A New Framework for Data Collection in Vehicular Networks

Ismail Salhi, Mohamed Oussama Cherif, and Sidi Mohammed Senouci
Orange Labs, 2 Avenue Pierre Marzin, 22307, Lannion Cedex, France

***Abstract*— Vehicular Sensor Networks (VSNs) are an emerging paradigm in vehicular networks. This new technology uses different kind of sensing devices available in new vehicles, to gather information about the driver's environment (speed, acceleration, temperature, seats occupations, etc.) in order to provide a safer, more efficient and more comfortable driving experience. In this paper, we focus on a particular VSN architecture, where the ad hoc network is operated by a telecommunication/service provider (WiMax access point, 2.5/3G base station) to combine non-valuable individual sensed data and extract from them effective feedbacks about the situation of the road in a geographical area (traffic density, unusual traffic behavior, etc.). In operated VSNs, providers tend to reduce the traffic load on their network, using unlicensed spectrum communication medium (IEEE 802.11p, for example). To do so, we propose CGP (Clustered Gathering Protocol), a cross layer protocol based on hierarchical and geographical data collection, aggregation and dissemination mechanisms. We analyze the performances of CGP using a simulation environment and realistic mobility models. We demonstrate the feasibility of such solution and show that CGP offers the operator precious information without overloading his network.**

Keywords- VANET, dissemination, ITS, data collection, data aggregation, hybrid architecture, operated network.

I. INTRODUCTION

Over the last decade, the nature of wireless communications has evolved rapidly. The introduction of 3G and WLAN technologies and the recent standardization of WiMax have helped to realize the vision of ubiquitous connectivity. Currently, much research effort is focusing on exploiting this "always-on" feature for use in Transportation Systems. The primary objective of ITS (Intelligent Transportation Systems) is to improve traffic safety, efficiency, and travelling comfort.

In this work, we focus on the main component in Intelligent Transportation Systems, which is vehicular communication. Indeed, many car manufacturers are installing wireless connectivity equipments in their vehicles to enable communication between vehicles and also with the infrastructure. Vehicular Sensor Networks (VSNs) can be built on top of these vehicular networks by equipping vehicles with onboard sensing devices. In such case, sensors can gather a set of information like video data, speed, localization, acceleration, temperature, seat occupation, etc. Compared to traditional sensor networks, this recently emerged sensor network is not restricted by the power supply and the storage space. However, the typical scale of a VSN over wide geographic areas (e.g., millions of nodes), the volume of generated data (e.g., streaming data), and mobility of vehicles make it infeasible to adopt traditional sensor network solutions where sensed data tends to be systematically delivered to sinks using data-centric protocols such as Directed Diffusion [1].

Under such environment, an effective and efficient architecture for data collection and data exchange is more important. This work deals with such a system in a framework consisting of mobile vehicular sensor and road-side-units operated by a Telco or a service provider. Base stations are distributed over the road for collecting data from mobile vehicular sensor passing by. While mobile sensors on vehicles sense and send the information to road side units.

In our study, the VSN will be used by an infrastructure wireless network owner to gather "useless" individual information (see Fig.1) from each vehicle and to aggregate them inside an ad hoc wireless network using unlicensed spectrum links (IEEE 802.11, for example), to get a global view of the state of the road in a geographical area at a specific time, or to use these information as a database for a posterior treatment.

