolerant application, and its generalization is not straightforward. That is why we propose a new protocol, ADCD, with more parameters that make it more general; 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 earlier, existing solutions cannot always meet these requirements. 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 overhead.

The proposed ADCD concepts fit the identified constraints thanks to the data characterization and the corresponding enhanced diffusion.

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

III. ADVANCED DIFFUSION OF CLASSIFIED DATA

In this section, we describe our proposed protocol called ADCD. Its aim is to empower safety applications with VSNs. 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 communication means, accepting to collaborate with other vehicles for safety information distribution. This protocol allows for regular harvesting along with smart information sharing among 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 carried out 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 not only to do that but also to distribute information on the entire network if needed.

Advanced Diffusion of Classified Data is based on three main parts: (i) harvesting and data classification, (ii) broadcasters election 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, and 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 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.

Advanced Diffusion Of Classified Data defines an interval $[\sigma_{min}, \sigma_{max}]$ for the classes and modes. The most urgent messages and the longest in validity are characterized as class σ_{max} . A piece of information is represented by C_{xy} , where x represents the class and y the mode.

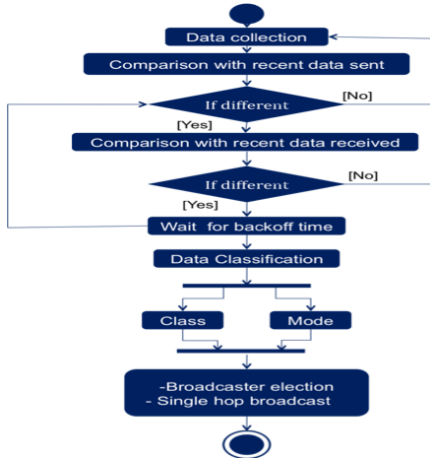


Fig. 1. Sending diagram.

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 I 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 ($200 \times 200m^2$, $300 \times 300m^2$, $400 \times 400m^2$, $600 \times 600m^2$ and $800 \times 800m^2$ to the first, second, third, fourth, and fifth class, respectively). Hence, in this example, a road density message should not exceed the perimeter of $200 \times 200m^2$.

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

B. Data Sharing

Fig.1 illustrates the process executed by each vehicle before transmitting the collected data. Each vehicle cooperates in the network by sending its collected data to other vehicles. For redundancy to be avoided, 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; and
- none of the vehicle's neighbors already distributed this information (i.e., a message from a neighbor containing the information to share had not 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.

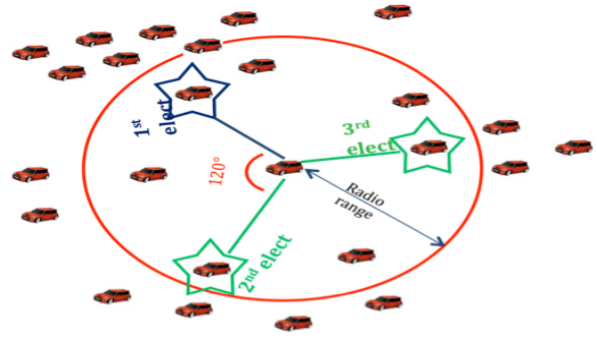


Fig. 2. Example of election process for three broadcasters to be elected.

For information among vehicles in VSNs to be shared, 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 [16]. 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 with the message, that is, 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 avoiding broadcasting redundancy at the reception.

The number of broadcaster nodes to be elected depends on the information class. For the example depicted in Table I, we can have, for instance, three configurations. One is for class 1, where a simple single-hop broadcast is sufficient while respecting the $200 \times 200m^2$ perimeter. The second, for classes 2 and 3, where the source selects three broadcaster nodes to cover the larger area covered ($300 \times 300m^2$ and $400 \times 400m^2$, 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 in Fig.2, 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.

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 Algorithm.1.

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 with the data in the message. If the information crossed the frontiers of its targeted

Algorithm 1 Election algorithm.

Input: $elect_number$, $\theta = \frac{360^\circ}{elect_number}$, $Y = 1$.
Output: $elect_table$.

