GeoVanet: A routing protocol for query processing in vehicular networks

Thierry Delot[a,*], Nathalie Mitton[b], Sergio Ilarri[c] and Thomas Hien[a]
[a]*University Lille North of France, LAMIH FRE CNRS, Valenciennes, France*
[b]*INRIA Lille-Nord Europe, Villeneuve, France*
[c]*IIS Department, University of Zaragoza, Zaragoza, Spain*

Abstract. In a vehicular ad hoc network (VANET), cars can exchange information by using short-range wireless communications. Along with the opportunities offered by vehicular networks, a number of challenges also arise. In particular, most works so far have focused on a push model, where potentially useful data are pushed towards vehicles. The use of pull models, that would allow users to send queries to a set of cars in order to find the desired information, has not been studied in depth.

The main challenge for pull models is the difficulty to route the different results towards the query originator in a highly dynamic network where the nodes move very quickly. To solve this issue, we propose GeoVanet, an anonymous and non-intrusive geographic routing protocol which ensures that the sender of a query can get a consistent answer. Our goal is to ensure that the user will be able to retrieve the query results within a bounded time. To prove the effectiveness of GeoVanet, an extensive experimental evaluation has been performed, that proves the interest of the proposal for both rural and urban areas. It shows that up to 80% of the available query results are delivered to the user.

Keywords: vehicular networks, mobile computing, query result routing

1. Introduction

Developing data management techniques that can be useful to build information services for drivers is a hot research topic (e.g., see [13,49]). A key driving force is that, 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) [36]. 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 communicate with one another by 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 [27].

To provide drivers with useful information, accessing data in vehicular networks has become a major issue. As we briefly describe in the following, push and pull approaches can be considered. The basic idea of these approaches is similar to the traditional push and pull approaches considered in the field of distributed and mobile databases [37,45]. The main difference is their application to a scenario

---

*Corresponding author: Thierry Delot, University Lille North of France, UVHC/LAMIH FRE CNRS 3304, Le Mont Houy, 59313 Valenciennes, France. Tel.: +33 3 27 51 19 56; Fax.: +33 3 27 51 19 40; E-mail: Thierry.Delot@univ-valenciennes.fr.

of vehicular networks, where there is no single server (several vehicles can act as servers for a given query or for certain data items) and each vehicle can communicate directly only with others within its communication range.

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) [11].

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 [1], consists of diffusing the queries to different data sources, either directly or by using multi-hop relaying techniques [27]. 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 [19]). 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) has been investigated [16,26]. 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. Waze users can 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 by each vehicle (i.e., the data 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 and non-intrusive solution for drivers to share data in vehicular networks, by describing a geographic mailbox-based routing protocol called GeoVanet. This paper extends and improves the initial proposal of GeoVanet presented in [14], including an extensive experimental evaluation, a discussion of the use of maiboxes, and more detailed explanations. Our goal with GeoVanet is to allow drivers 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. We present the algorithms needed to solve this routing problem. We use a 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 2 presents some related works. Our assumptions and our model are introduced in Section 3. Our proposal is detailed in Section 4. Section 5 presents our experimental evaluation. Finally, we conclude and present the perspectives of this work in Section 6.

2. Related work

Numerous recent research works have addressed the problem of information gathering in vehicular networks. In this section, we consider works focused on information dissemination, works that concern about query processing, and finally other works that are related to ours due to the similarity of some of the techniques proposed.

2.1. Works on data dissemination (*push-based approach*)

In [29], 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) 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.

In [10], 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).

As a final example of push-based approach for vehicular networks, *VESPA* (*Vehicular Event Sharing with a mobile P2P Architecture*) [9,11,12] is a system developed to support sharing information about

different types of events in inter-vehicle ad hoc networks. One of the interesting features of VESPA is that it has not been developed with a specific type of event or application scenario in mind, and therefore can deal with many types of events in a VANET (e.g., information about available parking spaces, accidents, emergency brakings, obstacles in the road, real-time traffic information, information relative to the coordination of vehicles in emergency situations, etc.). VESPA is based on the concept of *Encounter Probability* (*EP*) between a vehicle and an event (measure of the likelihood that the vehicle will meet the event) for both alerting the driver about important events [11] and for diffusion decisions in the proposed dissemination protocol [9]. EP values can be computed based on maps, which can increase the precision of the system in some situations [12].

2.2. *Works on query processing* (*pull-based approach*)

PeopleNet [33] 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 an 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 [25] 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 [51] 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 [47], 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 [13], 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 routing problem tackled in this paper.

In [48] the problem of searching documents in a vehicular network is considered. In this approach, the authors adapt the concept of *Distributed Hash Table* (*DHT*) [1] 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.

2.3. Other interesting related works

The problem of broadcasting/flooding has been widely studied (e.g., [28]), not only for the specific case of wireless ad hoc networks but for mobile environments in general (e.g., [46]). In [41], different routing protocols for vehicular networks are compared. Works such as [30,40] have studied the problem of routing a message towards a certain location.

As far as we know, only a few works have considered the problem of routing a message to a moving vehicle using multi-hop routing [22,24,44]. [24] tackles the problem of geographic routing in MANETs (mobile ad hoc networks) in scenarios where the communicating nodes are moving. In [22], the trajectories of the vehicles are considered for data delivery from infrastructure nodes to moving vehicles; however, the approach proposed requires knowledge about trajectories. In works such as [44], the problem of routing the results to a moving vehicle is also under study, but again some knowledge about the trajectories is needed. In this work, we apply some geographical routing but, contrary to related works, we consider mobile nodes and at the same time we want to preserve anonymity and avoid the need to know and manage information about trajectories. Using the mailbox-based approach presented in this paper we do not incur the overhead of managing trajectories and we are able to deliver the results more efficiently towards a static destination.

Finally, it is also interesting to mention that a mailbox approach has also been proposed in the context of mobile agent technology [43], in works such as [6]. 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 agent migrations and 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 networks, it preserves the anonymity of the user, and it is able to work independently of whether a fixed support infrastructure is available or not.

3. Preliminaries

In this work, we consider a vehicular ad-hoc network (VANET) in which vehicles (nodes) are aware of their geographical position and moving direction. 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.

3.1. Disseminating queries in a VANET

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 visit. 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 enough 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 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 to reach a sufficient coverage in the countryside.
– both when some road maps are available and when they are not.

3.2. Ad hoc communications vs. 3G wide-area communications

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 arise when considering a high number of drivers sharing an important amount of data with multimedia contents (e.g., pictures or videos of places, audio descriptions, etc.).

Moreover, new short-range communication protocols dedicated for vehicular applications are under standardization, such as the Wireless Access in Vehicular Environments (WAVE) protocol [32]. They will provide vehicles and pedestrians with 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.

Hence, in the rest of this paper we consider the use of ad hoc communications between vehicles and, only if available, with infrastructure nodes.

3.3. Use of mailboxes

With the proposed protocol, when a driver submits a query, a *mailbox* is assigned to the query. Several mailboxes will be available, which could communicate with each other (if necessary) by using a fixed network. The mailbox acts as a collector for the results of the query. Thus, it allows an indirect routing. The idea is that the answers will not be directly routed to the requester but to the mailbox assigned to his/her query, which is identified by a field inserted into the request. Different mailboxes can be set up as part of the fixed support infrastructure (thus, mailboxes could be considered as *infrastructure nodes*). The density of mailboxes in an area could affect the performance, as each mailbox will collect some of the query results and there may be a certain routing overhead to deliver the results to the mailbox and to retrieve the results from it. Thus, the location and distribution of mailboxes could be planned by following design principles inspired by those used in cellular telecommunication networks [39].