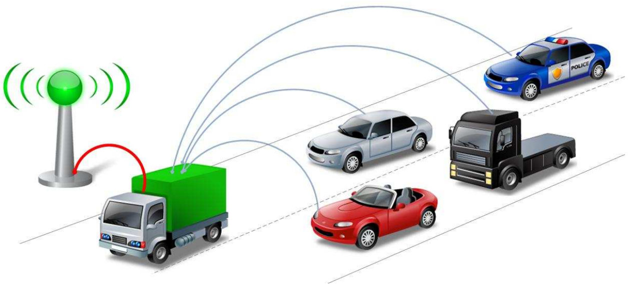

Figure 1. Figure 1 – Operated vehicular sensor network

The VSN in our case will be used to collect individual information from each vehicle and aggregate them inside the ad hoc wireless network (unlicensed spectrum). The aggregated information will be sent to the road side unit owned by the operator via licensed spectrum (WiMax or 2.5/3G). In fact, these sensors can generate massive amounts of sensed data and there is a need to collect, store, and retrieve them. The objective in such architectures for an operator/service provider is to reduce the use of its high-cost links. To do so, we present CGP (Clustered Gathering Protocol): a cross-layered protocol based on hierarchical and geographical data gathering, aggregation and dissemination. The goal of CGP is to gather data from all nodes in the vehicular ad hoc network in order to offer different kind ITS services. For example, a service-telecommunication provider can use CGP to provide:

- Real-time fuel consumption and pollution indicator, [5]
- A parking lots availability service, [3]

- A real-time traffic information service, by gathering all node's positions and velocities, [2]
- A geographical localization service for customers who want to follow their vehicles mobility (fleet management),
- Warning messages in a specific area, if unusual events happen (sudden speed decrease for example), [4]
- Surveillance applications such as proposed in [2] where nodes make videos of the road and detect and save the registration plates of vehicles around.

The rest of this paper is organized as follows. Section II presents vehicular sensor networks properties and the related network mechanisms. Section III describes CGP, the proposed data gathering protocol. Performance evaluation of CGP is presented in Section IV. Finally we conclude the paper and give some perspectives to our work in Section V.

II. BACKGROUND

This section provides an overview of vehicular sensor networks properties and their management mechanisms.

VSNs come out as a new brand of vehicular networks, whose purpose is the real-time gathering and diffusion of information. In [5], the author used a VSN for a better understanding of the traffic jam formation. They pointed out the fact that vehicular sensor networks are one of the costless solutions which tends to reduce traffic jams, CO2 emissions and fuel consumption. In [7], authors proposed the use of VSNs for security issues where agent nodes can look for a stolen car for example, by sending a query to all nodes that have crossed that vehicle. Another application of VSNs is the one proposed in [2] where the network provides the road users safer driving by sending alert messages in case of emergencies.

A VSN can be considered as a fusion of a Vehicle ad hoc network (VANET) and wireless sensor network (WSN). However, a VSN has some properties like: (i) higher capacity since the inboard sensors are supplied with more energy, storage and computing capabilities comparing to well known sensor networks, for example, (ii) huge amounts of data since a vehicle could be equipped by a lot of sensors (cameras, etc.), (iii) dynamic data-sinks management since data sinks could be mobile compared to traditional WSNs, and (iv) large scale connectivity since wide roads and grand avenues in urban environments can contain thousands of vehicles.

These specific characteristics have important implications for designing decisions in these networks. Thus, numerous research challenges (e.g. data dissemination, data aggregation, self-organization mechanisms) need to be addressed for vehicular sensor networks to be widely deployed.

A. *VSN Architectures*

As illustrated in Figure 2, data dissemination in vehicular sensor networks can be based upon three architectures:

- ***V2I***: infrastructure wireless links (GSM, UMTS, WiMax, etc.) are used to gather the data from VSN nodes.
- ***V2V***: where both the collection and the restitution of information are done within the vehicular network. This solution may be used for quick alert messages dissemination for example.
- ***Hybrid***: uses both V2V and V2I architectures.

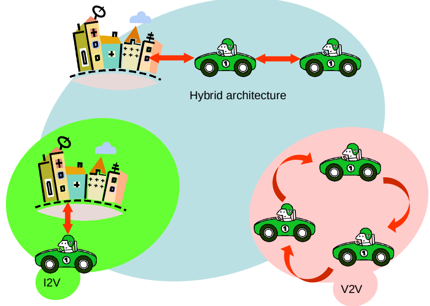

Figure 1- Vehicular communication architectures.

