Decentralized pull-based information gathering in vehicular networks using GeoVanet

Thierry Delot, Nathalie Mitton, Sergio Ilarri, Thomas Hien

HAL Id: inria-00599107
https://hal.inria.fr/inria-00599107

Submitted on 8 Jun 2011

Decentralized pull-based information gathering in vehicular networks using GeoVanet

Thierry Delot*, Nathalie Mitton†, Sergio Ilarri‡ and Thomas Hien*
*University Lille North of France, LAMIH FRE CNRS/UVHC 3304, 59313 Valenciennes, France
Email: Firstname.Lastname@univ-valenciennes.fr
†INRIA Lille-Nord Europe, University Lille 1, CNRS, 59650 Villeneuve d'Ascq, France
Email: nathalie.mitton@inria.fr
‡University of Zaragoza, IIS Department, Zaragoza, 50018, Spain
Email: silarri@unizar.es

Abstract—A vehicular ad hoc network (VANET) is a type of mobile network whose nodes are traveling cars and which communicate with one another using short-range wireless communications. These cars can exchange and share different information among them, which can lead to the development of interesting applications that require the cooperation of vehicles or using the vehicular network as a distributed computing platform. Along with the opportunities offered by vehicular networks, a number of challenges also arise.

In this paper, we focus on information gathering in vehicular ad hoc networks. Until now, only a few research works have addressed this problem. They have lead to solutions relying on push models, where potentially useful data are pushed towards vehicles. To the best of our knowledge, no work has tackled the use of pull models in VANETs. Such models would allow users to send queries to a set of cars in order to find the desired information. In order to propose such a query processing scheme, the main challenge to address is to route the different results towards their recipient in a highly dynamic network where the nodes move very quickly. To solve this issue, we propose GeoVanet, a DHT-based geographic routing protocol which ensures that the sender of a query can get a consistent answer. Our goal is not to compute the query result "instantaneously" but to ensure that the user will be able to retrieve it within a bounded time. To prove the effectiveness of GeoVanet, an experimental evaluation is provided in the paper. It shows that up to 80% of the available query results are delivered to the user.

I. INTRODUCTION

In the last decade, a number of small-sized wireless devices (e.g., PDAs or laptops) with increasing computing capabilities have appeared in the market at very affordable costs. These devices have started to be embedded into modern cars in the form of on-board computers, GPS navigators, or even multimedia centers. This has lead to the emergence of vehicular ad hoc networks (VANETs). In this kind of networks, cars traveling along a road can exchange information with other nearby cars. The lack of a fixed communication infrastructure, characteristic of ad hoc networks, implies that vehicles usually can communicate with one another using short-range wireless communications. Nevertheless, a piece of information can be disseminated and reach a far distance by using moving cars as intermediates, following multi-hop routing protocols [10].

To provide drivers useful information, accessing data in vehicular networks has become a major issue. Different existing schemes have therefore been considered. Some works rely on a *push model*. With such an approach, each vehicle receives information (e.g., about an emergency braking, a traffic congestion, an available parking space, etc.) from its neighbors and has to decide whether that information is relevant enough to be transmitted to the driver or not. The major difficulty for these solutions is how to disseminate data in the vehicular network so that vehicles receive the relevant information efficiently (timely and without unneeded overheads such as duplicate packets or irrelevant data) [16], [18].

Nevertheless, with such a push model, we cannot imagine that every data item will be communicated to every vehicle, as this would consume too much bandwidth and lead to high communication and processing efforts on the vehicles. On the contrary, only data about events that are potentially interesting for a large set of vehicles (e.g., an emergency braking or a traffic congestion) are diffused among the vehicles that the system estimates as potentially interested. Information about other events will not be disseminated, and so it is impossible to share information among a small set of interested vehicles, for example to build vehicular social networks.

Several works also use a *pull model* where a query is actually communicated to other vehicles in the vehicular network. This provides more flexibility in terms of the types of queries that can be considered, as opposed to the approaches based on a push model, since a query could in principle be diffused far away to retrieve remote data. This implies that vehicles should be able to understand, route, and process those queries. The basic idea, inspired by traditional Peer-to-Peer (P2P) systems, consists of diffusing the queries to different data sources either directly or using multi-hop relaying techniques such as [3], [10]. Then, each node can compute a partial query result based on its local data and then deliver it to the destination node. However, since no fixed data server or any kind of infrastructure is necessarily available in vehicular ad hoc networks, new techniques to access data are needed. Indeed, the mobility of nodes makes the management of an indexing structure, used in traditional P2P systems to decide how to route queries, impossible (as indicated in works such as [8]). These works must also face the problem of routing

the query results back to the query originator. This is indeed a challenge because the vehicle that issued the query can move in the meanwhile, and so routing the results based on simple geographic criteria may not be enough (it is difficult to know where the originator is currently located). Furthermore, since the vehicles keep moving, it is not even possible to ensure that there is at that moment a communication path to the originator node.

To overcome the issues related to the high mobility of nodes, the use of mobile telephony networks (as an alternative or a complement of a VANET) have been investigated. The queries can then be evaluated on a central server storing information sent by the vehicles. An example of a remote service for drivers is the *Waze* application (http://world.waze.com/), which is a social mobile application available on smartphones allowing Waze users to publish and consume real-time maps and traffic information. Maps can be provided to mobile users and traffic information is retrieved using a mobile telephony network. Anyway, the use of a central server leads to obvious scalability issues. The types of data shared are so limited to those interesting for a majority of the users. This approach can also cause confidentiality issues, since a driver may not accept to send his/her location for it to be stored by an untrusted peer.

To summarize, the different access techniques for vehicular networks mentioned previously are really interesting and have lead to the design of real systems. Nevertheless, they also impose severe limits. Indeed, whereas they are well-suited to the dissemination of information useful for a large set of users, they are not adapted to information-sharing in vehicular networks. Existing techniques only allow diffusing a small subset of the data shared on each vehicle (i.e., the one interesting for the majority of the vehicles nearby) to neighboring vehicles through the network. Different objectives justify such choices (e.g., bandwidth saving, difficulty to identify the set of recipients, etc.).