$elect_table[0]$ = ID of the neighbor with the highest density (list of potential sender nodes);
 List of areas = areas delimited according to $(\theta, \text{coordinates of the node in } elect_table[0])$;

while ($Y < elect_number$) **do**
 $elect_table[Y]$ = ID of the neighbor with the highest density in area (area[Y]);
 $Y = Y + 1$;

end while

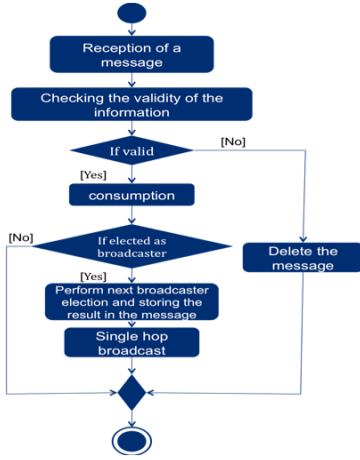


Fig. 3. Message consumption and rebroadcasting.

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 as illustrated in Fig.3.

IV. MARKOV CHAIN MODEL FOR ADVANCED DIFFUSION OF CLASSIFIED DATA

This section presents a modeling of ADCD framework, particularly the broadcasting process. The message is sent by a source node and retransmitted by a fixed number of broadcaster nodes. We use the Markov chain theory with a binomial distribution. We discuss each part of this modeling in the following.

A. Network model

We consider a network made up of N mobile nodes located in a closed area. An event occurs in this area, and a source node is responsible for its dissemination; we suppose that all the nodes are affected by its reception because of the smallness of the area. The limitations of this dissemination are due to the transmission range that determines if two nodes are neighbors or not and the mobility that makes the links appear and disappear frequently. We model this by a Markovian model, where the states are represented by the number of nodes that received the message and the edges are represented by the

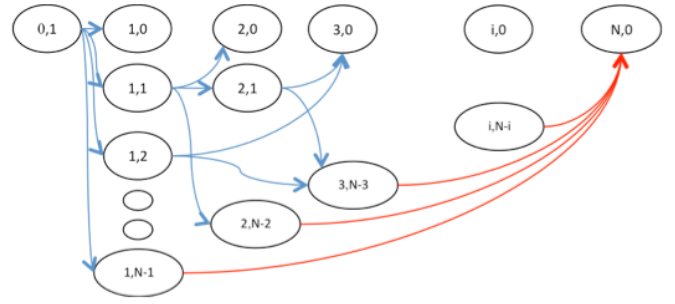


Fig. 4. Markov chain for Advanced Diffusion of Classified Data.

probabilities regarding the transition from one state to another, in other words, the variation in the number of nodes having received the message.

In the ADCD protocol, the nodes retransmit a message only once at most. To model this specificity, we compose a state in our Markov chain by (i, j) , such as $i \in \{0, 1, 2, \dots, N\}$ and $j \in \{0, 1, 2, \dots, N - i\}$. The first component i of a state represents the number of nodes that have already received the message and therefore have transmitted it once at most. The second component j is the number of nodes that have recently received the message among the remaining nodes $N - i$ and then able to retransmit it in the next step.

Possible state's transactions are from (i, j) to $(i + j, m)$ such as $m \in \{0, 1, 2, \dots, N - i - j\}$ and represent the future nodes affected by the reception. To determine the probabilities of this model, we assign a probability P to the existence of a link between two nodes that makes them neighbors and Q to its nonexistence or disappearance. This connectivity graph can be viewed as a Markov chain whose stationary probability for existing link is calculated as follows: $\pi = \frac{P}{P+Q}$. This value represents the probability of reception for a node during one step [17] and allows calculating the average node degree by $\pi \times (N - 1)$.

The Markov chain's modeling of our routing algorithm has $1 + \frac{N \times (N+1)}{2}$ states. The initial state $(0, 1)$ represents the case where the source is the only holder of the message, the final state $(N, 0)$ represents the case where all the nodes have received the message, and there are $\frac{(N-1) \times (N+2)}{2}$ remaining states representing all the possible transition states (i, j) . The Markov chain for ADCD is illustrated in Fig.4.

B. Transition probabilities

To calculate the probability of a transition from (i, j) state to $(i + j, m)$ state as described in equation (1), we consider this process as a binomial one where we have j attempts to transmit the message to the m nodes. The result can be a success or not between each couple of nodes. This binomial distribution has n and p as parameters, so that $m \sim B(n, p)$, where n represents the number of nodes that have not yet received the message, equivalent to $N - i - j$ in our model, and p is the probability of receiving a message sent by a number D of nodes during the same step, equivalent to $1 - (1 - \pi)^D$ in our case, according to the primitive given in [17].

TABLE II
MARKOV CHAIN'S MATRIX OF TRANSITION PROBABILITIES.