B. *Data Dissemination in VSN*

We found in the literature different approaches for data dissemination in a VSN:

1) *Opportunistic dissemination*

Due to the intrinsic network partitioning of VSNs, some works, such as [6], recommend the use of opportunistic diffusion of data, in which messages are stored in each intermediate node and forwarded to every encountered node till the destination is reached. Thus the delivery ratio is improved. However, this kind of mechanisms is not suitable for non-delay tolerant applications.

2) *Geographical dissemination*

Due to the fact that end to end paths are not constantly present in a VSN, a geographic dissemination is used in [2] by sending the message to the closest node toward the destination till it reaches it. Another way to do geographic dissemination is given in [6] where the authors show how to use geo-casting to deliver messages to several nodes in a geographical area.

3) *Peer-to-peer dissemination*

In a P2P solution, the source node stores the data in its storage device and do not send them in the network till another node asks for them. In [2], such architecture is proposed for delay tolerant applications.

4) *Cluster-based dissemination*

For a better delivery ratio and to reduce broadcast storms, a message has to be relayed by a minimum of intermediate nodes to the destination. To do so, nodes are organized on a set of clusters, in which one node or more (Cluster Head) gathers data in his cluster and send them after to the next cluster. Cluster-based solutions provide less propagation delay and high delivery ratio with also bandwidth fairness. In [4] the authors use a distributed clustering algorithm to create a virtual backbone that allows only some nodes to broadcast messages and thus, to reduce significantly broadcast storms.

We are interested in this paper, in the cluster based dissemination mechanisms combined with the geographical and opportunistic approaches.

C. *Data Aggregation in VSN*

Data aggregation is a well known concept in Wireless Sensor Networks (WSN); it allows the nodes to merge, update or delete some information because they might be duplicated, similar or expired. There are several aggregation mechanisms proposed in the literature for WSNs that can also be used in VSNs:

In [8], a timestamp aggregation technique is developed upon an opportunistic dissemination solution. In this case, if a node receives an information, it can decide if it is valid or not, by checking its sending time. Authors of [9] use a ratio-based and a cost-based algorithm to choose which information is important to aggregate and to estimate the error that can introduce a message in the data. Another approach in data aggregation is introduced on [10]. The authors use Flajolet-Martin Sketch [11] to estimate similar data in a set of N entities and to merge these information.

In this paper, we use a simple aggregation mechanism based on both timestamps and mean value calculation. Indeed, each set of nodes in a specific area will calculate an average value of collected data dropping some non-needed information (messages from another area, from other orientation, etc.)

III. CGP - Clustered Based Protocol

CGP is a cross-layered gathering, dissemination and aggregation protocol, based on a geographical clustering in a hybrid vehicular architecture. It collects costless information from every node within the network, aggregates them via free frequency radio communication links, and sends them to a provider via higher cost links, to provide real-time feedbacks about the road environment, using a specific platform.

CGP executes on the top of a hybrid vehicular sensor network architecture consisting of IEEE 802.11 clusters connected to the provider via cellular links. Additionally, it involves a GPS (Global Positioning System) synchronization mechanism between vehicles to improve the dissemination reliability and to figure out nodes localization. An application of CGP would be a real-time traffic feedback service. In this case, CGP will gather information about position, speed and orientation of every vehicle in a specific area via unlicensed spectrum links and send them to the provider via a cellular medium. Hence, the provider can estimate with accuracy the real-time traffic situation. The main purpose of CGP becomes to the reduction of the use of higher costs links (cellular links) with no loss of precision on the collected data from the VSN.

A. CGP geographical Environment

CGP uses a geographical clustering dissemination solution. The road is divided into S virtual segments (Figure 3) with the same length L. In each segment a cluster head is elected. It gathers data from all nodes in its segment and aggregates them, and sends the result to the next segment. For each BS, the number of segments is calculated depending on its communication range R (R=S×L) which represents the BS gathering perimeter.

B. Hypothesis

In the perspective of providing an efficient and scalable gathering and dissemination protocol, taking into consideration nodes' mobility and distribution, we decided to use a cluster-based solution for the V2V dissemination, and a geographic dissemination mechanism for the V2I communication.