In this paper, we present an anonymous solution for drivers to share data in vehicular networks. Our goal is to allow one driver to query information shared by other vehicles. Therefore, a query has to be disseminated in the network in a bounded time. Then, once computed on remote nodes in the VANET, the results have to be delivered to the node that issued the query. In the following, we focus on the algorithms needed to solve this routing problem. We use a DHT-based model which identifies a fixed geographical location where a mailbox is dedicated to the query to allow the user to retrieve his/her results in a bounded time.

The rest of the paper is organized as follows. Section II presents some related works . Our assumptions and our model are introduced in Section III. Our proposal is detailed in Section IV. Section V presents our experimental evaluation. Finally, we conclude and present the perspectives of this work in Section VI.

II. RELATED WORK

Numerous recent research works have addressed the problem of information gathering in vehicular networks. Some of them focused on information dissemination. For instance, in [11], the authors focus on urban environments. They explain that the network connectivity is a limiting factor for information dissemination, since chains of vehicles are needed for broadcasting and a low traffic density may become a problem. The authors so make a clear distinction between data transportation via locomotion (i.e., vehicles carry data to areas where they can be disseminated, which is usually called store-and-forward [1], [5] or carry and forward [22]) and via wireless communications, emphasizing the problem of lack of network connectivity that may occur depending on the density of equipped vehicles in an area. The main idea is that *Stationary Supporting Units* (*SSUs*) are needed to alleviate this problem, and different heuristics are proposed to decide the best locations to place them. This is the main contribution of the work presented in [11], rather than proposing data dissemination protocols.

In [5], the use of a *propagation function* is proposed to decide the route that a message has to follow in order to reach a target spatial area. The originator of a message defines an appropriate propagation function (e.g., by considering traffic conditions for the current time frame), which can be interpreted as a "gravitational field" where the message is attracted towards areas of minimum potential. The route traversed by the message is thus the result of evaluating the propagation function at each routing hop. On the basis of this propagation function, different dissemination approaches (both deterministic and probabilistic) are proposed and compared. The metrics used for performance evaluation are: *message delivery* (ratio between the nodes receiving the message and the total number of nodes, where *delivery-IN* –within the target area– should be closed to 100% and *delivery-OUT* –outside the target area– should be minimized) and *network traffic* (total number of messages transmitted). There is no need of keeping information about neighbors, which would be very expensive. It is interesting to observe that the direction of a mobile node is considered in one of the analyzed dissemination strategies: *DFD-FSFD* (*Direction-aware Function Driven Feedback-augmented Store & Forward Diffusion*). However, how to define appropriate data propagation functions for different scenarios, which is a key element for the dissemination strategy, is not studied in such paper. Moreover, although the experimental results are promising, for both dense and sparse networks, more experiments with more mobility models (other than the Manhattan mobility model) are needed to represent realistic traffic patterns, as the authors themselves claim in relation to the protocol DFD-FSFD.

Besides, several works have addressed query processing in vehicular networks. Thus, *PeopleNet* [14] is an infrastructure-based proposal for information exchange in a mobile environment. However, it relies on the existence of a fixed network infrastructure to send a query to an area that may contain relevant information. Even though, once the query has arrived in the target area, it uses epidemic query dissemination through short-range communications within the area (to save economic costs), the answer to the query is communicated again to the

originator using the fixed network (e.g., by sending an SMS or email). Thus, problems related to query routing and result routing do not appear in this context. Moreover, this proposal does not focus on vehicular networks.

FleaNet [9] is a virtual market organized over a vehicular network. It proposes a *mobility assisted query dissemination* where the node that submitted the query periodically advertises it only to its one-hop neighbors, which will see if they can provide some answers from information stored on their local caches. With this approach the query spreads only due to the motion of the vehicle that submitted the query. This avoids overloading the network with many messages. It also solves the problem of routing back the results to the query originator, since it is only at one-hop distance. However, it is not general enough to process some kinds of queries, since the vehicle that submits the query must move near the vehicles which store the information needed.

Roadcast [21] is a content sharing scheme for VANETs. As in FleaNet, a vehicle can only query other vehicles that it encounters on the way. So, the problems of query and result routing do not arise either. In this case, keyword-based queries are submitted by the users and the scheme proposed tries to return the most popular content relevant to the query, as this content is likely to be useful for more vehicles in the future. Thus, not necessarily the most relevant data are returned for a query, as the popularity of the data is also taken into account.

In [19], a combination of pull and push is considered for *in-network query processing in mobile P2P databases*. When two mobile nodes encounter each other, they first exchange queries and results (pull phase). Then, they broadcast other popular data items that may help the other peer to improve its capabilities as source of relevant data in the future (push phase). Once again, multi-hop query/result routing is not considered.

In [6], the authors focus on multi-scale query processing in vehicular networks. They consider both the pull and push models and claim that different types of data sources (e.g., the local data cache containing data diffused by neighboring vehicles, remote web services, etc.) can be interesting for the evaluation of a single query. The authors explain that these data sources provide complementary information that increases the probability to answer the users' queries. However, no specific solution is proposed for the problem tackled in this paper.

In [20] the problem of searching documents in a vehicular network is considered. In this approach, the authors adapt the concept of *Distributed Hash Table* (*DHT*) [2] to a mobile environment and propose a *Hybrid Retrieval* (*HR*) approach which, based on the expected costs, adapts itself to choose between a flooding scheme and a DHT scheme for indexing and searching. The geographic space is divided into regions, such that each region must keep certain data. So, when a vehicle leaves a region it must transfer the documents belonging to the region to some other vehicle within the region. Thus, the purpose and approach of this work is quite different from those traditionally found in the context of vehicular networks. Besides, the problem of routing the results to the query originator is not discussed.

Finally, it is also interesting to mention that a mailbox approach has also been proposed in the context of mobile agent technology [17], in works such as [4]. Mobile agents are programs that can move from one computer to another and resume their execution on the target computer. Given the mobility of mobile agents, achieving an efficient communication among the agents is a challenge, as some mechanism is needed in order to route a communication to the location where an agent is currently executing. Moreover, the communication should eventually succeed even if the agent moves during the communication. Thus, the goal of the mailbox-based communication schemes is to increase the efficiency and reliability of the communication protocols by decoupling the agent migrations and the message delivery. However, although there is some relation, it should be noticed that the application context is totally different from the one considered in this work (distributed environments with mobile agents vs. vehicular networks). Thus, our proposal focuses on query processing in vehicular network, it preserves the anonymity of the user, and it is able to work independently of whether a fixed support infrastructure is available or not.