It should be noted that an alternative approach would be to consider that a mailbox is not a physical entity but just refers to an area. In such a case, vehicles in that area (or close to it, if there is no vehicle inside) would be responsible for maintaining the results available in the area. If the vehicles holding the results are going to leave the area, they would need to relay the information to other vehicles in the area. This alternative option would require a deep study on how to maintain the results alive in the area until they are picked up by the vehicle that submitted the query. As it is complementary to our approach (based on physical fixed entities called mailboxes) and the problem of keeping information alive in an area has been studied in other works [7], we do not study it further in this paper.

The use of the mailbox avoids the great difficulty that would arise if we intended to route the results directly back to the moving querying vehicle, as its location would be changing constantly and for privacy reasons no information about the route of the vehicle (if known) should be disseminated with the query. Indeed, even if the estimated route is available, routing the results would still be challenging, as it would be difficult to estimate the precise location of the destination vehicle at that moment. So, the advantage is that the mailbox is static, it has a fixed location and therefore routing the results towards it does not involve the aforementioned problem, especially the privacy issue. In this way, our scheme preserves the identity of the user.

Of course, due to this indirect routing, when the initiator of the request wishes to retrieve the answers to his/her query, he/she will either need to drive towards the mailbox (until the vehicle is within the communication range of the mailbox) to obtain the results using ad hoc communications (querying the mailbox can be considered a form of *V2I –vehicle-to-infrastructure–* communication [3]) or ask the mailbox remotely about the answers. In this latter case, the collecting mailbox will send the results by using the proposed routing algorithm (described in Section 4). The target of this routing process can be directly the user (assuming that he/she will not move during the process, as otherwise this would not be the most appropriate approach, due to the challenges involved in routing to a moving node), who in this case needs to provide his/her location, or another mailbox chosen by the user based on his/her planned

trajectory. Even if the results are routed directly towards the user, the collecting mailbox can first send the results (through the fixed network) to another mailbox that is located near the location of the user, in order to reduce the efforts invested in V2V communications (this routing process is transparent to the user).

It should be noted that the user could indicate with the query that he/she wants the mailbox to automatically send the results after the request expires (*push-based retrieval of results*). Alternatively, he/she could explicitly ask about the results when he/she is near any other mailbox (*pull-based retrieval of results*); since mailboxes are connected, this other mailbox will be able to contact the one which has collected the results and to retrieve them.

We assume that the available mailboxes and their locations are known by the vehicles, as this information can be easily obtained through a simple initial registration process when the software module implementing the proposed routing protocol is installed on the vehicles. Now the question is: how is an appropriate mailbox assigned to a query? In the following, we analyze three possible alternatives.

Option 1: Selection of mailbox by the user

The driver himself/herself could select a target mailbox among those available. For example, if he/she knows quite well his/her driving plans, he/she could choose a mailbox that is near his/her destination in order to facilitate the retrieval of the results (the vehicle could just query the mailbox when entering its communication range). Alternatively, he/she could choose a mailbox along his/her planned route, in order to pick up the results on his/her way to the final destination.

The main advantages of this strategy are its simplicity and the fact that the user is in control of the mailbox that is assigned for his/her query. However, it has also important disadvantages. First, the user has the responsibility of selecting an appropriate mailbox. Second, this strategy does not ensure load-balancing, as there may be popular mailboxes that are frequently selected as collectors of results of queries; this will not only increase the load at the mailbox itself but also the routing overhead near it. Third, this strategy may compromise the privacy of the querying vehicle. For example, based on the mailboxes selected for queries it may be possible to infer information about locations frequently visited by that vehicle. Another privacy concern is that if the mailbox is in an area with not many vehicles it may be easier to infer which vehicle submitted a certain query, as it will be more difficult to preserve a *k-anonymity* [42] with a high k (in this context, k-anonymity would imply that the issuer of a query could be any among at least k vehicles). Besides, it could happen that the driver does not know exactly where he/she is going or does not want to bother himself/herself with selecting an appropriate mailbox. In these cases, the system should automatically determine a mailbox for the query by using one of the other two options that we discuss in the following.

Option 2: Automatic assignment of mailbox based on geographic criteria

With this strategy, the idea is to exploit some information about the contents of the query in order to decide a suitable mailbox. In particular, it can be reasonably assumed that most queries of interest in the context of mobile users are associated to a certain *spatial scope*, that delimits the range of the area that is interesting for the query (e.g., in the so-called *range queries* and their variants) [21]. The scope of a query could be defined by indicating a simple rectangular or circular range or based on the concept of *location granules* proposed in [20]. So, for example, if a driver is approaching the city center and wants to retrieve information about available parking spaces, then the scope of the query will be the city center and its surroundings. If the user is asking about that area, it is likely that he/she is going towards it. So, allocating a mailbox located in that area seems a good idea.
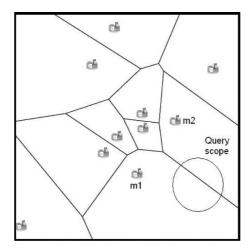
Fig. 1. Example of map of mailboxes and their Voronoi cells.

As there could be several mailboxes in a certain area, the one assigned to the query is selected based on the overlapping between that area and a set of Voronoi cells [2,50] defined by the locations of the mailboxes (see Fig. 1). For each location within each Voronoi cell there is no closer mailbox than the one defining that cell. So, the intersection between the spatial scope associated to a query and the Voronoi cells is obtained. For each intersecting cell, the percentage of overlapping is computed and finally the mailbox in the cell with the largest overlapping is selected (ties are solved randomly); for example, for the spatial scope indicated in Fig. 1, mailbox *m1* would be selected. The Voronoi diagram would be computed off-line. For simplicity, we can assume that a single Voronoi diagram is managed. Nevertheless, it could be partitioned to reduce the space needed to store it in the vehicles. On the one hand, a vehicle routing the results only needs to know the identifier of the target mailbox and its location, which can be embedded in the query.[1] On the other hand, a vehicle submitting a query needs to compute the target mailbox. For this, it could download from a server (if it does not have this information already locally available) only the part of the complete Voronoi diagram corresponding to the spatial scope associated to the query. In practice, a vehicle will eventually store locally fragments of the Voronoi diagram corresponding to the areas it frequently queries about, which will avoid the need to download this information from a server.

Of course, the procedure described here for an automatic assignment is just a heuristic approach and it may fail in assigning the best mailbox in some cases. For example, the user could want to query something about an area in order to decide whether he/she would like to go there or not; in this case, he/she will need to retrieve the query results before going there, and therefore assigning a mailbox within the area may not be the optimal choice. On the other hand, we consider that this strategy is flexible in the sense that different heuristics can be incorporated in order to decide the most appropriate mailbox in a given area (e.g., the decision can also be based on the current load of the candidate mailboxes).

The main disadvantage of this approach is that some mailboxes in popular areas could get easily overloaded, although (as mentioned above) there is some flexibility in the assignment of mailboxes and so load-balancing criteria could be considered. Another problem with this strategy is that privacy could

---

[1]Having information about nearby Voronoi cells would also be useful, as the vehicle could route the results to the target mailbox through a nearby mailbox, but for simplicity we do not consider this possibility here.