States	(0,1)	(1,0)	(1,1)	(1,2)	(2,0)	(2,1)	(3,0)
(0,1)	0	$(1-\pi)^2$	$2\pi(1-\pi)$	π^2	0	0	0
(1,0)	0	1	0	0	0	0	0
(1,1)	0	0	0	0	$1-\pi$	π	0
(1,2)	0	0	0	0	0	0	1
(2,0)	0	0	0	0	1	0	0
(2,1)	0	0	0	0	0	0	1
(3,0)	0	0	0	0	0	0	1

$$P[(i, j)(i + j, m)] = P_{\text{Binomial}}(m, 1 - (1 - \pi)^D, N - i - j) = \left(\frac{(N-i-j)!}{m! \times (N-i-j-m)!} \right) \times (1 - (1 - \pi)^D)^m \times (1 - \pi)^{D(N-i-j-m)} \quad (1)$$

such as

$$D = \begin{cases} \text{fixed number of electedvbroadcasters} & \text{if } D > j \\ j & \text{Otherwise,} \end{cases}$$

Advanced Diffusion of Classified Data uses an adaptive broadcast algorithm based on the election of a fixed number of broadcasters to reduce the overhead without neglecting the reception ratio. In order to study the performance of such a strategy, we adapt it to a Markovian model by setting D as the number of broadcasters chosen for a given application. Table II gives an example for a Markov chain corresponding to a network with three mobile nodes and a number of broadcasters fixed to 1 (because of the restricted number of nodes). In this case, the transition probabilities matrix includes seven states.

V. NUMERICAL ANALYSIS

In order to adapt the number of selected broadcasters to an application or, as in our case, chosen regarding the class and mode of information, we made a study of the evolution of the delivery ratio versus time, the probability of a complete transmission for all the concerned nodes, and finally the overhead induced. Each criterion is detailed in the following.

A. Numerical results

1) *Delivery ratio*: The delivery ratio is a general criterion for all applications, and each one has a set of minimum values corresponding to the importance for the vehicles to be aware of the exchanged information. Generally, the vehicles concerned by an event are limited to an area, and this makes the number of nodes and the extent of the message transmission proportional to the vehicles' density; this is why we restrict the number of nodes N to 20 during our calculus.

Fig 5 illustrates the number of informed nodes according to different π parameters and network connectivity.

Without taking into consideration the time parameter, the results essentially depend on the network connectivity. However, this allows giving a general idea of the ratio, and it is a sufficient criterion for the delay-tolerant applications.

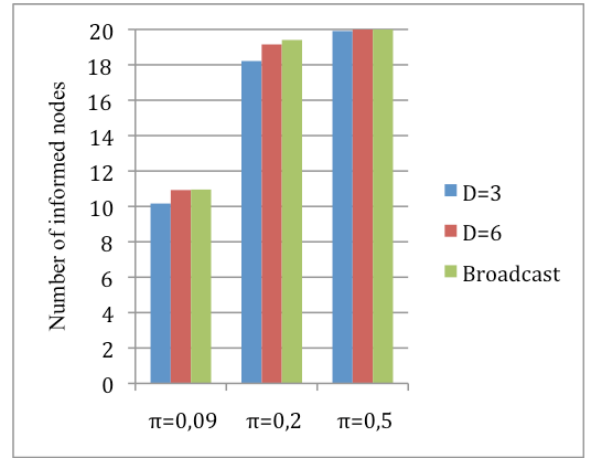


Fig. 5. Delivery ratio for the different diffusion strategies.

TABLE III
MOST PROBABLE NUMBER OF HOPS FOR A GLOBAL RECEPTION.

Minimum number of hops	Number of broadcasters		
	3	6	All the neighbors
$\pi=0.09$	6	5	4
$\pi=0.2$	4	3	3
$\pi=0.5$	3	2	2

2) *Transmission speed*: This metric is calculated using Dijkstra's algorithm as a reference in order to determine through the matrix of transitions the number of hops between the initial state and the final one with the highest probability, which also determines the most probable way of transmission.

The results are presented in Table III for $N=20$. We remark that there is no significant difference between the use of three and six broadcasters.