III. PRELIMINARIES

In this work, we consider a *vehicular ad-hoc wireless network* (*VANET*), in which vehicles (nodes) are aware of their geographical position and moving direction [12], [15]. VANETs are highly mobile networks whose nodes are vehicles traveling along a road or a highway, and which communicate among themselves using short-range (e.g., 100-200 meters) wireless communication devices (such as Wi-Fi or UWB). Using these *vehicle-to-vehicle* (*V2V*) communications provides some interesting advantages. The most important benefit is probably that there is no need to deploy an expensive and wide-coverage dedicated support infrastructure. Moreover, the users will not be charged by the use of such a network. Finally, there are applications that require a quick and direct exchange of data (without intermediate proxies or routers) between two vehicles within range of each other, for example in the area of safety.

In this paper, we assume that users of the VANET may need to spread queries in the network for which the answer is needed in a bounded time (but not necessarily immediately). For instance, let us consider a tourist visiting a city or a region and searching for information about the interesting places to see. Some of the vehicles receiving the query broadcasted by this user may belong to tourists sharing information about the sites they have visited. If the shared information matches the user's needs, it has to be delivered to him/her. In this scenario, the user does not require an immediate answer as long as information is gathered before visiting the city/region. As opposed to traditional query processing techniques, whose objective is to deliver the query result as quickly as possible, our goal here is to guarantee that the maximum amount of results will be delivered in a bounded time (e.g., with a deadline set to tomorrow). This allows spending more time

searching for relevant information, since it is here distributed over a large set of nodes. Our contribution in this paper resides in an anonymous DHT-based geographic routing protocol, called GeoVanet, which ensures the routing of query results in highly mobile and dynamic networks. Our protocol is suitable:

- both when an infrastructure is available and when it is not. Indeed, we assume that cities can be easily equipped (at each crossroad, for instance) with reliable infrastructure nodes, whereas it would be much more costly in the countryside to reach a sufficient coverage.
- both when some road maps are available and when they are not.

At this stage, the reader may wonder why we do not consider the use of mobile telephony networks (e.g., 3G) to share information between drivers but only a partial infrastructure when it is available. We indeed believe that the use of such mobile telephony networks presents several drawbacks:

1) First of all, not all devices are able to communicate using mobile telephony networks (e.g., connection not available everywhere, no subscription to a mobile telephony operator, etc.).
2) As opposed to short-range communication networks, mobile telephony networks are not free. The users may then not agree to be charged to provide others useful information. So, the fees related to the use of mobile telephony networks do not match with the cooperative applications considered in this paper.
3) As concerns privacy preservation, users may not accept that the personal data that they share with other users in their vicinity (e.g., information about the places they visited with the date of the visit, personal comments or pictures they may appear on) may be stored in a remote repository. However, the use of mobile telephony networks implies the use of such a repository where shared data will be stored.
4) Finally, bandwidth limitations and scalability issues may be encountered when considering a high number of drivers sharing important quantities of data with multimedia contents (e.g., pictures or videos of the places, audio descriptions, etc.).

Hence, in the rest of this paper we only consider the use of ad hoc communications between vehicles. Moreover, new short-range communication protocols dedicated for vehicular applications are under standardization, such as the Wireless Access in Vehicular Environments (WAVE) protocol [13]. They will provide vehicles and pedestrians the ability to communicate with each other and with the road-side infrastructure with a bandwidth up to 27 Mbps. This will also ease the exchange of important quantities of data between mobile units, for example to share multimedia contents.

Finally, even though we consider the use of digital maps for routing purposes in this paper, we assume that the destination of the vehicles is not known. The knowledge of the destination is for example required for route guidance using existing navigation devices, but these are used only occasionally, when the driver does not know his/her route. In our case, the system is *anonymous*, *not intrusive* and should work *without requiring information about the destination* of the cars.

IV. Proposed Routing Protocol

In this section, we present GeoVanet a DHT-based geographic routing protocol which spreads efficiently a request in a vehicular network in such a way that the sender can easily and quickly get a consistent answer within a bounded time. The DHT allows an indirect routing, i.e. the answers will not be directly routed to the requester but to a *rendezvous point* which is identified by a hashing key inserted into the request. Upon reception of a request, a vehicle able to answer it will route its answers towards this rendezvous point (and not towards the request initiator). When the initiator of the request wishes to retrieve the answers to its query, he/she will either drive to the rendezvous point or ask the rendezvous point about the answers. This feature makes GeoVanet preserve the sender's anonymity, since he/she does not need to give her identity within this process.

Note that the rendezvous point can be chosen in different manners. It may be a fixed node or a geographical area where the vehicles within have to maintain the information. Hashing keys can be distributed in several ways. For instance, they can be *(i)* such that every request from the same initiator will be stored at the same place to allow it to retrieve the answer to all his/her requests from the same place, or *(ii)* such that if several vehicles send the same request all theirs answers are gathered at the same place. These variants do not impact the way GeoVanet is running or the requester's identity.

In the rest of this section, we detail how GeoVanet works for every step of the process, that are: *(i)* the query broadcasting, *(ii)* the result delivery to the rendezvous point, and *(iii)* the retrieval of the data by the requester at the rendezvous point.

When a driver needs to broadcast a request, it sets up a query message QUERY =[request, exp-date, key], in which:

- 'request' is the core of the request i.e., 'what are the interesting sites to visit in Paris?'.
- 'exp-date' is the date by which the answer is expected.
- 'key' is used to determine the location where the answer should be sent and retrieved (i.e., the rendezvous point).

A. Notations

We denote by $N(u)$ the set of vehicles a vehicle u can directly communicate with. Since nodes are aware of their location, direction and speed, they are able to determine wether they are aiming or not to a rendezvous point d. We note $dir_{\to d}(u)$ the variable representing this relative movement of u towards d. $dir_{\to d}(u) > 0$ means that u is getting closer to d while $dir_{\to d}(u) < 0$ means that u is getting further from d. $dir_{\to d}(u) = 0$ means that u is static. In this latter case, u may be an infrastructure node.