Table 1
Strategies for assigning a mailbox for a query: summary

Strategy	Load balancing	Privacy	Automatic assignment	Flexibility
1) The user selects	$\emptyset$	$\emptyset$	$\emptyset$	$\sqrt{}$
2) Geographic-based	$\approx$	$\emptyset$	$\sqrt{}$	$\approx$
3) Hash-based	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\emptyset$

be more easily violated because there exists a relation between the query and the mailbox assigned, and therefore probably also between the destination of the driver and the mailbox. The density of mailboxes and their Voronoi cells may have an impact on performance. Anyway, as commented later, we advocate a different approach for the assignment of mailboxes.

Option 3: Automatic assignment of mailbox based on hashing

Another solution is to use a hash function applied on a query identifier (a query identifier could be assigned by the vehicle launching the query, for example by concatenating the GPS current time and the location of the vehicle) and use the result returned by the hash function to allocate a mailbox randomly to the query. So, the idea would be to exploit a *Distributed Hash Table* (*DHT*) [1].

The main problem of this approach could be that the mailbox assigned can be located anywhere. Thus, the use of a uniform hash function will destroy the spatial locality information [52] (at least, unless some kind of *proximity-aware or locality-aware hash function* [38] is used). However, this can also be seen as an advantage, as the random assignment of mailboxes makes this approach less susceptible to privacy attacks (there is no apparent relation between a query/driver and the mailbox assigned, as the hash function may select a far away mailbox, a nearby one, or one in between). Besides, the use of a uniform random hash function will provide load-balancing. This approach leaves no flexibility for choosing a mailbox, as the one determined by applying the hash function must be selected. If the number of mailboxes available changes, the hashing must be re-adapted to the new situation.

It is interesting to mention that other works have proposed the use of distributed hash tables in mobile ad hoc networks and vehicular networks (e.g., [15,23]).

Conclusions

As described above, different strategies can be used to allocate mailboxes. The first strategy relies on the decision of the driver, whereas the two other strategies try to automatically locate a suitable mailbox. The first two strategies try to minimize the effort required by a vehicle to retrieve the results from the mailbox, whereas the last strategy attempts to achieve a good load-balancing and preserve the privacy of the user. A summary of the comparison among the three strategies considered is shown in Table 1, where $\emptyset$ represents a missing feature, $\sqrt{}$ represents an existing feature, and $\approx$ represents a feature partially supported.

In order to maximize the privacy of the user and offer a good load-balancing, we advocate the third option, which is based on hashing. Thus, a *key* field (the query identifier) is attached to each query message, which can be mapped to a mailbox identifier by applying a hash function. However, it is important to emphasize that the strategy used to choose the mailboxes does not have an impact on the routing process of GeoVanet.

GeoVanet preserves the sender's anonymity, since the driver does not need to provide his/her identity or intended route during the process. This is important because privacy protection is a major concern in vehicular networks [5,8] and in any kind of location-based system [17].

3.4. No need of digital road maps

To conclude this section, we would like to emphasize that, even though we consider the use of digital road maps for routing purposes with one of the possible implementations of the protocol that we propose (see Section 5.3.2), we do not assume that the destination of a vehicle submitting a query is known. The knowledge of the destination is for example required for route guidance when 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, non-intrusive and works without requiring information about the destination of the cars.

4. Proposed routing protocol

In this section, we present GeoVanet, a 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 use of mailboxes allows an indirect routing, as explained in the previous section. The routing protocol proposed consists of four steps:

1. Query broadcasting: the query is spread over the network.
2. Query processing: the query is processed locally on the vehicles that receive it.
3. Reply delivery: the answers are routed from the vehicles to the mailbox.
4. Data retrieval: the request initiator retrieves the results to his/her query.

In the rest of this section, we first introduce our notation and the two types of messages considered. Then, we detail how GeoVanet works for every step of the process, particularly the query broadcasting, the result delivery to the mailbox, and the retrieval of the results from the mailbox by the requester.

4.1. Basic notation

We denote by $N(u)$ the set of vehicles a vehicle u can directly communicate with. Since the nodes are aware of their location, direction and speed, they are able to determine whether they are aiming or not towards mailbox d. We denote $dir_{\rightarrow d}(u)$ the variable representing this relative movement of u towards d. This relative movement can be evaluated either in the Euclidean plane (if no maps are available) or based on maps, if available. Indeed, a vehicle following a road may get closer to a point based on maps and get farther based on Euclidean distances if the road makes some detours. $dir_{\rightarrow d}(u) > 0$ means that u is getting closer to d while $dir_{\rightarrow d}(u) < 0$ means that u is getting further from d. $dir_{\rightarrow d}(u) = 0$ means that u is static. In this latter case, u may be an infrastructure node.

4.2. Types of messages

We consider different kinds of messages: Query messages and Reply messages. On the one hand, a Query message represents the request issued by a vehicle. This message is as follows: 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' (query identifier) is used to determine the location where the answer should be sent and retrieved (i.e., the mailbox).

On the other hand, the Reply message is created by a vehicle which is able to answer the query and has to send its answers to the mailbox. This message is as follows: Reply = [request, data, exp-date, key], in which:

– 'data' is the core of the answer, e.g., 'Visit the Eiffel Tower'.
– 'exp-date' is the expiration date of the request to be answered. It is used to stop the routing in case the date expires before the reply reaches the mailbox.
– 'key' (query identifier) is used to determine the location of the destination of the reply. It is the same as in the Query message.

In the following, we describe the four steps of the proposed protocol.

4.3. Step 1: Query broadcasting

When a driver needs some information, he/she issues a Query message that is spread over the network. Several solutions may be considered to spread the query expressed by a user over the vehicular ad hoc network. More precisely, the following strategies may be considered:

– *Flooding*, where each vehicle receiving a message (i.e., the query to disseminate) relays it to its neighbors, without any limitation on the number of diffusions.
– *Contention-based forwarding*, where the principle "the farthest broadcasts" is applied to restrict the number of vehicles relaying the message. 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 [9].
– *Dissemination using a fixed infrastructure*. The use of fixed relays (usually called *Stationary Supporting Units –SSUs–* [29], *Road-Side Units –RSUs–* [31], or stationary gateways [34]) can help disseminate the query and keep it alive in some areas where the density of vehicles may not be high enough to ensure its propagation.

In each of the above mentioned strategies, the diffusion stops by the expiration date associated to the query. Any of these strategies, with their particular advantages and disadvantages, could be used for query broadcasting. Thus, works on data/query dissemination are complementary to our approach. We will compare these strategies in the context of GeoVanet in Section 5.2.

4.4. Step 2: 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.

4.5. Step 3: Delivery of the query results

Upon reception of a Query message, a vehicle u first checks the expiration date. If the message is still valid, u forwards it as needed (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 procedure then used by the vehicles to forward the Reply message R to the destination node d is described in Algorithm 1. In Algorithm 1, R can either be issued by the node u or by another node.

In this latter case, the node u is a forwarder of the original query. Algorithm 1 runs as follows. Every Δt seconds, while the Reply message R is still valid (i.e., while the expiration date that was set in the query and associated to R is valid), u checks its direction. If it aims at d ($dir_{\rightarrow d}(u) > 0$), it keeps the message as long as it either reaches the mailbox d itself or it ends up getting further from the mailbox. In the former case, u delivers directly the reply message to d. In the latter case, u has to pass R to a more promising vehicle. It then chooses the closest one among its neighbors v ($v \in N(u)$) that are either infrastructure nodes ($dir_{\rightarrow d}(v) = 0$) or aiming towards d ($dir_{\rightarrow d}(v) > 0$), 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 mailbox 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. To determine whether a vehicle is driving towards a mailbox or not, either simple geographic computations or computations based on road maps could be considered (evaluated in Sections 5.3.1 and 5.3.2, respectively).