3) *Probability of a complete transmission*: This probability represents the case where all $(N - 1)$ nodes have received the message, that is, the probability of being in the state $(N, 0)$. This metric is significant for safety applications whose content is extremely important and its transmission has to be completely achieved. In order to study the variation versus the time, we calculate this probability after each τ time step so that one step matches one hop as follows in equation (2).

$$P_{\text{globaltransmission}}(\tau) = \text{initial vector} \times \text{matrix of transaction} \times \text{final vector} \quad (2)$$

In our case, the source node is the only holder of the message at the beginning, and this is thus represented by $\text{initialvector} = [100\dots0]$; the end of our model is when we reach the final state $(N, 0)$ represented by a $\text{finalvector} = [000\dots01]^{-1}$.

The results for this metric are presented in Fig.6 and Fig.7 for $N = 20$. Again, to vary the network connectivity, we first consider a number of elected broadcasters D equal to 3, then D equal to 6, and finally consider that all the neighbors are broadcasters (i.e., full broadcast). Fig.6 illustrates the obtained results for $\pi = 0.2$, whereas Fig.7 is for $\pi = 0.5$.

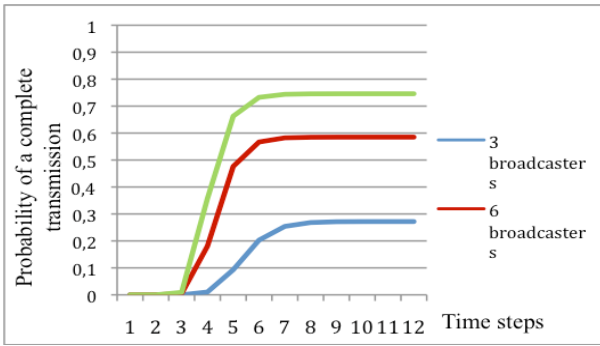


Fig. 6. Probability of a complete transmission with $\pi = 0.2$.

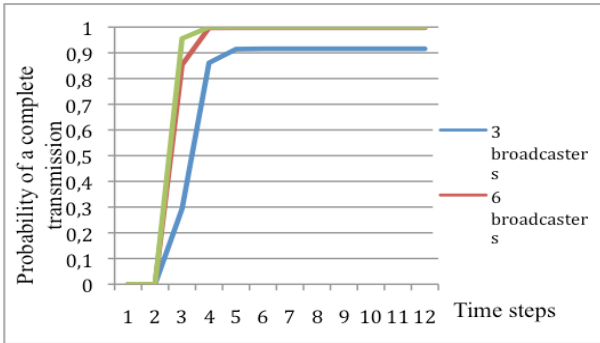


Fig. 7. Probability of a complete transmission with $\pi = 0.5$.

The connectivity degree of the network is the parameter that has the strongest impact on the results, but we can nevertheless notice that the number of broadcasters also affects these probabilities and the transmission's speed. For a weak connectivity, the probability is very low, what shrinks the deployment of many applications; in this case, we advocate the use of a broadcast to maximize the performance. For an average connectivity, the probability is higher than 0.5 with the use of six broadcasters or more; the choice of the number of broadcasters will only depend on the relevance of the transmitted information. The last case is when the connectivity is strong; the results for the three diffusion strategies are similar, and then the chosen broadcasters' number should be minimal and without any damage for the delivery's ratio and speed.

4) *Overhead*: This parameter is generally tied to the delivery ratio. To minimize it so much can severely decrease the ratio; this is why we have to find a compromise. In this study, we consider that any repeated reception at the same node is an overhead. This can occur even with a few broadcasters.

We calculate the number of repeated receptions with the following equation (3), so that this value depends on the probability of double reception for $(i+j-1)$ already informed nodes. The transmission is made by the D nodes selected as broadcasters, taking into account the probability of transition from state (i, j) to state $(i+j, m)$. To obtain the number of duplicate copies for all the possible transitions, we multiply the result of this probability by the number of nodes that are likely

TABLE IV
INDUCED OVERHEAD (NUMBER OF REPEATED RECEPTION).

Overhead	Number of broadcasters		
	3	6	All the neighbors
$\pi=0.09$	19.16	24.41	26.72
$\pi=0.2$	31.43	57.78	95.85
$\pi=0.5$	39.58	64	109.86

to have an overhead, which means those that have already received a copy of the message.

$$\begin{aligned}
 \text{Overhead} = & \sum_{y=1}^{i+j-1} P_{\text{Binomial}}(y, \pi, i-j-1) \\
 & \times D \times (i+j) \\
 & \times P_{\text{Binomial}}(m, 1-(1-\pi)D, N-i-j) \quad (3)
 \end{aligned}$$

There are several ways between the first state $(1, 0)$ and the last state $(N, 0)$ to obtain an average for the overhead; we compute a proportionate average regarding the probability or realization of each way. Table IV presents these results.