B. Query broadcasting

Several solutions may be considered to spread the query expressed by one user over the vehicular ad hoc network. More precisely, the following strategies may be investigated:

- *Flooding*, where each vehicle receiving a message (i.e., the query to disseminate) relays it to its neighbors, without any limitation of the number of diffusions.
- *Contention-based forwarding*, where the principle "the farthest broadcasts" is applied to restrict the number of vehicles relaying the message [7]. In our case, in order to diffuse the query in every direction, the number of relaying vehicles should not be limited to one but only the farthest vehicle in each direction should broadcast, as done in [16].
- *Dissemination using fixed hotspots*. The use of fixed relays can help keeping the information alive in some areas where the density of vehicles would not be high enough.

In each of the above mentioned strategy, the diffusion stops by the expiration date associated to the query.

C. Processing of the query

In this paper, we do not concern about the query language used to formulate the query. Indeed, our proposal is independent from it. Numerous candidates are therefore available (e.g., SQL, XQuery, etc.). We also assume that the time needed to process the query on the vehicles holding interesting information is not significant when compared to the time needed to route the query result towards its destination.

D. Delivery of the query result

Upon reception of a Query message, the vehicle u first checks the expiration date. If the message is still valid, u forwards it (for query broadcasting) and checks whether it has information to provide to the Query sender. If so, u retrieves the location of the destination d that should gather every Reply message R. To do so, it uses the *key* of the Query.

The algorithm then used by the vehicles to forward the Reply message R to the destination node d is described in Algorithm 1. It runs as follows. Every Δt seconds, while the query message is still valid, u checks its direction. If it aims at d ($dir_{\to d}(u) > 0$), it keeps the message as long as it either reaches the rendezvous point d itself or it ends up getting further. In the former case, u delivers directly the reply message to d. In the latter case, u has to get rid of R. It then chooses among its neighbors v ($v \in N(u)$), that are either infrastructure nodes ($dir_{\to d}(v) = 0$) or aiming towards d ($dir_{\to d}(v) > 0$), the closest one, which will run the same algorithm, and so on until either R is delivered to d or it expires. The choice of the closest neighbor driving in the direction of the rendezvous point aims at maximizing the probability that the connection time between the two vehicles exchanging the query result is long enough to allow the complete transfer.

E. Retrieval of the results

When the query has expired, the node S that issued the Query Q will contact the rendezvous point d to retrieve every answer that has been gathered. S can either move towards d to query it directly or it can apply the same algorithm

(Algorithm 1). However, in this latter case, it would have to provide its identity or/and location to allow the report to be routed to it.

Algorithm 1 GeoVanet(u,d,R) - Run at vehicle u holding a Reply R to deliver to d.

```
 1: ToBeDelivered:=TRUE
 2: while ToBeDelivered == TRUE do
 3:    Every Δt seconds
 4:    if Q.exp-date ≤ Current-date then
 5:       ToBeDelivered := FALSE {Q has expired.}
 6:    else
 7:       if dir→d(u) > 0 then
 8:          {Vehicle u is moving towards the target position. It keeps
             the message.}
 9:          if d ∈ N(u) then
10:             Send Reply to d
11:             ToBeDelivered := FALSE {Destination is within the
                range of u. u successfully delivers its Reply. }
12:          else
13:             ToBeDelivered := TRUE {u keeps on holding Q.}
14:          end if
15:       else
16:          {Node u is getting further from d or it does not move,
             so it has to release Q.}
17:          N' = {v} such that v ∈ N(u) and dir(v)→d ≥ 0
18:          Send Q to the v ∈ N' which minimizes distance(u, v)
19:          {u selects the closest neighbor which is either an
             infrastructure node or moving towards d.}
20:          ToBeDelivered := FALSE
21:       end if
22:    end if
23: end while
24: RETURN;
```

V. EXPERIMENTAL EVALUATION

In order to evaluate our solution with an important number of vehicles over a significant period of time, we have tested it using a simulator. We needed a testing system that could simulate realistic vehicles' movements and wireless exchanges and support the implementation of both the query dissemination strategies and GeoVanet to route the set of the computed query results. Therefore, we chose to reuse the simulator we developed in the context of the VESPA project[1], which is a simulator dedicated to the evaluation of any data management system designed for vehicular networks.

A. Experimental Settings

During the simulations, the vehicles drive from a random departure point to a destination point through roads defined according to real maps. The choice of roads used for each vehicle to reach its destination is computed using the Dijkstra's shortest path algorithm. To calibrate our simulations (e.g., communication times, communication ranges, etc.), we extracted values from tests performed in real environments using our VESPA prototype.

As shown in Fig. 1, we have considered during our experimentations a real road network of an area surrounding the

---

[1]See http://www.univ-valenciennes.fr/ROI/SID/tdelot/vespa/

city of Lille (France), by using digital maps provided by Tele Atlas (http://www.teleatlas.com). We chose to present in this paper the results obtained for such an urban area because this clearly corresponds to the worst scenario for our GeoVanet algorithm. Indeed, with such a configuration, vehicles keep changing direction all the time, which considerably increases the probability that useless changes of the carrier of a message arise.

Fig. 1. Geographic area considered during the experimentations

Each simulation is 16000 seconds long and includes about 25000 vehicles. The evolution of the number of vehicles during the simulations is presented in Fig. 2. We tried to model a realistic flow of vehicles with several phases where the traffic increases or decreases. After a warm-up phase, one of the vehicles starts disseminating a query in the vehicular network. Each time a query reaches a vehicle holding a query result, this result is routed towards the rendezvous point (i.e., the mailbox) using GeoVanet, so that it can be eventually delivered to the user. We considered that only 2% of the vehicles carry a query result (we maintain constant this percentage during the whole simulation). Finally, the communication range used by the vehicles to exchange data is set to 200 meters.