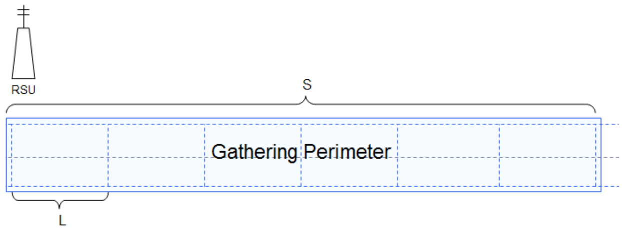

Figure 3- Gathering perimeter of a base station.

As we described in section 2.B, a cluster-based dissemination provides equity which is very important in a sensor network where every node has to send its data. It also reduces significantly broadcast storms and thus avoids collisions. Concerning the VSN-Telco communications, we use a geographical dissemination, to send aggregated data from the VSN to the road side unit, because it's the best way for a cluster head to send its data without using the high cost links.

For CGP execution, nodes must meet the following criteria:

- Nodes must be equipped with an IEEE 802.11 device for V2V communications and a cellular link device (Cell phone or Carbox) for V2I communication,
- Nodes must be equipped with a GPS device for localization and synchronization,
- Nodes know the road map and the BS placement,
- Each node must be able to determine its speed, position and orientation.)

C. CGP overview

Figure 4 shows the different steps of the execution of CGP. Each step starts periodically and has a predefined duration.

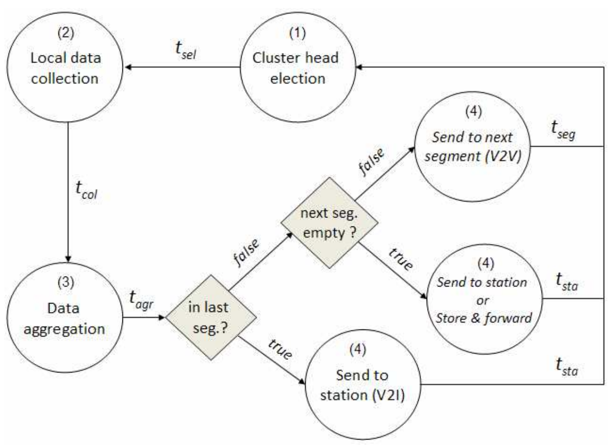

Figure 2 - CGP overview

The step (1) in the algorithm is the cluster head election phase; it allows a set of nodes in a segment, to decide which one of them will gather the information in the next phase. In (2), the elected cluster head collects the information from all the nodes in its segment, then it aggregates them (phase 3). When a cluster head reaches the step (4) of CGP, if it is in the closest segment to the station, it sends the collected data; else, it either sends to the next cluster head toward the BS, or sends them directly to the station or makes a store & forward, depending on the initial configuration of CGP (more details in sections 4, 5, 6).

1) Cluster head election

Since CGP is a distributed algorithm, all nodes must be able to know who is the cluster head in their segment. To do so, every node uses a back-off-based algorithm to announce that he is the cluster head. The cluster head election algorithm is described bellow:

- All nodes have to calculate their eligibility by estimating their position at the end of the election period

$$Pos(i, t+t_{selection}) > seg\_end \Rightarrow i \neq CH$$

- CH_ANNOUNCE message contains the identifier of the sender and its position,
- Each node is the cluster head till it gets a CH_ANNOUNCE message from another node or till the cluster head election period ends,
- Nodes calculate a back-off time to announce that they are cluster head to their neighbours using a CH_ANNOUCE message,
- If a node *i* receives a CH_ANNOUNCE message from *j* and *j* has a better position (closer to the end of segment), then *i* cancels its CH_ANNOUNCE message,
- The back-off duration is a random bounded integer that depends on the node proximity to the segment end position. It is calculated using the following formula :

$$t_{backoff}(i) = Rand(0, t_{collect}) + Priority(Pos(i,t), seg\_end)$$

Where:

tbackoff(i) is the back-off time of a node *i*,

Rand(x,y) is a function that returns a uniform distributed random integer bounded between x and y,

Priority(x, Seg_end) is a function that returns a period correlated with the distance between the node and the segment end position. Thus, the closer is the node to the segment end, the shorter will be the period,

Pos(i,t) is the node's (i) position at an instant t,

tselection is the cluster head selection duration,

tcollect is the local gathering period duration.