5) *Overview*: Fig.8 joins the three studied metrics and gives them a broader view about the specificities of each diffusion strategy. In our application example, we chose to shrink the number of broadcasters to classes 1, 2, and 3, because of their restricted area of diffusion and important frequency of events. This allows reducing the overhead. However, for the diffusion strategy of messages of classes 4 and 5, the area is much greater and the reception delay more critical. Thus, the number of broadcasters should be greater to guarantee a certain level of performances. That is why the delivery ratio has been chosen instead of the overhead reduction. According to the obtained results, the optimal number of broadcaster nodes under such conditions must be between 3 and 4.

B. Simulation results

In order to validate our theoretical model, we reuse the same parameters and calculate the delivery ratio and the generated overhead. We used Network Simulator 2 (NS2) [18] for our simulation and Simulation of Urban Mobility (SUMO) [19] to generate the vehicular mobility patterns. In order to simulate the connectivity changes, we varied the size of the considered area until obtaining the corresponding average node degree of $\pi \times (N-1)$ for each π used in the theoretical study. The number of nodes N is set to 20. For a weak connectivity, the area used is $1000 \times 1000m^2$ large; for an average connectivity, it is equivalent to $800 \times 800m^2$; and for a high connectivity, a $500 \times 500m^2$ area is used.

Fig.9 and Fig.10 confirm our previous results. The first one presents the results regarding the delivery ratio; the use of six broadcasters and the full broadcast produce a good delivery ratio. To confirm these choices, Fig.10 presents the overhead induced for each strategy, and we notice an important difference. We can conclude that the use of a maximum number of broadcasters is not always the best strategy. We thus argue that the optimal policy to broadcast its messages can be chosen by analyzing the application's needs and more precisely its sensitivity to the delivery ratio.

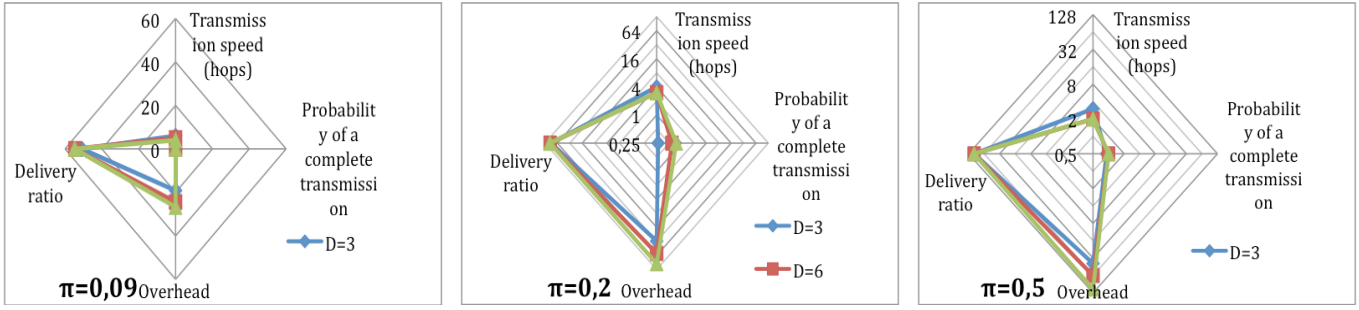


Fig. 8. Numerical results for the application's criteria.

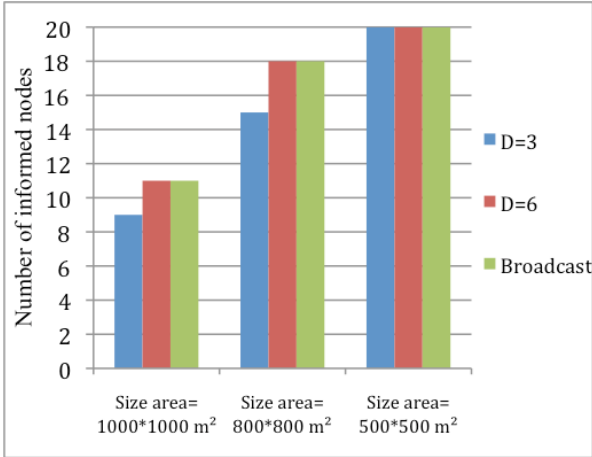


Fig. 9. Simulation results for the delivery ratio.