B. Evaluation of the query dissemination

In this paper, we do not focus on the effectiveness of the query dissemination process. Nevertheless, we present in this section some results obtained for the query dissemination. Our objective here is to show that the percentage of vehicles contacted among those holding interesting information for a particular query is high enough for our approach to collect a suitable number of results. More precisely, we considered four dissemination strategies based on the ideas introduced in Section IV-B:

- *Flooding*. With this strategy, each vehicle receiving the query from one of its neighbors relays it.
- *Contention-based forwarding*. This strategy aims at avoiding problems like the "broadcast storm", which may
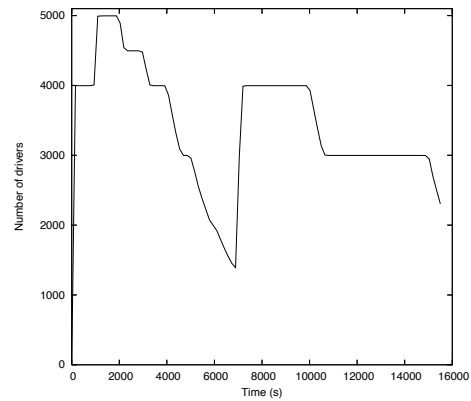

Fig. 2. Evolution of the number of vehicles during the simulation

happen with the previous flooding strategy. Therefore, we apply the contention-based forwarding principle described in Section IV-B to limit the number of vehicles relaying the query in the network.
- *Flooding using fixed hotspots*. In this configuration, we considered 4 fixed hotspots around the city of Lille. These hotspots receive the query formulated by the user and then regularly broadcast it in the network. Each vehicle receiving the query diffused by one of the access points then relays it.
- *Contention-based forwarding using hotspots*. This strategy relies on the same infrastructure than the previous one, but in this case the contention-based forwarding principle is applied when the vehicles receiving the query from the hotspot relay it.

For each strategy, the broadcast frequency is set to 400 seconds. During our experimentations, we evaluated the percentage of vehicles that hold a result and have received the query[2]. The results are presented in Fig. 3. They are quite good, since the percentage of relevant vehicles receiving the query is between 60% and 70% whatever the strategy used.

We also measured, for each strategy, the number of messages diffused. The results are depicted in Fig. 4. As expected, we observe that the traffic load is higher for the strategies based on flooding and on relays. The more messages are sent, the more vehicles are expected to be reached. However, the more messages are sent, the more expensive is the strategy in terms of energy spending and traffic overhead. There is thus a trade-off between the cost and the number of reached vehicles to consider when choosing the broadcasting strategy.

C. Evaluation of the delivery of the results using GeoVanet

In this section, we present the results obtained for the evaluations performed using GeoVanet to route the query results towards the fixed mailbox. Once spread in the vehicular network, the query generates different results which are routed

---

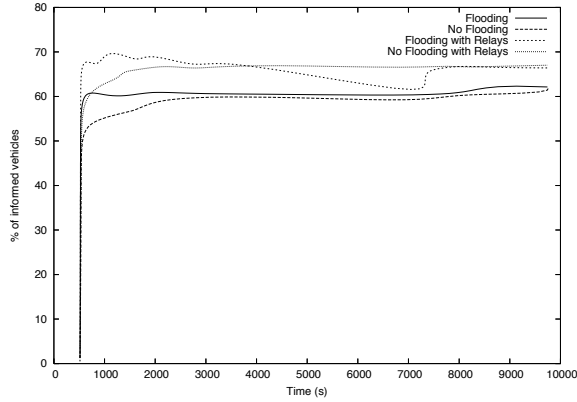[2]Videos illustrating the dissemination of the query in the network for the different strategies considered can be seen at the following address: http://www.univ-valenciennes.fr/ROI/SID/tdelot/GeoVanet/videos/

Fig. 3. Evolution of the percentage of vehicles that hold a result and have received the query over time

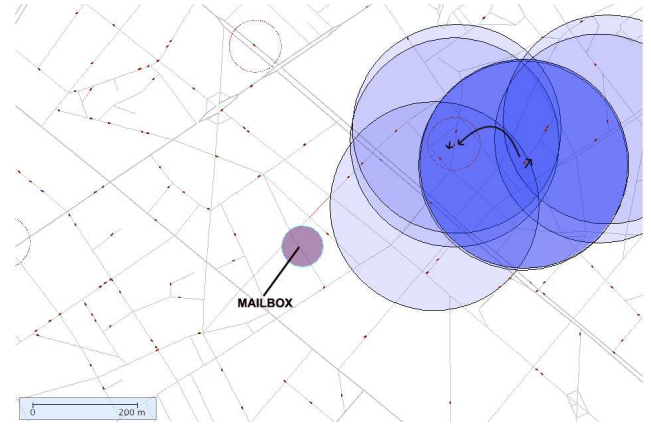Fig. 5. Graphical interface of the simulator for the results delivery step

also several other values (even if the results are not presented below due to space constraints) but no important changes were observed according to the variation of this parameter.

1) Implementation of GeoVanet on top of geographic vectors: First, we rely on geographic vectors to estimate the direction of the vehicles and determine if they are carrying the result towards the target (i.e., the mailbox) or not. Hence, we assume that vehicles are equipped with GPS receivers allowing to store previous locations of a vehicle to try to predict further ones. Obviously, the distance between the former (or reference) position and the current one plays a critical role. As illustrated in Fig. 6, a close reference position can be used to generate a vector indicating the instantaneous direction of the vehicle (called *direction vector* in Fig. 6) whereas a more distant one can be used to represent a vector that gives an indication of the global trajectory of the vehicle (called *mobility vector* in Fig. 6).
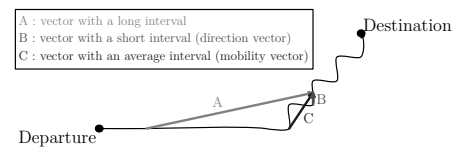

Fig. 4. Number of messages diffused for the dissemination strategies considered

Fig. 6. Influence of the reference position on the geographic vector

towards this mailbox. In Fig. 5, we present the graphical interface provided by our simulator during the routing phase of the query result. The small red rectangles correspond to vehicles driving on the roads. The rectangles surrounded by a dotted circle represent vehicles holding interesting information for the query considered. Finally, the larger circles correspond to the different messages exchanged between vehicles when the carrier of a result has to be changed (e.g., discovery of the neighbors' driving direction, diffusion of the result, etc.).