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 R.exp-date ⩽ Current-date then
5:          ToBeDelivered := FALSE {R 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 R 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 R.}
14:             end if
15:          else
16:             {Node u is getting further from d or it does not move, so it has to release R.}
17:             N' = {v} such that v ∈ N(u) and dir(v)→d ⩾ 0
18:             {u selects the closest neighbor which is either an infrastructure node or moving towards d.}
19:             Send R to node v ∈ N' which minimizes distance(u, v)
20:             ToBeDelivered := FALSE
21:          end if
22:       end if
23:    end while
24:    RETURN;
```

Notice that some variants to this basic scheme could be considered to try to maximize the performance of the protocol. For example, it is possible to consider several carriers simultaneously to route a message towards its destination (this will be evaluated in Section 5.3.3). As another example, even if the current vehicle is driving towards the mailbox, it could be more efficient to relay the message to another vehicle that could approach it faster. This last optimization would require extra knowledge about the status of the neighbors of a vehicle and so it is not further explored in this paper.

4.6. Step 4: Retrieval of the results

When the query has expired, the node S that issued the query Q will contact the mailbox d to retrieve every answer that has been collected. As explained in Section 3.3, S can either move towards d to query

it directly or it can send a message to d asking for the results. However, in this latter case, it would have to provide its location to allow the result to be routed to it (and it should remain at the same location until the results have arrived); the same algorithm (Algorithm 1) would be used first to route the message asking for the results from S to d and then to route the results from d to S.

Notice that, in some cases, it may be even possible for the vehicle that submitted the query to contact the mailbox directly by using a 3G connection. However, even if this connection may also be available for other vehicles to route their results to the mailbox, using ad hoc communications during step 3 is convenient and economical, which will also encourage the participation of the vehicles. Thus, our approach does not assume the existence of a support infrastructure (other than mailboxes).

5. 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 (see http://www.univ-valenciennes.fr/ROI/SID/tdelot/vespa/), which is a simulator dedicated to the evaluation of any data management system designed for vehicular networks. We have successfully used this simulator in several works before (e.g., [9,11,12,14]).

In this section, we first describe the experimental settings and then different experiments that we have performed to evaluate our approach. It should be noted that in a highly dynamic distributed environment like the one presented in this paper it is not possible to guarantee that a complete answer has been obtained (the traditional *closed world assumption* gives way to an *open world assumption* [4]), and therefore an interesting metric to consider for evaluation will be the amount or percentage of interesting results retrieved.

5.1. Experimental settings

During the simulations, the vehicles drive from a random departure location to a random destination location through roads defined according to real maps. The choice of roads used for each vehicle to reach its destination is computed by 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. 2, we have considered during our experimentations a real road network of two different areas by using digital maps provided by Tele Atlas (http://www.teleatlas.com). The first one (Fig. 2(a)) is an urban area, specifically the area surrounding the city of Lille (France) while the second one (Fig. 2(b)) is a rural area, specifically a countryside area in the North of France. We chose to present results obtained for such areas because they clearly correspond to two different scenarios for our GeoVanet algorithm. Indeed, on the one hand, in an urban area vehicles keep changing direction all the time, which considerably increases the probability that useless changes of the carrier of a message arise. On the other hand, in a rural area vehicles keep the same direction for a long time but the density of vehicles is much lower than in a city and thus next forwarders are rarer.

For each area considered, GeoVanet is evaluated by measuring the quality of the query dissemination and the retrieval of results. In the latter case, GeoVanet is run both when nodes are aware of existing road maps and when they are not.
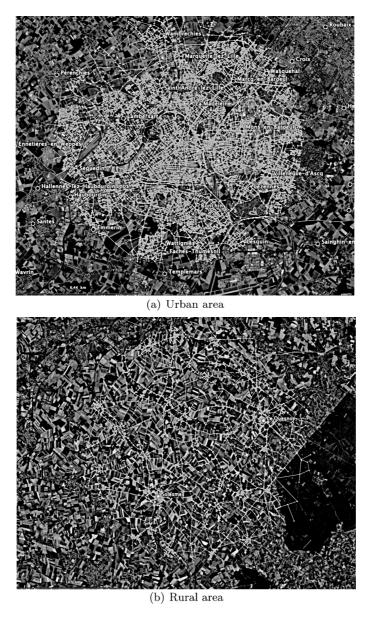
(a) Urban area

(b) Rural area

Fig. 2. Geographic areas 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. 5(a). 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 mailbox by 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.

5.2. Evaluation of the query dissemination

In this paper, we do not focus on the effectiveness of the query dissemination process, as this is complementary to the GeoVanet protocol. 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 consider four dissemination strategies based on the ideas introduced in Section 4.3:

- *Flooding*. With this strategy, each vehicle receiving the query from one of its neighbors relays it immediately.
- *Contention-based forwarding*. This strategy aims at avoiding problems like the "broadcast storm" [18,35], which may happen with the previous flooding strategy. Therefore, we apply the contention-based forwarding principle described in Section 4.3 to limit the number of vehicles relaying the query in the network. In order to better adapt the process to the environment, we apply two different contention-based forwarding strategies depending on whether the vehicle density is high or not. In an urban environment, where a message is more likely to be received by another vehicle when sent, the message is sent once (*one-time contention-based forwarding*). In a rural environment, the message is regularly sent till it has been received (*periodic contention-based forwarding*).
