RESEARCH　　　　　　　　　　　　　　　　　　　　　　　　　　**Open Access**

Efficient load balancing and QoS-based location aware service discovery protocol for vehicular ad hoc networks

Kaouther Abrougui[1,2*], Azzedine Boukerche[1,2] and Hussam Ramadan[3]

Abstract

Service discovery studies in vehicular networks that guarantee QoS requirements to service requesters are very important. To the best of authors' knowledge, none of the existing service discovery protocols in vehicular networks have been provided in the literature that guarantee QoS to service requesters and to the vehicular network. For efficient service discovery in vehicular networks, it is very important to provide users with services that suit better to their requests while balancing the load on service providers. Moreover, for service discovery protocol integrated with routing protocols, it is important to guarantee load balancing on routing paths between service providers and service requesters. In this article, we present a QoS aware location-based service discovery protocol for vehicular networks. Our protocol guarantees load balancing on service providers, and routing paths between service providers and service requesters. It permits also to choose service providers and routing paths between service providers and service requesters that satisfy some performance attributes specified by service requesters. We present our QoS aware protocol, prove its correctness, report on its performance evaluation, and discuss our experimental results we have obtained using realistic scenarios.

1. Introduction

Recently, we have noticed an increasing interest in the study of service discovery protocols in vehicular networks. To the best of authors' knowledge, most of the proposed service discovery schemes [1,2] do not take into account the load balancing and the QoS requirements features. In this article, we propose a load balancing, QoS-aware and location-based service discovery protocol for vehicular networks. The location-based service discovery protocol (LocVSDP) [3] permits the discovery of location-aware and time-sensitive services in Vehicular Networks. It integrates service information into the network layer and uses diverse channels. Our proposed load balancing and QoS-based location aware discovery protocol (QoSLocVSDP) would permit to a service requester to connect to the appropriate less congested service provider through the less congested routing path. It permits also to satisfy some QoS requirements specified in clients requests.

The remainder of this article is organized as follows: Section 2 describes the related studies. Section 3 presents the system model. Section 4 describes our proposed load balancing and QoS-based service discovery protocol. Section 5 presents the proof of correctness. Section 6 presents the message complexity computation and the bandwidth usage. Section 7 describes the performance evaluation of the proposed technique. Finally, Section 8 concludes the article.

2. Related studies

Very little study has been done on the design of load balancing and QoS-based service discovery protocol in vehicular networks [4]. In [4], a server localization architecture for heterogenous wireless sensor networks is presented. The authors present a two-tier server architecture. The first tier comprises the servers and the second one comprises standard sensor nodes acting as server locators. Their localization and communication infrastructure uses Q-NiGHT [5], which is an improved version of the geographic hash table (GHT) [6]. The presented architecture permits the load balancing of queries to the servers and it is tolerant to server faults.

* Correspondence: abrougui@site.uottawa.ca
[1]DIVA Research Center, University of Ottawa, Ottawa, ON K1N 6N5, Canada
Full list of author information is available at the end of the article

Springer

In order to guarantee load balancing in their protocol, all servers that provide the same service are registered at the same server locator. Service requests are processed using a specific strategy; the server is chosen either randomly or in a round robin way. A few number of recent works in ad hoc and vehicular networks have focused on migratory services in order to accomplish the location-awareness in their protocols [7-9]. Some recent projects have tried to deploy real services over mobile ad hoc and vehicular networks [10-12]. Only few distributed location-based service discovery protocols have been designed with the characteristics of vehicular networks in mind [1,2]. Dikaiakos et al. [1] proposed location-aware services over vehicular ad hoc networks using car-to-car communication: Their protocol provides time-sensitive information about the traffic conditions and the available services on the roadside. Klimin et al. [2] proposed a hybrid service discovery protocol in ad hoc vehicular networks. Their approach combines the proactive dissemination of advertisement messages and the reactive propagation of discovery requests. Their protocol is based on the geocast addressing of control messages.