As explained in Section IV-E, we assume that the mobile user who formulated the query will be able to download the set of results collected in the mailbox after the deadline set for the query has expired. In the following, we consider two different configurations to evaluate our GeoVanet algorithm. In the first one, GeoVanet is implemented on top of geographic vectors, whereas the second one exploits the information on the road network provided by digital road maps. In the following, we consider a value of Δt equal to 10 seconds. We evaluated

To illustrate the efficiency of our GeoVanet algorithm, we performed numerous experimentations. In the following, we first study the impact of the size of the vector. Thus, Fig. 7 shows the evolution of the percentage of results collected in the mailbox over time according to the reference position considered to generate the vector. With the urban area we considered, it appears that the use of a vector providing a good indication of the vehicle's direction is the best choice to decide when the carrier of the information has to be changed. Moreover, we observe that we are able to successfully route more than 60% of the results transmitted to the target by the vehicles contacted. Besides, let us note that the time needed

to collect the majority of the results is very limited since most results are already available after one hour. This is satisfactory and proves that GeoVanet can be used even in scenarios where the time limit could be closer than the one we considered for the touristic application we introduced in Section III.
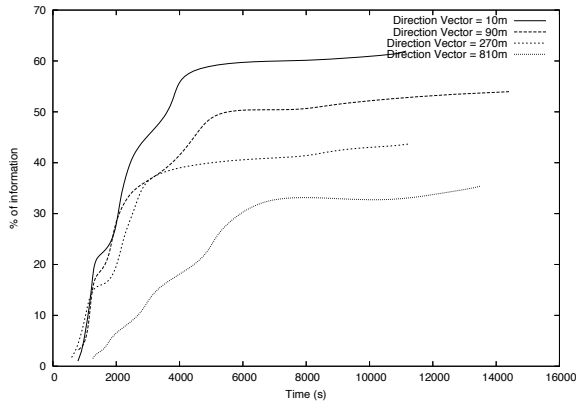

Fig. 7.   Percentage of information collected over time using variable length vectors

Obviously, the choice of the reference position (i.e., the size of the vector) also has an incidence on the network traffic. Indeed, the shorter the vector is, the higher the probability to decide to change the carrier. Nevertheless, as shown in Fig. 8, the accumulated number of diffusions does not significantly increase for the shorter vectors that provide an indication of the vehicle's direction.
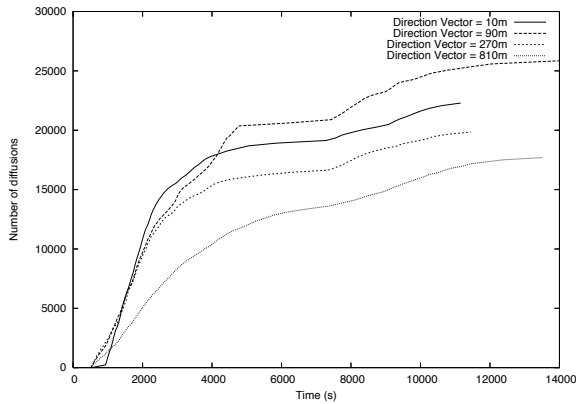

Fig. 8.   Evolution of the total number of diffusions over time

During our experimentations, we also observed the number of hops (i.e., changes of carrier) needed for a result to reach the mailbox between the moment when the vehicle carrying the result receives the user's query and the moment when the result joins the mailbox. Fig. 9 so presents the percentage of information collected according to the number of hops. Whatever the size of the vector, the number of hops needed to reach the mailbox is quite high (i.e., more than 30 hops to

navigate at most 7 kms for 50% of the results collected). Once again, these results can be explained by the urban environment considered in the evaluation, where the vehicles frequently change direction.
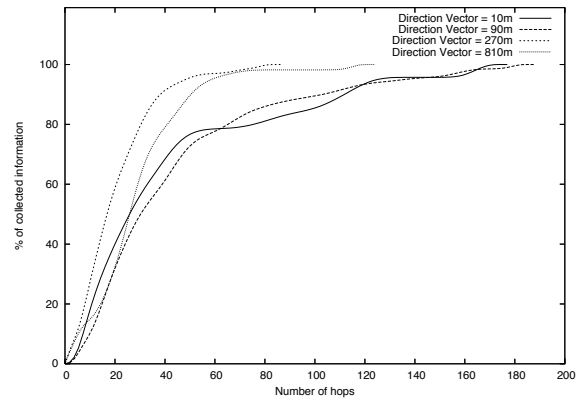

Fig. 9.   Percentage of recovered information according to the number of hops

As depicted in Fig. 7, we do not collect all the query results transmitted by vehicles reached with the query. One reason why we lose results using geographic vectors is due to the presence of loops. Due to the road topology, useless changes of carrier may arise because a car is temporarily driving in the wrong direction. More precisely, we observe in Fig. 10 that a vehicle (i.e., the black rectangle surrounded by a circle) may decide at one moment that it is not driving in the direction of the mailbox and choose to pass the result to carry to another vehicle. However, the first vehicle may be driving in the wrong direction only for a short period of time, and so deciding to change the carrier may be a bad decision. Moreover, the carrier may even choose a vehicle following exactly the same path and only temporarily driving towards the mailbox. Such loops may lead to delays or even to a loss of results which may never reach the target. This may also generate an overhead in terms of the number of changes of carrier and hops needed to reach the mailbox.
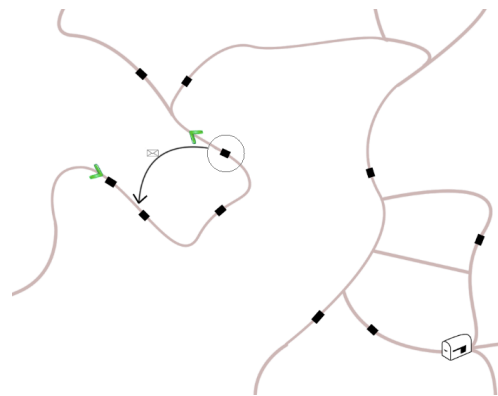

Fig. 10.   Illustration of possibles loops

2) Implementation of GeoVanet on top of digital road maps:
Using maps, each vehicle detects every $\Delta t = 20\ s$ that it is
not driving in the right direction if the shortest path between
this vehicle and the target (i.e., the mailbox) increases. In that
case, GeoVanet advocates a change of carrier in order to reach
the target. Even if we consider here digital maps, geographic
vectors are also necessary to determine the closest vehicle
driving towards the mailbox, which should become the new
carrier.