- *Flooding using a fixed infrastructure*. In this configuration, for urban settings we considered 4 fixed hotspots around the city of Lille. Once these hotspots have received a query they regularly broadcast it in the network. For rural settings, a few hotspots are set in the center of the town or village (in main places like the town hall or the church), in order to facilitate the dissemination when the density of vehicles is low. Each vehicle receiving the query diffused by one of these access points then relays it.
- *Contention-based forwarding using an infrastructure*. 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.

When applicable, the broadcast frequency is set to 400 seconds. During our experimentations, we evaluated the percentage of vehicles that held a result and received the query. Some 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/. The experimental results are presented in Fig. 3. All the strategies behave quite well in the urban area, since the percentage of relevant vehicles receiving the query is between 60% and 70% whatever the strategy used. In such an environment, every strategy acts similarly. On the contrary, in a rural environment, the classical flooding performs very poorly compared to other strategies that reach between 70% and 90% of the vehicles owning information of interest. This is due to the low density of vehicles. With the flooding strategy, once a message is relayed, if there is no vehicle within communication range to receive it, the message is lost. By using the periodic contention-based strategy, a message is periodically sent till being reached by another vehicle, which improves the performance of the diffusion. In the same way, when there are some fixed hotspots relaying the message when a vehicle is within range, the results are better.

We also measured, for each strategy, the number of messages diffused. The results are depicted in Fig. 4. 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 consumption and traffic overhead. Therefore, there is a trade-off between the cost and the number of vehicles reached, that must be considered when choosing the query broadcasting strategy. As expected, we observe that the network
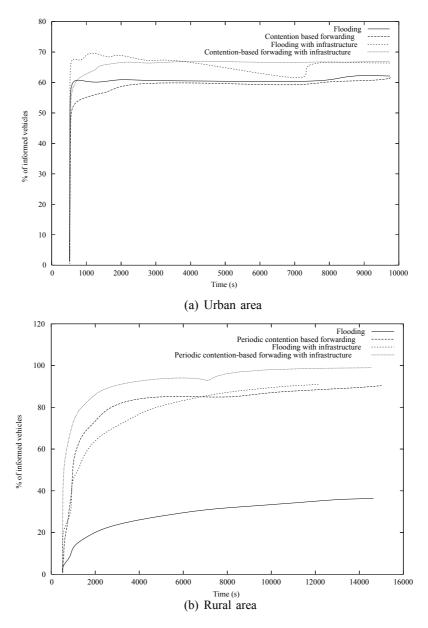
(a) Urban area

(b) Rural area

Fig. 3. Evolution of the percentage of vehicles holding results that have received the query over time.

traffic load for the strategies based on flooding is higher in an urban environment, since the density of vehicles is high. In a rural environment there are less diffusions with the plain flooding strategy because the message is quickly lost. Another interesting point is the difference in scale of the number of diffusions between both environments (compare the Y-scale of Figs 4(a) and 4(b)). This can be explained by the fact that in a rural environment a periodic contention-based forwarding is needed, unlike in an urban environment, to alleviate the problems derived from the low density of vehicles. Thus, in a rural environment, a vehicle will keep sending regularly a message till reaching a new vehicle not already aware of it.

(a) Urban area

(b) Rural area
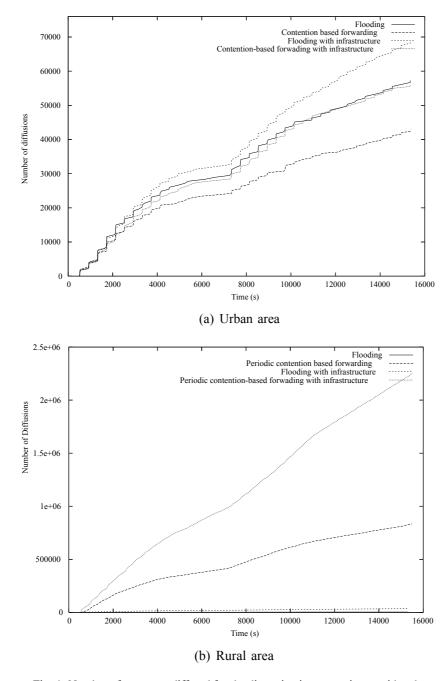
Fig. 4. Number of messages diffused for the dissemination strategies considered.

5.3. 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 a fixed mailbox. Once spread in the vehicular network, the query generates different results which are routed towards the corresponding mailbox. For this evaluation, we focus on
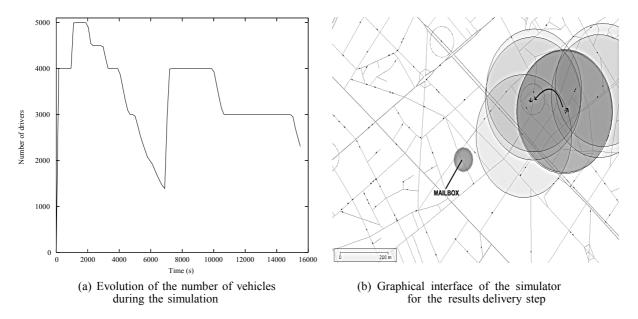
(a) Evolution of the number of vehicles
during the simulation

(b) Graphical interface of the simulator
for the results delivery step

Fig. 5. Simulation settings.

the case where no infrastructure (other than the target mailbox) is available, which can be considered a worst-case situation. In Fig. 5(b), we present the graphical interface provided by our simulator during the routing phase of the query result. The small 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 4.6, 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 expiration. 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 (vehicles are not aware of road maps), whereas the second one exploits information about the road network provided by digital road maps. In the following, we consider a value of Δt (see Algorithm 1) equal to 10 seconds. We evaluated also several other values (even if the results are not presented due to space constraints) but no important changes were observed with the variation of this parameter.

5.3.1. Implementation of GeoVanet on top of geographic vectors

First, we rely on geographic vectors to estimate the direction of the vehicles and determine whether they are carrying the results towards the target (i.e., the mailbox) or not. Hence, we assume that vehicles are equipped with GPS receivers supporting the storage of previous locations to try to predict future 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).

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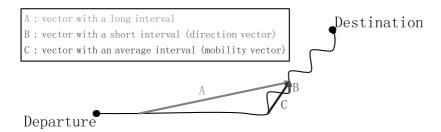
Fig. 6. Influence of the reference position on the geographic vector considered.

percentage of results collected in the mailbox over time according to the reference position considered to generate the vector. In both environments, it appears that the use of a vector providing a good indication of the vehicle's direction (i.e., a direction vector) 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 in the urban area and more than 40% in the rural area. 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 much smaller than the one we considered for the touristic application we introduced in Section 3.

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, whatever the environment. Nevertheless, as shown in Fig. 8, the accumulated number of diffusions does not significantly increase for the shorter vectors, that provide a precise indication of the direction of the vehicle.