An example of the cluster head election is given in Figure 5. As we can see, node (a) is elected because it is the closest eligible node to the segment end. Node (b) is not eligible because it won't still be into the segment at the end of the local data collection phase.

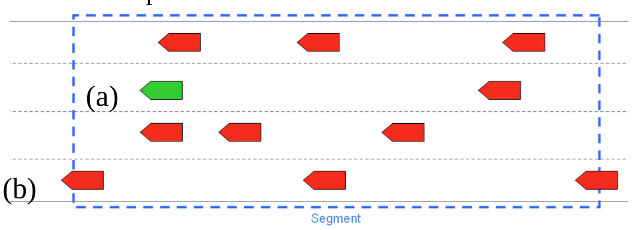

Figure 3 – Cluster head election.

2) Local data gathering

During this phase, all nodes in the segment send with unicast their sensed data to the cluster head, using a mechanism similar to DCF (Distributed Coordination Function, presented in IEEE 802.11).

i. Each node wait a random bounded back-off time,

ii. At the end of the back-off time, a node send a request to send to the cluster head,

iii. The cluster head acknowledges the reception by sending a Clear to Send message,

iv. The node sends its data to the cluster head.

An example of this step of the algorithm is shown in Figure 6.

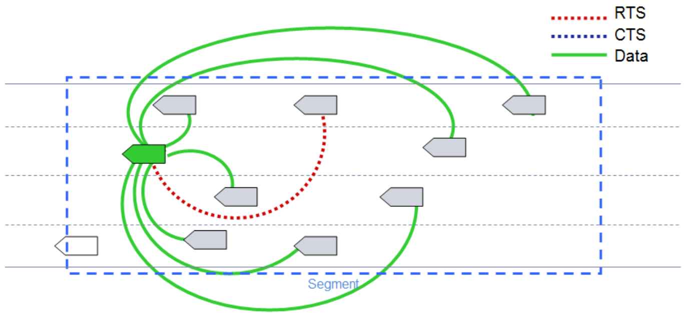

Figure 4 – local data gathering.

3) Data aggregation

Each cluster head aggregates the collected data in its segment. In our particular case study, each node sends its identifier, position and orientation. Thus, the cluster head will calculate the number of nodes in its segment, and the average speed.

4) Inter-Segment dissemination

When a cluster head is not in the closest segment to the base station, it automatically broadcasts its data to the next cluster head in the direction of the BS. The cluster head who receives these data, aggregates them with its own segment data. An illustration of this phase is given figure 7.

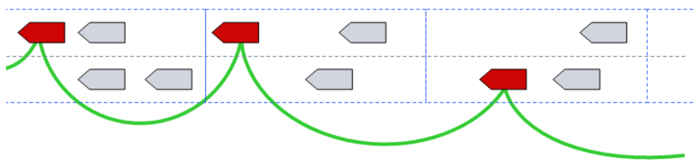

Figure 5 – Inter-Segment dissemination.

5) Segment-BS communication

There are two situations where a cluster head can send its information to the base station: (i) if there is no cluster head in the following segment and (ii) if it is on the closest segment to the BS. In both cases, the node aggregates the data and sends them to the BS using a cellular link. Then, the station will carry the information to the provider via its backbone network. Thus, the provider can take advantage of these information to have a real time vision of the road traffic.

6) Store & Forward

CGP can be configured to do store & forward instead of sending directly to the base station when the next segment is empty. In such situation, the node keeps the data in its memory during a parametric time, and waits for a cluster head in the next segment or till the node is in the closest segment to the base station. An example is Figure 8, where the node (a) cannot reach the next segment. In this case, it will store its data till it reaches the next closest segment to the station.