Fig. 11 shows the evolution of the results collected in the
mailbox for different thresholds. A threshold of 0% means that
the carrier of the result is changed as soon as an increase of
the shortest path is detected. More generally, with a threshold
set to $x\%$, the change is initiated only if the increase is greater
than $x\%$. We observe in Fig. 11 that the percentage of results
collected is higher if the change is initiated immediately when
the shortest path increases. Using digital maps, the quantity of
results collected is even higher than using geographic vectors.
Indeed, the percentage of results collected is close to 80% in
the best case. As concerns the time limit, we observe similar
results as using geographic vectors. Indeed, most of the results
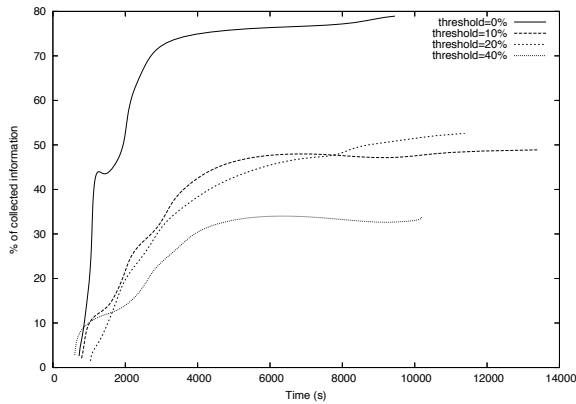are available at most one hour after the first dissemination of
the query.

Fig. 11.   Percentage of information collected over time using digital maps

As we did for the implementation of GeoVanet on top of
geographic vectors, we measured the number of messages dif-
fused while routing query results towards the mailbox. Fig. 12
shows a comparison between the results obtained for maps
(*Map-based approach*) and vectors (*GV-based approach*). We
considered in each case the parameter leading to the highest
percentage of collected results (i.e., *10 m* for the size of the
vector and 0% for the threshold). The exploitation of digital
maps leads to a reduction of the number of messages diffused.

Obviously, one solution to improve the percentage of results
collected in the mailbox, one solution is to increase the number
of relays at each hop. Fig. 13 shows that up to 90% of the
results available can be collected if the two closest vehicles
driving in the right direction are used as carriers. However,
this also leads to a severe increase of the network load, as
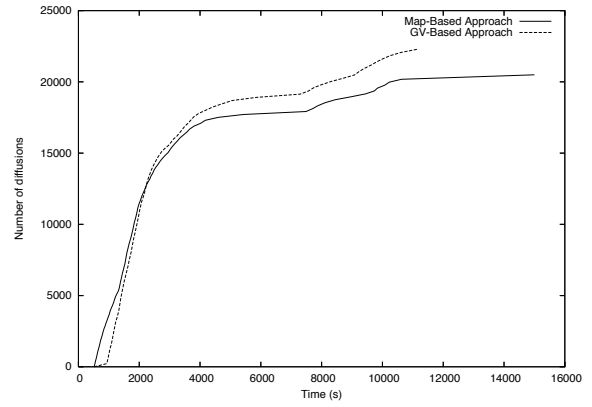

Fig. 12.   Comparison of the evolution of the number of messages diffused
over time between maps and vectors

depicted in Fig. 14, and to a lot of duplicates received at the
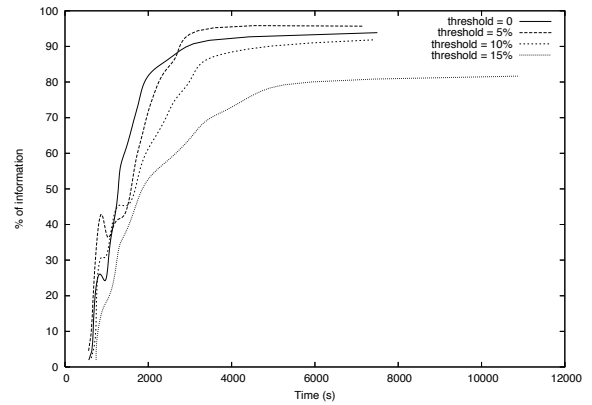mailbox.

Fig. 13.   Percentage of information collected over time by duplicating the
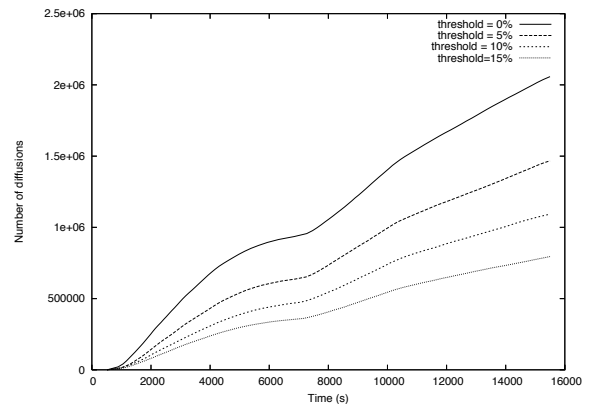carriers

Fig. 14.   Evolution of the number of messages diffused over time by
duplicating the carriers

Fig. 15 shows the number of hops needed to reach the mailbox considering the use of digital maps and geographic vectors. We kept the same parameters as for our previous study on the number of messages. We can observe that the number of hops is strongly reduced by using the information provided by digital maps, which also explains the reduction of the number of messages diffused observed in Fig. 12.
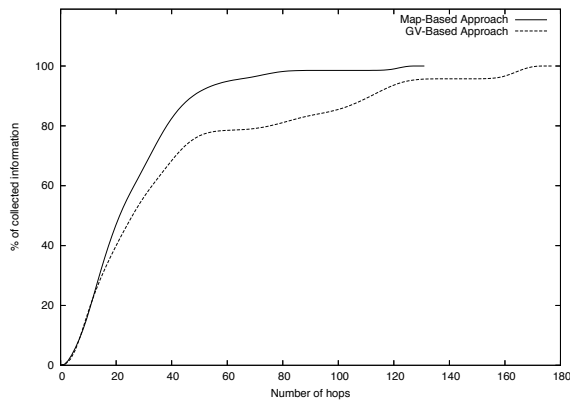

Fig. 15. Comparison of the percentage of recovered information according to the number of hops between maps and vectors

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we presented GeoVanet, our solution to process queries in vehicular networks and retrieve relevant data items shared by other nodes in the vehicular networks. GeoVanet therefore allows users to obtain these items within a bounded time. Through numerous simulations, we have proved the effectiveness of our solution, especially when it is implemented on top of digital maps.