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. 9 that a vehicle (i.e., the 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.

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 query and the moment when the result joins the mailbox. Figure 10 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 km for 50% of the results collected) and similar in both environments. In the urban environment, these results can be explained by the facts that the vehicles frequently change direction and the traffic is dense, and thus a message is very likely to reach a vehicle heading in the right direction. In the rural environment, this can be explained by the low density of vehicles and the use of distance vectors. Because of the low density of vehicles, as soon as a vehicle apparently heading towards the right direction is passed by, it
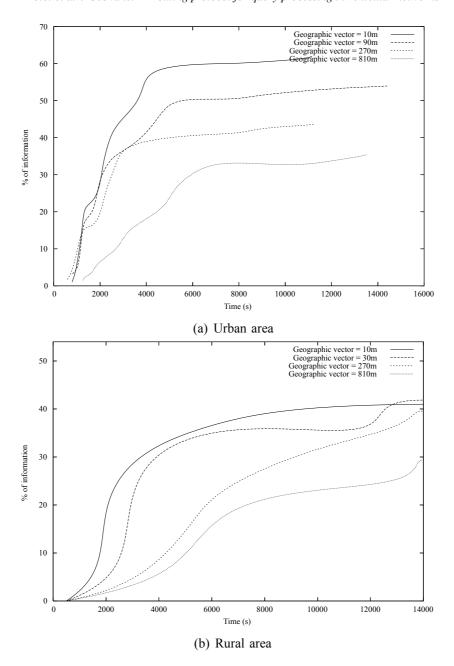
(a) Urban area

(b) Rural area

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

takes charge of the message. However, since the routes are longer, the route may actually head towards the wrong direction. Since the density is low, a large amount of time may pass before crossing another vehicle.

5.3.2. Implementation of GeoVanet on top of digital road maps

Using maps, each vehicle detects every $\Delta t = 20s$ that it is not driving in the right direction if the shor-

(a) Urban area

(b) Rural area
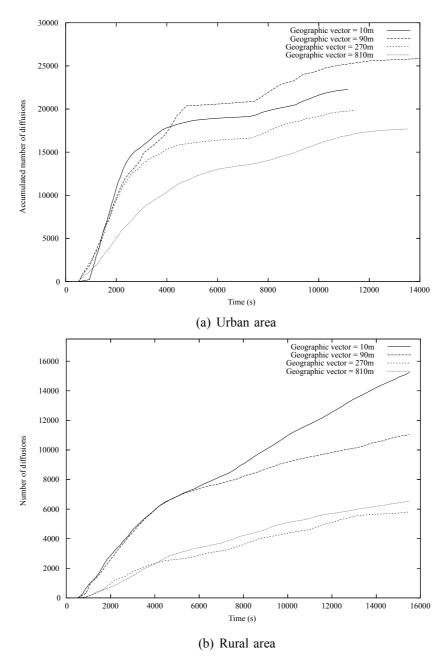
Fig. 8. Evolution of the total number of diffusions over time.

test 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.

Figure 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
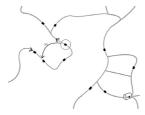
Fig. 9. Illustration of possible loops.

(a) Urban area

(b) Rural area
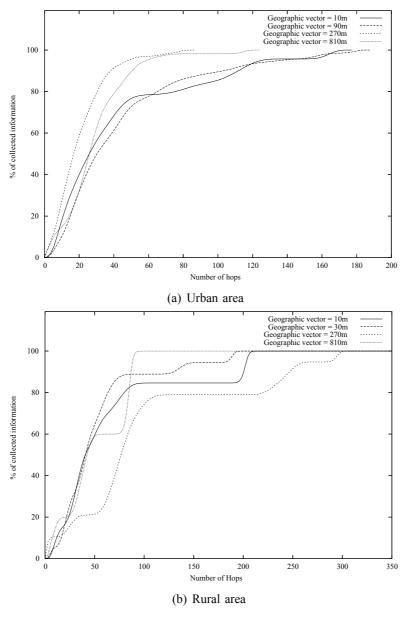
Fig. 10. Percentage of recovered information according to the number of hops.

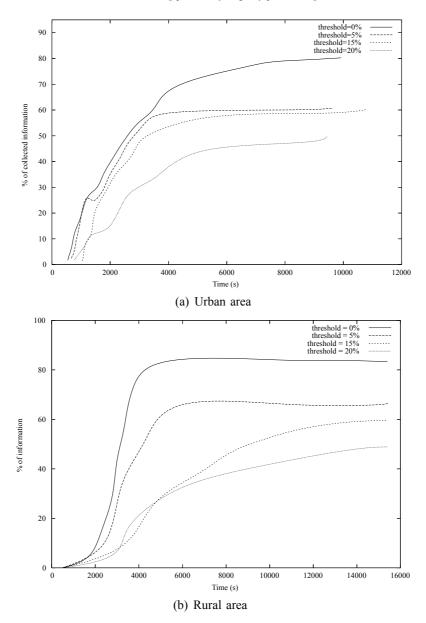(a) Urban area

(b) Rural area

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

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, as this leads to the fastest reaction possible when a wrong route is being taken. Using digital maps, the quantity of results collected is higher than using geographic vectors. Indeed, the percentage of results collected is close to 80% in the best case. The time needed is similar to the case using geographic vectors. So, most of the results are available at most one hour after the first dissemination of the query.

As we did for the implementation of GeoVanet on top of geographic vectors, we measured the number
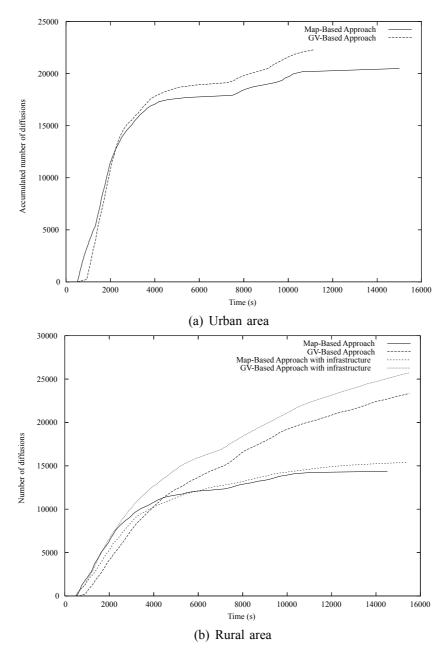
(a) Urban area

(b) Rural area

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

of messages diffused while routing query results towards the mailbox. Figure 12 shows a comparison between the results obtained for maps (*Map-based approach*) and vectors (*GV-based approach*). We considered and compared the best configurations. The exploitation of digital maps leads to a reduction of the number of messages diffused whatever the environment and independently of whether an infrastructure is deployed or not. This is due to the fact that loops are avoided and thus messages reach the destination with a lower number of hops.
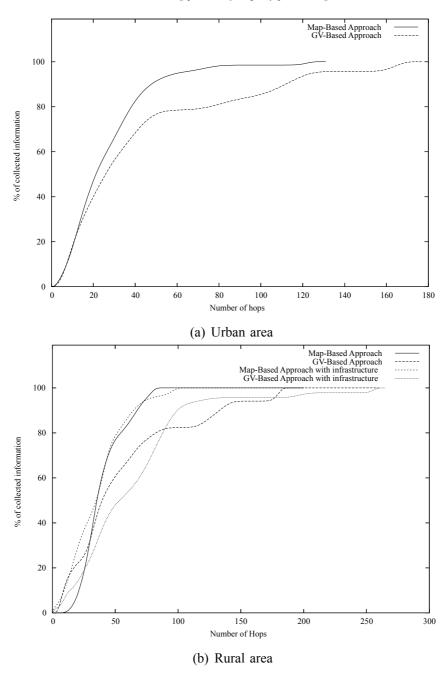
(a) Urban area

(b) Rural area

Fig. 13. Comparison of the percentage of recovered information according to the number of hops between maps and vectors.

Figure 13 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 confirms our assumptions regarding the reduction of the number of messages diffused observed in Fig. 12, whatever the environment.
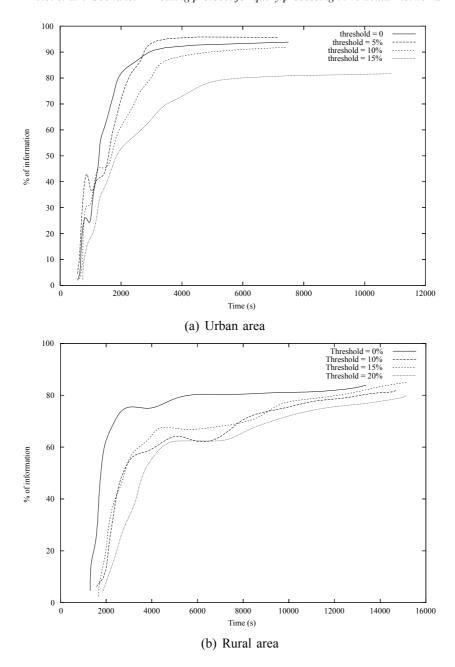
(a) Urban area

(b) Rural area

Fig. 14. Percentage of information collected over time by duplicating the carriers.

5.3.3. Evaluation of another possible optimization: Multiple carriers

Obviously, one solution to improve the percentage of results collected in the mailbox is to increase the number of message carriers at each hop. Figure 14 shows that, with the approach based on maps, up to 90% in the urban area and 80% in the rural area of the results available can be collected if the two closest vehicles driving in the right direction are used as carriers (instead of only one carrier as in our basic approach). However, this also leads to a severe increase in the network load, as depicted in Fig. 15, and
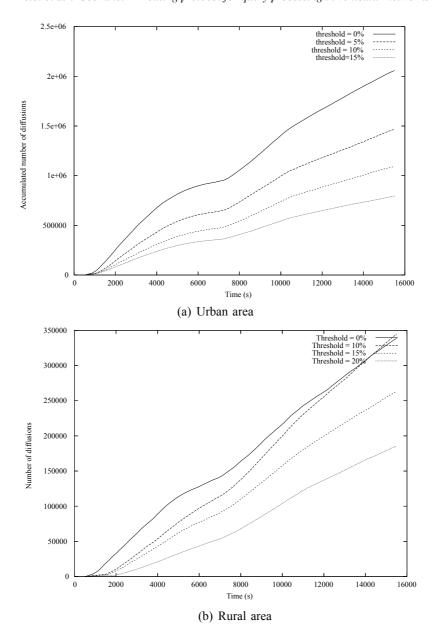
(a) Urban area

(b) Rural area

Fig. 15. Evolution of the number of messages diffused over time by duplicating the carriers.

to many duplicates received at the mailbox, especially in the urban area (where the density of vehicles is high). Therefore, this optimization should be used in case the density of vehicles is low and retrieving a very high percentage of results is needed.

6. Conclusions and future work

In this paper, we presented GeoVanet, an approach that supports query processing in vehicular networks

by providing an efficient and effective protocol for routing the results of the queries. GeoVanet therefore allows users to obtain the interesting data items within a bounded time. Through numerous simulations, we have proved the effectiveness of our solution in several environments and with different assumptions. The performance of Geovanet is high, especially when its implementation exploits the information available on digital road maps.

In the future, we plan to study in detail the impact of the number of mailboxes and their location. We would also like to study the potential interest of partitioning and distributing a single mailbox in different locations. Another interesting research direction will be to consider mixed environments where some vehicles may be equipped with digital maps and others are not.

Acknowledgments

References

[1] S. Androutsellis-Theotokis and D. Spinellis, A survey of peer-to-peer content distribution technologies, *ACM Computing Surveys* **36**(4) (2004), 335–371.