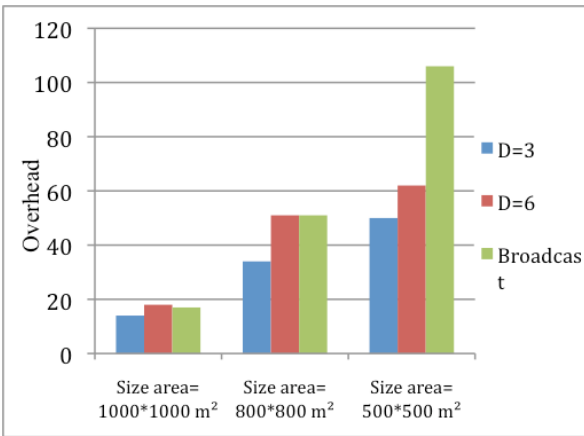


Fig. 10. Simulation results for the overhead.

VI. PERFORMANCE EVALUATION

In order to evaluate the performance of our protocol, ADCD, and compare it with other existing protocols, we used Network Simulator 2 [18]. The used simulation parameters and their values are given in Table V.

Advanced Diffusion of Classified Data is compared with two other diffusion algorithms. The first one is the classical broadcast where each receiving node performs one rebroadcast

TABLE V
PARAMETER VALUES FOR THE SIMULATION.

Number of nodes	100, 200, 300
MAC layers protocols	IEEE 802.11
Bandwidth	11Mbps
Propagation model	Two-ray ground reflection
Transmission range	250m
Mobility model	SUMO [19]
Traffic environment	Urban
Area size	9000 × 9000m ²
Maximum speed	13.9m/s
Maximum simulation time	300s

per received message. The second one is an adapted version of MobEyes, where a node transmits a message each 12s containing a maximum of five received or harvested summaries. To increase the effectiveness of the information sharing using MobEyes, we set the k-hop to 3. We consider a message as redundant if the five 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 as depicted in Table I; their class and mode belong to the interval [1, 5].

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

We select two performance metrics for our evaluation: (i) the percentage of useful reception according to the target diffusion area and (ii) 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 time that differs according to its content. In Fig.11, we represent the percentage of received messages by the concerned nodes as a function of time. For reliability, we used confidence intervals estimated using standard deviation.

We notice that ADCD quickly informs all concerned nodes compared with the adapted version of MobEyes. In terms of

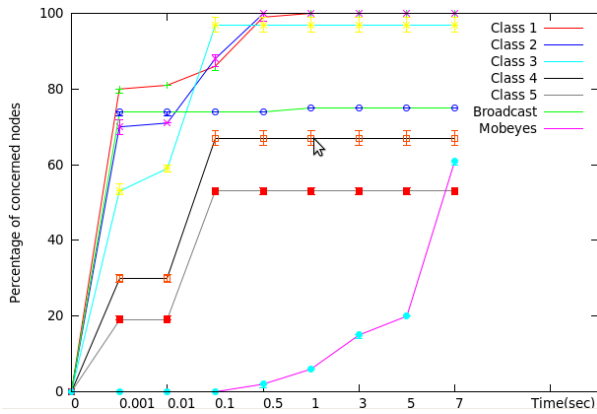


Fig. 11. Reception rate according to the thresholds.

TABLE VI
PERCENTAGES OF CONCERNED NODES.

Classes	ADCD	Broadcast	Adapted MobEyes
Class 1	100%	77%	58%
Class 2	100%	76%	74%
Class 3	100%	75%	61%
Class 4	65%	52%	60%
Class 5	50%	35%	35%

promptness, ADCD gives better results than both concurrent protocols. In addition, it reaches 100% of concerned nodes in the cases of classes 1, 2, and 3 in less than 0.5s. However, in the cases of classes 4 and 5, the concerned nodes receive, respectively, more than 65% and 50% of messages after 0.1s. We also note that extending the broadcasting time does not allow improving 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 among classes). MobEyes, however, requires more time to distribute its messages because of its design features, which also leads to a low percentage of informed vehicles corresponding more precisely to 61%. More detailed results about these percentages are given in Table VI.

B. Overhead

Each message transmitted to a node that is not relevant to or after the expiry of its validity period is uninteresting and is considered as overhead. Fig.12 shows the wide gap between the overhead generated by the three algorithms.

We note that ADCD provides less overhead compared with the full broadcast and the adapted version of MobEyes. 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 significantly reducing the redundancy.