3. System model

Our vehicular network system comprises two types of components: roadside routers (RRs), and road vehicles (RVs). *RRs* form a wireless backbone of the vehicular network. We suppose that vehicles move on a two-dimensional plane. Lanes are L m long and l m large. RRs are located on the intersection of straight lanes and perpendicular lanes. The total number of RRs in our model is nRR. The vehicle arrival process in our model follows a poisson process with an arrival rate (λ). We define the density (*Density*) of the network as the number of vehicles per meter square in the vehicular network. The vehicular network is divided into road sections that have the same density than the network. The average speed of a vehicle is S. We assume the random variable N that defines the number of vehicles in a road section of surface x. The probability that there are n vehicles in a road section of surface (x) is given by

$$P\left[N(x) = n\right] = \frac{\left(\frac{x}{S}\lambda\right)^n}{n!}e^{-\lambda\frac{x}{S}} \qquad (1)$$

4. The load balancing and QoS-based location discovery protocol

In the following, we describe our proposed load balancing and QoSLocVSDP. Then, we discuss the features of the QoSLocVSDP and we compare it to the original LocVSDP.

4.1. Protocol description

QoSLocVSDP has four main phases: *QoS-based service advertisement phase*, *QoS-based service request propagation phase*, *QoS-based leader election and service reply generation phase*, and *QoS-based service reply propagation phase*.

4.1.1. QoS-based service advertisement phase

In the QoS-based service advertisement phase, service providers send QoS-based advertisement messages to their neighboring road components (RRs and RVs). A QoS-based advertisement (QoSAdv) message contains both routing and service information. The service provider adds its correspondent QoS attributes to the QoSAdv message. This latter could be intercepted by vehicles or by neighboring RR. Each RR that receives the QoSAdv message adds the service information to its service table and the routing information to its routing table. If the service exists already in the service table or the route exists already in the routing table, then the service table or the routing table is updated, respectively.

4.1.2. QoS-based service request propagation phase

A QoS-based request (QoSReq) message is sent by a service requester that wants to find a service provider in a desired location called region of interest (RI) and with QoS requirements. The QoSReq message is propagated until it reaches the RI. It contains both routing information and discovery information. The QoSReq message is generated and sent by the service requester. Only RRs and vehicles that are closer to the RI than the sending road component, will propagate the QoSReq message. Once the QoSReq message reaches the RI, the QoS-based leader election phase starts.

4.1.3. QoS-based leader election and service reply generation phase

In this phase, a leader is elected in the RI. When an RR RRi receives a QoSReq message, it generates a QoS-based election (QoSElec) message and broadcasts it inside the RI. Every RR that receives a QoSElec message generates its own election message and broadcasts it to its neighbors inside RI. In the QoSElec message, the RR RRi adds its distance to the center of RI. If RRi has been elected as leader for another QoS-based service request, it adds the radius of RI to the distance to RI and sends it in the QoSElec message, such that it will not be elected as leader again until all RRs inside RI have been elected at least once as leaders. After the exchange of QoSElec messages between RRs inside RI, every RR chooses its parent in the RI as the RR that has the smallest value of the distance to the origin of RI. After the exchange of parent messages a spanning tree inside the RI is constructed. The root of the spanning is elected as leader. QoS-based local reply (QoSLocalRep) messages are sent by every RR to its parent in the spanning tree.

A QoSLocalRep message includes all the service providers that satisfy the QoS-based service request sent by the service requester. The leader RR will receive the information about all the service providers inside the desired RI that satisfy the requester. At this stage, it generates the QoS-based reply message that will be propagated to the service requester.

4.1.4. QoS-based service reply propagation phase
The QoS-based service reply (QoSRep) message is generated by the leader RR in the desired RI. It is propagated toward the service requester. Since the service requester could be moving during the discovery process, the QoSRep message is geocast toward the service requester. Only road components that are closer to the expected region of the service requester forward the QoSRep message. Moreover, only road components that satisfy the QoS requirements specified by the requester can forward the QoSRep message. This is because the requester can use the reverse path received in a service reply message to communicate with the chosen service provider.

4.2. QoSLocVSDP discussion and comparison to the original LocVSDP
Load balancing and QoS are very important features for service discovery in vehicular networks. In fact, in vehicular networks that have multiple service providers of the same service, it is very important to balance the load between the different service providers while satisfying requesters performance attributes in order to guarantee a good service quality.