In the future, we plan to study the possible use of several mailboxes and the impact of their position. Indeed, here it is crucial to ensure that the mailbox can be easily reached, so that a maximum number of results can be collected. Another interesting research direction will be to consider mixed environments where some vehicles may be equipped with digital maps and other are not.

REFERENCES

[1] C.J. Adler. Information dissemination in vehicular ad hoc networks. Master's thesis, University of Munich, April 2006.
[2] S. Androutsellis-Theotokis and D. Spinellis. A survey of peer-to-peer content distribution technologies. *ACM Computing Surveys*, 36(4):335–371, 2004.
[3] S Biswas and R. Morris. ExOR: opportunistic multi-hop routing for wireless networks. *ACM SIGCOMM Computer Communication Review*, 35(4):133–144, 2005.
[4] J. Cao, X. Feng, J. Lu, and S. K. Das. Mailbox-based scheme for designing mobile agent communication protocols. *Computer*, 35:54–60, 2002.
[5] P. Costa, D. Frey, M. Migliavacca, and L. Mottola. Towards lightweight information dissemination in inter-vehicular networks. In *Third International Workshop on Vehicular Ad Hoc Networks (VANET'06)*, pages 20–29. ACM, September 2006.
[6] T. Delot, S. Ilarri, M. Thilliez, G. Vargas-Solar, and M. Bellengier. Highly mobile query processing. In *International Conference on Ambient Systems, Networks and Technologies (ANT'10)*, 2010.
[7] H. Füβler, J. Widmer, M. Kasemann, M. Mauve, and H. Hartenstein. Contention-based forwarding for mobile ad hoc networks. *Ad Hoc Networks*, 1(4):351–369, November 2003.
[8] H. Huang, J. H. Hartman, and T. N. Hurst. Efficient and robust query processing for mobile wireless sensor networks. *International Journal of Sensor Networks*, 2(1/2):99–107, 2007.
[9] U. Lee, J. Lee, J.-S. Park, and M. Gerla. FleaNet: A virtual market place on vehicular networks. *IEEE Transactions on Vehicular Technology*, 59(1):344–355, 2010.
[10] C. Lochert, H. Hartenstein, J. Tian, H. Füßler, D. Hermann, and M. Mauve. A routing strategy for vehicular ad hoc networks in city environments. In *Intelligent Vehicles Symposium (IV'03)*, pages 156–161. IEEE Computer Society, 2003.
[11] C. Lochert, B. Scheuermann, M. Caliskan, and M. Mauve. The feasibility of information dissemination in vehicular ad-hoc networks. In *Forth Annual Conference on Wireless On demand Network Systems and Services (WONS'07)*, pages 92–99. IEEE Computer Society, January 2007.
[12] J. Luo and J-P. Hubaux. A survey of research in inter-vehicle communications. In *Embedded Security in Cars – Securing Current and Future Automotive IT Applications*, pages 111–122. Springer, 2005.
[13] Y. L. Morgan. Notes on DSRC & WAVE standards suite: Its architecture, design, and characteristics. *IEEE Communications Surveys Tutorials*, 12(4):504 –518, 2010.
[14] M. Motani, V. Srinivasan, and P. S. Nuggehalli. PeopleNet: engineering a wireless virtual social network. In *11th Annual International Conference on Mobile Computing and Networking (MobiCom'05)*, pages 243–257. ACM, 2005.
[15] S. Olariu and M. C. Weigle, editors. *Vehicular Networks: From Theory to Practice*. Chapman & Hall/CRC, 2009.
[16] N. Cenerario T. Delot and S. Ilarri. Vehicular event sharing with a mobile peer-to-peer architecture. *Transportation Research Part C: Emerging Technologies, ISSN 0968-090X*, 18(4):584–598, August 2010.
[17] R. Trillo, S. Ilarri, and E. Mena. Comparison and performance evaluation of mobile agent platforms. In *The Third International Conference on Autonomic and Autonomous Systems (ICAS'07), Athens (Greece)*, pages 41–46. IEEE Computer Society, ISBN 978-0-7695-2859-5, June 2007.
[18] B. Xu, A. M. Ouksel, and O. Wolfson. Opportunistic resource exchange in inter-vehicle ad-hoc networks. In *Fifth International Conference on Mobile Data Management (MDM'04)*. IEEE Computer Society, 2004.
[19] B. Xu, F. Vafaee, and O. Wolfson. In-network query processing in mobile P2P databases. In *17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS'09)*, pages 207–216. ACM, 2009.
[20] Q. Xu, H. Tao Shen, Z. Chen, B. Cui, X. Zhou, and Y. Dai. Hybrid retrieval mechanisms in vehicle-based P2P networks. In *Ninth International Conference on Computational Science (ICCS'09)*, volume 5544 of *Lecture Notes in Computer Science*, pages 303–314. Springer, 2009.
[21] Y. Zhang, J. Zhao, and G. Cao. Roadcast: A popularity aware content sharing scheme in VANETs. In *29th IEEE International Conference on Distributed Computing Systems (ICDCS'09)*, pages 223–230. IEEE Computer Society, 2009.
[22] J. Zhao and G. Cao. VADD: Vehicle-assisted data delivery in vehicular ad hoc networks. In *25th IEEE International Conference on Computer Communications (INFOCOM'06)*, pages 1–12. IEEE Computer Society, April 2006.