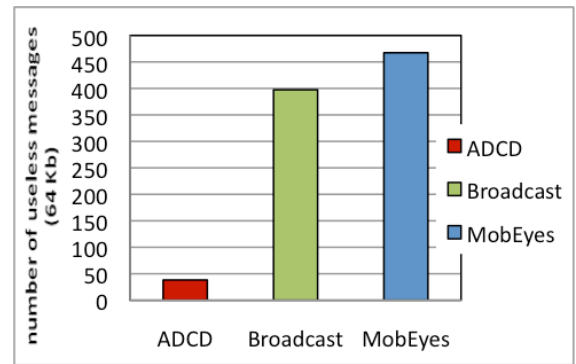


Fig. 12. Overhead generated by 300 nodes.

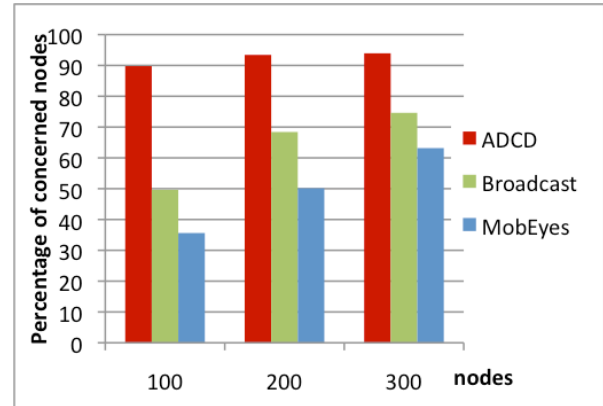


Fig. 13. Percentage of reception with 100, 200, and 300 nodes.

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.13, which illustrates the average percentage of concerned nodes receiving messages according to the nodes' density. We remark that, even if the nodes' density increases, the percentage of concerned nodes remains stable (and greater than 90%) in the case of ADCD. However, in the case of the two other evaluated protocols (Broadcast and adapted version of MobEyes) with 100 nodes, the percentage of concerned nodes cannot even reach 50% and 40% in the case of the Broadcast and MobEyes, respectively. When the nodes' density increases, the Broadcast and MobEyes concerned nodes stay lower than 75% and 60%, respectively. These results confirm the efficiency of ADCD for the target diffusion even with a small density that can limit the possibilities of rebroadcast in a large area.

Fig.14 shows the overhead according to the node density. We notice the linear increase of the ADCD overhead compared with the exponential increase experienced by the two other evaluated protocols, the Broadcast and the adapted 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 applications for the VSNs. In our case, it was even impossible to push simulations

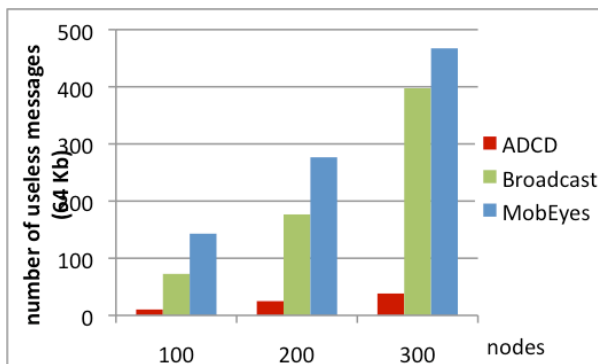


Fig. 14. The overhead generated by 100, 200, and 300 nodes.

with larger density because of the overhead induced by the Broadcast and the adapted MobEyes.

VII. CONCLUSION

The main objective of our study was to develop a new protocol to ensure a trade-off between the reception gain of data and the latency because 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 and validity period) and using an adapted broadcasting strategy validated by a theoretical study. The analytical model based on Markov chain was proposed to optimize the ADCD broadcasting parameters. The performance evaluation has been performed, and the results obtained using both analytical evaluation and simulations showed that ADCD allows reducing the overhead by up to 90% compared with 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 (suburban, rural, highways, etc.). Taking into account security concerns by using the distributed security architecture [20] in order to ensure the messages confidentiality and integrity is another important field of investigation.