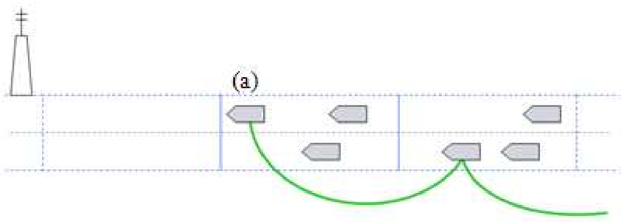

Figure 6- Store & forward example.

D. *CGP parameters*

Due to physical constraints and for a proper execution of CGP, we need to calculate the maximum segments number MAX_SEG_NUM that the BS can manage without loss of accuracy in the data. Table 1 describes CGP parameters.

TABLE I – CGP PARAMETERS

Parameter	Description
FULL_DURATION	Full duration of a CGP cycle, from the cluster head election to the dissemination to the BS
CHD_DURATION	Duration of the cluster head election phase
GAT_DURATION	Local data gathering phase duration
AGG_DURATION	The aggregation phase duration
DIS_DURATION	Inter-segment dissemination phase duration (with ACK), including the dissemination toward the base station
CGP_WAIT_ACK	The period that a C.H needs to wait, till it decides to send directly its data to the BS
COL_INT	interval between two complete CGP cycles
SEG_LEN	length of a single segment
LAN_NUM	number of Lanes in a one-way road
VEH_SPC	average inter-vehicles distance
VEH_WID	mean vehicles width
VEH_NUM	Maximum number of vehicles in a segment
TRA_RGE	radio transmission range of a vehicle
HL_PT	average hello propagation time
PK_PT	average data propagation time
IS_PT	average inter-segment packet propagation time
BS_PT	average propagation time for BS communications

We have:

$$\text{MAX\_SEG\_NUM} = \frac{\text{COL\_INT} - \text{CHD\_DURATION} - \text{GAT\_DURATION}}{\text{IS\_PT}}$$

$$\text{MAX\_SEG\_NUM} = \frac{\text{COL\_INT} - (\text{VEH\_NUM} * \text{HL\_PT}) - [(\text{VEH\_NUM} - 1) * \text{PK\_PT}]}{\text{IS\_PT}}$$

IV. PERFORMANCES EVALUATION

To validate and evaluate CGP, we have chosen Qualnet 4.5 simulation environment. We also extended and adapted the mobility model proposed in [11] to our needs. Our tool generates realistic random vehicles' displacements.

A. *Assumptions*

1) *Spatiotemporal environment*

We execute CGP on a straight road section partitioned into 18 equal segments, as depicted in Figure 9. The base station that covers all the section is present at one end point of the road. All the key parameters of our simulation are summarized in the following table:

TABLE II – SIMULATION SETUP

SIMULATION / SCENARIO		MAC / CGP	
Simulation time	600s	MAC protocol	802.11b
Map size	2500x2500 m2	Capacity	2 Mbps
Mobility model	VanetMob. [11]	Trans. Range	~266 m
Number of seg.	18	HL_PT	~0.1 s
Nodes	50 - 1000	PK_PT	~0.2 s
Vehicle velocity	0 – 108 km/h	IS_PT	~0.1 s
Segment length	100 m	FULL _DURATION	5 s
Road length	1.8 km	CHD_DURATION	1 s
Road width	15 m	GAT_DURATION	3 s
Number of lanes	2	AGG_DURATION	~0.1 s
Store & forward	Not used	DIS_DURATION	1 s

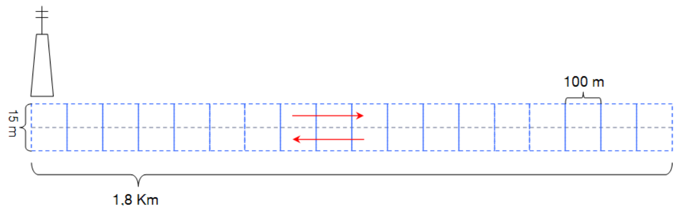

Figure 7 – Spatial environment.