In our service discovery protocol, we design novel techniques for a load balanced and QoS aware service discovery protocol in vehicular networks. First, we guarantee that service requests are handled equitably by service providers in the desired RI. For this purpose, it is not necessary that the service provider located closer to the center of the RI that satisfies the service request, but any service provider located inside the RI and that is less loaded than the other service providers in the same RI can handle the service request. Second, since our protocol is integrated into a routing protocol, we guarantee that routes between service providers and service requesters are not congested. For this purpose, we proposed a novel technique that guarantees load balancing on RRs in the vehicular network. In addition to guaranteeing load balancing among the different components of the vehicular network, our QoS-aware discovery protocol finds service providers and routing paths between service providers and service requesters that satisfy some attributes specified in the user request. Such attributes could be for example streaming ratio (for example, in a streaming service, the user specifies the streaming rate required...).

In the following, we discuss our proposed techniques for load balancing and QoS aware service discovery protocol in vehicular networks.

4.2.1. Load balancing on service providers
We define the load on a service provider SP as the number of requests handled by the SP during the simulation time. In the LocVSDP [3], the service requester receives in the service reply the information of the closest requested service provider to the center of the desired RI. If many service requesters specify the same RI, then the same service provider SP_0 would be returned to these service requesters, even if there are other service providers located in the desired RI. Thus, the service provider SP_0 would be overloaded while the other service providers located in the RI are underloaded, which can affect the performance of the overloaded service provider. In order to overcome this weakness and balance the load on service providers in the vehicular network, we proposed the following load balancing technique: When a service client sends its request to the desired RI, the RRs inside the RI compute a spanning tree with a unique leader RR. Then all RRs send their local replies to the leader of the spanning tree starting from the leaves. In their local replies, every RR adds the load on each service provider in its range. The leader RR receives all the information about SPs including their loads. In the reply returned to the service requester, the leader sends the information of service providers ranked from the less loaded one to the most loaded one. Then the service requester connects to the service provider that satisfies the desired QoS requirements. The chosen service provider updates its load and acknowledges its neighboring RRs about its new load that would be considered for future requests.

4.2.2. Load balancing on RRs leaders inside RI
In the LocVSDP protocol [3], a leader is elected in the RI specified in a user request. The elected leader RR is the closest RR to the center of the RI. If many requesters specify the same RI then the load on the same elected leader for many requests would be high. In order to balance the load on RRs in the RI we proposed the following mechanism: an RR RR_0 that participates in an election process and is elected as leader stores the request number and the information regarding the RI (i.e., center and range). Besides, it computes its distance to the center of the RI *DistanceRI*, it stores this value in the variable *realDistanceRI* and initiates a variable *fakeDistanceRI* with the value of *DistanceRI*. If RR_0 participates in another election process for the same RI it updates the value of *fakeDistanceRI* such that *fakeDistanceRI = fakeDistanceRI + range* where *fakeDistanceRI* is the previously computes fake *DistanceRI*; and *range* is the range of the RI and is a fixed value. RR_0 sends *fakeDistanceRI* in the message *election_IDReq_msg*. Thus,

RR_0 will not be elected again as leader for a new service request in the same RI. Thus, the second closest RR to the center of the RI would be elected as leader. Similarly, the new elected leader executes our mechanism such that it will not be elected as leader for the next similar user request. Since *fakeDistanceRI* is computed by adding a fixed value to the real *DistanceRI*, our process guarantees that all RRs located in the same RI are elected as leaders at least once before anyone of them is elected again as leader for the second time. As a result, our technique guarantees load balancing on RRs in any RI requested by many service clients.

4.2.3. Load balancing on routing path between service providers and service requesters

In the LocVSDP protocol [3], the routing path established between a service requester and its correspondent service provider is the first path from which the elected leader has received the first request message or parent message. In our QoS and load balancing LocVSDP, this first path is not necessarily chosen. If the first path is congested, our load balancing and QoS-based protocol is able to generate an alternative path less congested than the first path in order to establish an efficient communication between the service provider and the service requester. In the following, we explain how our load balancing and QoS-based LocVSDP protocol generates the less congested path while propagating the service reply.