[2] F. Aurenhammer, Voronoi diagrams – a survey of a fundamental geometric data structure, *ACM Computing Surveys* **23** (1991), 345–405.

[3] P. Belanovic, D. Valerio, A. Paier, T. Zemen, F. Ricciato and C.F. Mecklenbrauker, On wireless links for vehicle-to-infrastructure communications, *IEEE Transactions on Vehicular Technology* **59**(1) (2011), 269–282.

[4] A. Bonifati, P.K. Chrysanthis, A.M. Ouksel and K.-U. Sattler, Distributed databases and peer-to-peer databases: past and present, *SIGMOD Record* **37**(1) (2008), 5–11.

[5] M. Burmester, E. Magkos and V. Chrissikopoulos, Strengthening privacy protection in VANETs. In *2008 IEEE International Conference on Wireless & Mobile Computing, Networking & Communication* (*WiMob'08*), pages 508–513. IEEE Computer Society, 2008.

[6] J. Cao, X. Feng, J. Lu and S.K. Das, Mailbox-based scheme for designing mobile agent communication protocols, *Computer* **35**(9) (2002), 54–60.

[7] A.A.V. Castro, G.D.M. Serugendo and D. Konstantas, Hovering information – self-organising information that finds its own storage, In *2008 IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing* (*SUTC'08*), pages 193–200. IEEE Computer Society, 2008.

[8] P. Cencioni and R.D. Pietro, A mechanism to enforce privacy in vehicle-to-infrastructure communication, *Computer Communications* **31** (2008), 2790–2802.

[9] N. Cenerario, T. Delot and S. Ilarri, A content-based dissemination protocol for VANETs: Exploiting the encounter probability, *IEEE Transactions on Intelligent Transportation Systems*, 2011. 12 pages, to appear, DOI: 10.1109/TITS.2011.2158821.

[10] 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, 2006.

[11] T. Delot, N. Cenerario and S. Ilarri, Vehicular event sharing with a mobile peer-to-peer architecture, *Transportation Research Part C: Emerging Technologies* **18**(4) (2010), 584–598.

[12] T. Delot, S. Ilarri, N. Cenerario and T. Hien, Event sharing in vehicular networks using geographic vectors and maps, *Mobile Information Systems* **7**(1) (2011), 21–44.

[13] T. Delot, S. Ilarri, M. Thilliez, G. Vargas-Solar and S. Lecomte, Multi-scale query processing in vehicular networks, *Journal of Ambient Intelligence and Humanized Computing*, 2011. 14 pages, to appear, published online (SpringerLink): 05 June 2011, DOI: 10.1007/s12652-011-0058-y.

[14] T. Delot, N. Mitton, S. Ilarri and T. Hien, Decentralized pull-based information gathering in vehicular networks using GeoVanet, In *12th International Conference on Mobile Data Management* (*MDM 2011*), pages 174–183. IEEE Computer Society, 2011.

[15] N. Dutta, A peer to peer based information sharing scheme in vehicular ad hoc networks, In *11th International Conference on Mobile Data Management* (*MDM 2010*), pages 309–310. IEEE Computer Society, 2010.

[16] M. Gerla and L. Kleinrock, Vehicular networks and the future of the mobile internet, *Computer Networks* **55**(2) (2011), 457–469.

[17] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi and K.-L. Tan, Private queries in location based services: anonymizers are not necessary, In *ACM SIGMOD International Conference on Management of Data* (*SIGMOD'08*), pages 121–132. ACM, 2008.

[18] A.M. Hanashi, I. Awan and M. Woodward, Performance evaluation with different mobility models for dynamic probabilistic flooding in MANETs, *Mobile Information Systems* **5**(1) (2009), 65–80.

[19] 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) (2007), 99–107.

[20] S. Ilarri, C. Bobed and E. Mena, An approach to process continuous location-dependent queries on moving objects with support for location granules, *Journal of Systems and Software* **84**(8) (2011), 1327–1350.

[21] S. Ilarri, E. Mena and A. Illarramendi, Location-dependent query processing: Where we are and where we are heading, *ACM Computing Surveys* **42**(3) (2010), 1–73.

[22] J. Jeong, S. Guo, Y. Gu, T. He and D.H. Du, TSF: Trajectory-based statistical forwarding for infrastructure-to-vehicle data delivery in vehicular networks. In *30th IEEE International Conference on Distributed Computing Systems* (*ICDCS 2010*), pages 557–566. IEEE Computer Society, 2010.

[23] E. Kang, M.J. Kim, E.-J. Lee and U.-M. Kim, DHT-based mobile service discovery protocol for mobile ad hoc networks, In *Fourth International Conference on Intelligent Computing* (*ICIC'08*)*: Advanced Intelligent Computing Theories and Applications – With Aspects of Theoretical and Methodological Issues*, volume 5226 of *Lecture Notes in Computer Science*, pages 610–619. Springer, 2008.

[24] E. Kulla, M. Hiyama, M. Ikeda, L. Barolli, V. Kolici and R. Miho, MANET performance for source and destination moving scenarios considering OLSR and AODV protocols, *Mobile Information Systems* **6**(4) (2010), 325–339.

[25] U. Lee, J.-S. Park, E. Amir and M. Gerla, FleaNet: A virtual market place on vehicular networks, *IEEE Transactions on Vehicular Technology* **59**(1) (2010), 344–355.

[26] I. Lequerica, P.M. Ruiz and V. Cabrera, Improvement of vehicular communications by using 3G capabilities to disseminate control information, *IEEE Network: The Magazine of Global Internetworking* **24**(1) (2010), 32–38.

[27] F. Li and Y. Wang, Routing in vehicular ad hoc networks: A survey, *IEEE Vehicular Technology Magazine* **2**(2) (2007), 12–22.

[28] T.-H. Lin, H.-C. Chao and I. Woungang, An enhanced MPR-based solution for flooding of broadcast messages in OLSR wireless ad hoc networks, *Mobile Information Systems* **6**(3) (2010), 249–257.

[29] C. Lochert, B. Scheuermann, M. Caliskan and M. Mauve, The feasibility of information dissemination in vehicular ad-hoc networks, In *Fourth Annual Conference on Wireless On demand Network Systems and Services* (*WONS'07*), pages 92–99. IEEE Computer Society, 2007.

[30] C. Maihöfer, A survey of geocast routing protocols, *IEEE Communications Surveys & Tutorials* **6**(2) (2004), 32–42.

[31] G.K. Mitropoulos, I.S. Karanasiou, A. Hinsberger, F. Aguado-Agelet, H. Wieker, H.-J. Hilt, S. Mammar and G. Noecker, Wireless local danger warning: Cooperative foresighted driving using intervehicle communication, *IEEE Transactions on Intelligent Transportation Systems* **11**(3) (2010), 539–553.