B. *Simulation Scenarios*

1) *Scenario 1: Per node dissemination*

In this scenario, each node sends its collected data (speed, position, etc.) individually and periodically to the base station using the provider's cellular network. The aggregation in this case, is done at the Telco provider level. (See Figure 10)

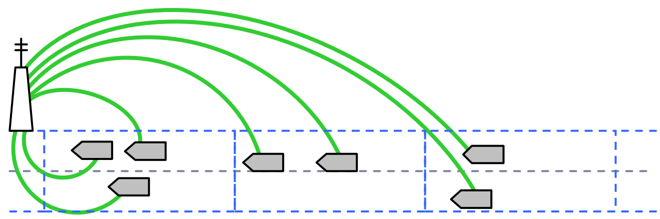

Figure 8 – Per node dissemination scenario.

2) *Scenario 2: Per Cluster Head dissemination*

In this scenario (see Figure 11), the local data gathering and aggregation are done at the segment level, as described in CGP. The aggregated data (average speed, number of nodes, etc.) are sent to the base station directly from the cluster head of each segment. The Telco provider will only aggregate the data from each segment.

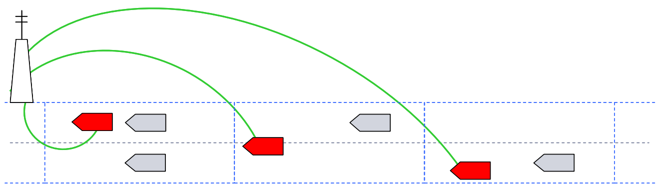

Figure 9 – Per cluster head dissemination scenario.

3) *Scenario 3: Complete CGP dissemination*

As depicted in Figure 12, CGP will be integrally executed in this scenario, from the cluster head election to the data dissemination to the provider.

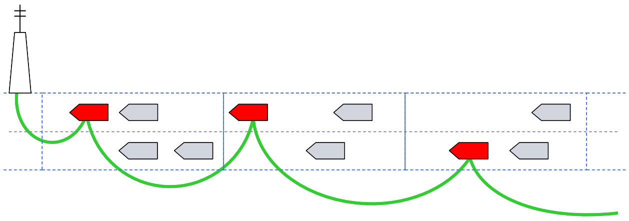

Figure 10- Complete CGP dissemination

C. Simulation Results

We calculate the number of messages sent to the base station via the provider's cellular network during 600 seconds. Thus, we can see in which scenario the data collection is the greediest in terms of cellular network usability.

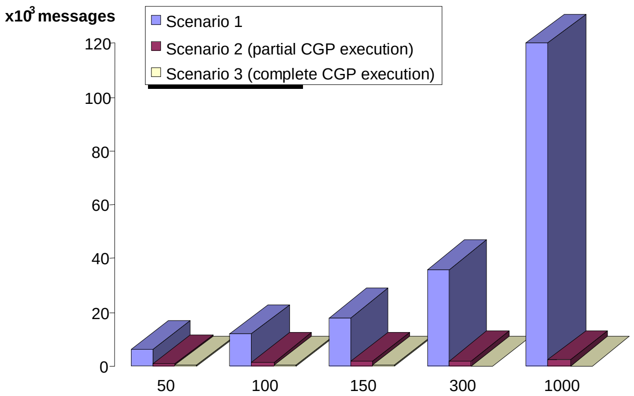

Figure 11 – Numbers of V2I messages

As we can see from this chart, the clear difference between scenario 1, where all nodes use the provider's links, and the V2V scenarios where nodes use the vehicular sensor network to send their data to the provider. The number of messages is reduced by 91.2 % in scenario 2 and by 99.16 % in scenario 3.

TABLE III – Overhead on V2V and V2I communications

Num. nodes	Number of messages (V2V)		Number of messages (V2I)	
	Scenario 2	Scenario 3	Scenario 2	Scenario 3
50	7 213	7 630	836	320
300	46 745	48 670	2 016	160
1000	237 967	240 127	2 128	120

Table 3 shows a minor variation of the overhead in V2V communications between scenario 2 and 3. Thus, we can see that with a negligible overhead, a complete execution of CGP is preferred because it reduces more (over 8%) the provider's links utilization, particularly in urban environment where large number of nodes is handled.