REFERENCES

- [1] N. Haddadou, A. Rachedi, and Y. Ghamri-Doudane, "Advanced diffusion of classified data in vehicular sensor networks," in *Wireless Communications and Mobile Computing Conference (IWCMC'11)*, Istanbul, Turkey, Jul. 2011.
- [2] 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*, vol. 13, no. 5, pp. 52–57, Nov. 2006.
- [3] U. Lee and M. Gerla, "A survey of urban vehicular sensing platforms," *Elsevier Computer Networks Journal*, vol. 54, no. 4, pp. 527–544, Mar. 2010.
- [4] A. Casteigts, A. Nayak, and I. Stojmenovic, "Communication protocols for vehicular ad hoc networks," *Wireless Communications and Mobile Computing, Special Issue: Wireless Mesh and Other Emerging Wireless Network Technologies*, vol. 11, no. 5, pp. 567–582, May 2011.
- [5] F. Ros, M. Ruiz, and I. Stojmenovic, "Reliable and efficient broadcasting in vehicular ad hoc networks," in *IEEE Vehicular Technology Conference (VTC' 2009 Spring)*, Barcelona, Spain, Apr. 2009.

- [6] C. Palazzi, F. Pezzoni, and P. Ruiz, "Delay-bounded data gathering in urban vehicular sensor networks," *Elsevier Journal of Pervasive and Mobile Computing, Special Issue on Vehicular Sensor Networks and Mobile Sensing over Wide-Scale Deployment Environments*, vol. 8, no. 2, pp. 180–193, 2011.
- [7] U. Lee, E. Magistretti, M. Gerla, P. Bellavista, P. LiÁo, and K. Lee, "Bio-inspired multi-agent data harvesting in a proactive urban monitoring environment," *Special Issue on Bio-Inspired Computing and Communication in Wireless Ad Hoc and Sensor Network*, vol. 7, no. 4, pp. 725–741, Sep. 2009.
- [8] T. Fukuhara, T. Warabino, T. Ohseki, K. Saito, K. Sugiyama, T. Nishida, and K. Eguchi, "Broadcast methods for inter-vehicle communications system," *Proceedings of IEEE Wireless Communications and Networking Conference*, vol. 4, pp. 2252–2257, 2005.
- [9] C. F. Chiasserini, E. Fasoloz, R. Furiatoz, R. Gaetax, M. Garettoy, M. Gribaudox, M. Serenox, and A. Zanellaz, "mart broadcast of warning messages in vehicular ad hoc networks," in *Workshop Interno Progetto NEWCOM (NOE' 05)*, Turin, Italy, Nov. 2005.
- [10] U. Lee, E. Magistretti, B. Zhou, M. Gerla, P. Bellavista, and A. Corradi, "Efficient data harvesting in mobile sensor platforms," in *IEEE International Conference on Pervasive Computing and Communications (PERCOM'06)*, Pisa, Italy, Mar. 2006.
- [11] A. Benslimane, "Optimized dissemination of alarm messages in vehicular ad-hoc networks (vanet)," *High Speed Networks and Multimedia Communication*, vol. 3079, pp. 655–666, Jul. 2004.
- [12] M. T. Moreno, D. Jiang, and H. Hartenstein, "Broadcast reception rates and effects of priority access in 802.11based vehicular ad hoc networks," in *ACM International Workshop on Vehicular Ad hoc Networks (VANET' 04)*, Philadelphia, USA, Oct. 2004.
- [13] M. Gerla and L. Kleinrock, "Vehicular networks and the future of the mobile internet," *Computer Networks*, vol. 55, no. 2, pp. 457–469, Feb. 2011.
- [14] F. Bai and B. Krishnamachari, "Exploiting the wisdom of the crowd: localized, distributed information-centric vanets," *IEEE Communications Magazine*, vol. 48, no. 5, pp. 138–146, May 2010.
- [15] U. Lee, E. Magistretti, M. Gerla, P. Bellavista, and A. Corradi, "Dissemination and harvesting of urban data using vehicular sensing platforms," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 2, pp. 882–901, 2009.
- [16] S. Ni, Y. Tseng, Y. Chen, and J. Sheu, "The broadcast storm problem in a mobile ad hoc network," in *Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking (MobiCom' 99)*, New York, USA, 1999.
- [17] J. Whitbeck, V. Conan, and M. D. de Amorim, "Performance of opportunistic epidemic routing on edge-markovian dynamic graphs," *IEEE Transactions on communications*, vol. 59, no. 5, pp. 1259–1263, May 2011.
- [18] The NS-2 website. [Online]. Available: <http://www.isi.edu/nsnam/ns/>
- [19] The SUMO website. [Online]. Available: <tp://sourceforge.net/apps/mediawiki/sumo/>
- [20] A. Rachedi and A. Benslimane, "A secure and resistant architecture against attacks for mobile ad hoc networks," *Journal of Security and Communication Network (SCN)*, vol. 3, no. 2-3, pp. 150–166, 2010.