[32] Y. Morgan, Notes on DSRC & WAVE standards suite: Its architecture, design, and characteristics, *IEEE Communications Surveys & Tutorials* **12**(4) (2010), 504–518.

[33] 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.

[34] V. Namboodiri, M. Agarwal and L. Gao, A study on the feasibility of mobile gateways for vehicular ad-hoc networks, In *First ACM International Workshop on Vehicular Ad Hoc Networks* (*VANET'04*), pages 66–75. ACM, 2004.

[35] S.-Y. Ni, Y.-C. Tseng, Y.-S. Chen and J.-P. Sheu, The broadcast storm problem in a mobile ad hoc network. In *Fifth Annual ACM/IEEE International Conference on Mobile Computing and Networking* (*MobiCom'99*), pages 151–162. ACM, 1999.

[36] S. Olariu and M.C. Weigle, editors, *Vehicular Networks: From Theory to Practice*, Chapman & Hall/CRC, 2009.

[37] M. T. Özsu and P. Valduriez, *Principles of Distributed Database Systems*, chapter 1 "Introduction", pages 1–40. Springer, New York, NY, USA, 2011. Third edition. See Section 1.3 "Data Delivery Alternatives" (pages 5–7).

[38] M.T. Özsu and P. Valduriez, *Principles of Distributed Database Systems, Third Edition*, chapter "Peer-to-Peer Data Management", pages 611–655. Springer, 2011.

[39] M.G.C. Resende and P.M. Pardalos, *Handbook of Optimization in Telecommunications*, Springer, New York, NY, USA, 2006.

[40] R. Saqour, M. Shanuldin and M. Ismail, Prediction schemes to enhance the routing process in geographical GPSR ad hoc protocol, *Mobile Information Systems* **3**(3–4) (2007), 203–220.

[41] E. Spaho, L. Barolli, G. Mino, F. Xhafa, V. Kolici and R. Miho, Implementation of CAVENET and its usage for performance evaluation of AODV, OLSR and DYMO protocols in vehicular networks, *Mobile Information Systems* **6**(3) (2010), 213–227.
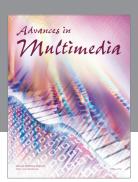
[42] L. Sweeney, k-anonymity: a model for protecting privacy, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **10** (2002), 557–570.

[43] R. Trillo, S. Ilarri and E. Mena, Comparison and performance evaluation of mobile agent platforms, In *Third International Conference on Autonomic and Autonomous Systems* (*ICAS'07*), pages 41–46. IEEE Computer Society, 2007.

[44] O. Urra, S. Ilarri, T. Delot and E. Mena, Mobile agents in vehicular networks: Taking a first ride, In *Eight International Conference on Practical Applications of Agents and Multi-Agent Systems* (*PAAMS 2010*), volume 70 of *Advances in Intelligent and Soft Computing*, pages 118–124. Springer, 2010.

[45] A.B. Waluyo, B. Srinivasan and D. Taniar, Research in mobile database query optimization and processing, *Mobile Information Systems* **1**(4) (2005), 225–252.

[46] A.B. Waluyo, D. Taniar, B. Srinivasan and W. Rahayu, An enhanced global index for location-based mobile broadcast services. In *International Conference on Advanced Information Networking and Applications* (*AINA 2010*), pages 1173–1180. IEEE Computer Society, 2010.

[47] 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.

[48] Q. Xu, H.T. 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.

[49] K. Xuan, D. Taniar, M. Safar and B. Srinivasan, Time constrained range search queries over moving objects in road networks, In *Eight International Conference on Advances in Mobile Computing and Multimedia* (*MoMM 2010*), pages 329–336. ACM, 2010.

[50] K. Xuan, G. Zhao, D. Taniar, W. Rahayu, M. Safar and B. Srinivasan, Voronoi-based range and continuous range query processing in mobile databases, *Journal of Computer and System Sciences* **77**(4) (2011), 637–651.

[51] 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.

[52] R. Zimmermann, W.-S. Ku and H. Wang, Spatial data query support in peer-to-peer systems. In *28th Annual International Computer Software and Applications Conference* (*COMPSAC'04*) – *Workshops and Fast Abstracts*, volume 2, pages 82–85. IEEE Computer Society, 2004.

Thierry Delot (http://www.univ-valenciennes.fr/ROI/SID/tdelot/) is an Associate Professor at the University of Valenciennes since 2002. He is a member of the LAMIH laboratory (FRE CNRS 3304). He got a PhD in Computer Science at the University of Versailles in 2001. His research interests mainly concern mobile data management and query processing. Since 2007, Thierry is particularly interested in vehicular ad hoc networks.

Nathalie Mitton received the MSc and PhD. degrees in Computer Science from the INSA de Lyon in 2003 and 2006 respectively. She is currently an INRIA full researcher. Her research interests are mainly focused on theoretical aspects of self-organization, self-stabilization, energy efficient routing and neighbour discovery algorithms for wireless sensor networks as well as RFID middlewares. She is involved in several program and organization committees such as MASS 2011, AdHocNow 2011, MASS 2011, WiSARN and WiSARN-Fall 2011&2010, WWASN 2010&2009, SANET 2008&2007.
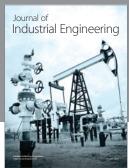
Sergio Ilarri received his B.S. and his PhD in Computer Science from the University of Zaragoza in 2001 and 2006, respectively. Now, he is an Associate Professor in the Department of Computer Science and Systems Engineering. For a year, he was a visiting researcher in the Mobile Computing Laboratory at the Department of Computer Science at the University of Illinois in Chicago, and he has also cooperated (through several research stays) with the University of Valenciennes and with IRIT in Toulouse. His research interests include data management issues for mobile computing, vehicular networks, mobile agents, and the Semantic Web.

Thomas Hien is the holder of a Master's degree in Computer Science from the University of Valenciennes, France, obtained in 2009. Since then, he is a member of the LAMIH laboratory (FRE CNRS 3304) devoted to the VESPA project as a research engineer. His work mainly concerns the evaluation of inter-v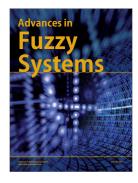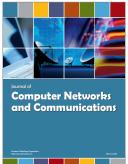ehicle communication systems in realistic environments.