TABLE IV – CGP Overhead variation vs. speed

Average speed (km/h)	Number of messages
0 – 20	34 413
20 – 50	34 659
50 – 80	35 805
80 – 110	35 165

Table 4 shows the number of CGP messages when we vary the speed average value for 200 vehicles. We can see that the number of messages is quite stable. Thus from table 3 and 4 we can deduct that CGP is convenient with different kinds of mobility models like countryside (high speeds, low density) or urban (low speeds, high density) model and also that nodes velocity does not affect CGP performances. Indeed, the number of messages does not increase significantly when the nodes' speed increases.

Hence, we can conclude from these results the significant contribution of CGP in terms of decreasing the number of messages upon a provider's network.

V. Conclusion

In this paper, a novel data gathering and dissemination system (CGP, Clustered Gathering Protocol) based on hierarchical and geographical dissemination mechanisms on vehicular sensor networks is proposed. Designed for hybrid VANET architecture, it allows telecommunication/service providers to get valuable information about the road environment in a specific geographical area, using V2V network to minimize the high-cost links usability and base stations to gather information from the vehicles.

Simulations results of CGP demonstrate the feasibility of the proposed approach; moreover, they show that CGP reduces considerably the provider's network usability without any loss of accuracy in the collected data. We are currently extending this work by performing other extensive simulation in order to study all the CGP parameters.

References

[1] C. Intanagonwiwat, R. Govindan, and D. Estrin. "Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks." *ACM MobiCom'00*, August 2000, Boston, Massachusetts.

[2] U. Lee, E. Magistretti, B. Zhou, M. Gerla, P. Bellavista, A. Corradi, "Efficient Data Harvesting in Mobile Sensor Platforms". *IEEE PerSeNS'06 Workshop*, March 2006, Pisa, Italy.

[3] R. Panayappan, J. M. Trivedi, A. Studer, A. Perrig, "VANET-based Approach for Parking Space Availability", *ACM 4th workshop on VANETs*, September 2007, Montreal, Quebec, Canada.

[4] L. Bononi,M. Di Felice, "A Cross Layered MAC and Clustering Scheme for Efficient Broadcast in VANETs", *IEEE MASS'07*, October 2007, Pisa, Italy.

[5] E. Guizzo. "Network of traffic spies built into cars in Atlanta". *IEEE Spectrum*, April 2004.

[6] H. Wu , R. Fujimoto , R. Guensler , M. Hunter, "MDDV: a mobility-centric data dissemination algorithm for vehicular networks", *ACM international workshop on VANETs*, October 2004, Philadelphia, USA.

[7] U. Lee, E. Magistretti, M. Gerla, P. Bellavista, A. Corradi, "Dissemination and Harvesting of Urban Data using Vehicular Sensing Platforms", submitted for publication.

[8] L. Wischhof, A. Ebner, H. Rohling M. Lott and R. Halfmann. "SOTIS - A Self-organizing Traffic Information System". *IEEE Vehicular Technology Conference*, April 2003.

[9] T. Nadeem, S. Dashtinezhad, C. Liao and L. Iftode. "TrafficView: Traffic Data Dissemination using Car-to-Car Communication". *ACM SIGMOBILE'04*, July 2004.

[10] C. Lochert, B. Scheuermann and M. Mauve. "Probabilistic Aggregation for Data Dissemination in VANETs", *ACM VANET*'07, September 2007, Montreal, Quebec, Canada.

[11] P. Flajolet and G. N. Martin. "Probabilistic counting algorithms for data base applications". *Journal of Computer and System Sciences*, September 1985, Orlando, FL, USA.

[12] J. Haerri, M. Fiore, F. Filali, C. Bonnet, "VanetMobiSim: generating realistic mobility patterns for VANETs", *ACM VANET'06*, Sept. 2006.