We define the cost of a routing path as the sum of loads on each component of the routing path. In order to generate the less congested path between a service provider and a service requester, we implement the following technique: after a service request is propagated to the desired RI and handled by the elected leader. A service provider is chosen and a unique reply is generated. In the LocVSDP, the unique reply is unicast to the service requester through the first path returned. In our QoS-based LocVSDP, the unique reply is geocast to the service requester from the elected RR leader. An intermediate *RR* forwards the received reply message if and only if, (i) the *RR* is closer to the service requester than the RR from which it has received the reply message and; (ii) the cumulated cost of the routing path of a request *ReqID* until the current RR is less than a previously cumulated cost of a routing path for the same request *ReqID*.

This way, it is guaranteed that the reply message is not flooded in the vehicular network and that the routing path returned to the service requester is the less congested one among all the possible routing paths between the service provider and the service requester. If the current intermediate RR has to propagate the reply message, it adds its ID to the reply message, its capacity, its load, and the link load between itself and the previous RR. The service requester sends a connection message to the service provider through the less congested routing path including its eventual load. Thus, each intermediate component updates its load accordingly.

4.2.4. Service provider quality requirement

Many service providers providing the same service could have different characteristics and quality (such as streaming rate, etc.). Service requesters could specify in their requests the quality of the service that they require using different attributes. In our proposed QoS-based service discovery technique, the quality attributes specified by a service request are handled at the elected RR leader. After the reception of all the local service replies from RRs inside the RI, the leader RR can determine the appropriate service provider that meets all the quality attributes defined by the service requester while maintaining the load balancing on service providers.

4.2.5. Intermediate components quality requirement

The intermediate components (RRs and links) between a service provider and a service requester can have different characteristics and quality. A service requester could require a desired quality in the routing path to a service provider. In our QoS-based LocVSDP, we handle the requester requirement during the reply propagation and routing path generation phase. Only routing paths that satisfy the QoS requirements are geocast in the vehicular network and are returned to the service requester.

5. Proof of correctness

In the following, we prove the correctness and completeness of our proposed QoSLocVSDP. Before we proceed further, we present the following lemmas that prove that our proposed QoSLocVSDP guarantees the load balancing of service queries among service providers and RRs, and that our QoSLocVSDP satisfies the QoS requirements imposed by the service requester in the different road components.

In the rest of this section, we assume that a unique cluster of RRs can be formed inside a desired RI, because the radius of this latter should not be very extended in order to preserve the proximity of a service provider to the center of the RI.

Lemma 1. *Load balancing on service providers inside the RI: Our QoSLocVSDP permits to balance the load on service providers while satisfying location-based service requests inside a predetermined RI.*

Proof: Assuming that RRs inside RI are connected, when a service request reaches the RI, a unique spanning tree comprising the RRs inside RI is constructed. The root of the spanning tree is elected as the leader RR. This latter receives all the local reply messages from its children. A local reply message comprises the information of service providers satisfying the request. It includes also the load on the returned service providers.

Since the unique elected leader RR_L inside the RI will receive all the local replies from RRs inside RI, it will be able to determine the appropriate less loaded service providers that satisfy the service request inside RI. The RR_L sends an aggregated service reply message to the service requester including the appropriate service providers sorted from the less loaded to the more loaded service provider. At the reception of the service reply message by the service requester, this latter will chose the less loaded service provider that satisfies other QoS requirements.

Lemma 2. *Load balancing on RRs inside a predetermined RI: Our proposed QoSLocVSDP permits to balance the load on RRs while electing a leader in the RI specified by a service requester.*

Proof: In order to balance the load of the RRs inside RI while processing the drivers' requests, an RR is elected again as leader in a predetermined RI only if all the RRs inside RI have been already elected as leaders. This is because in the proposed QoSLocVSDP, a leader RR is elected as leader if it has the smallest value of *fakeDistanceRI* when constructing the spanning tree for a specific request. In our proposed mechanism, once an RR inside RI is elected as leader and processes a driver's request, it increases the value of its *fakeDistanceRI* by the radius' value of RI. Thus, the previously elected leader RR will have the largest value of *fakeDistanceRI* in the next service request processing, and will not be elected as leader again until all the RRs inside RI are chosen as leader for different service requests in the same RI. Consequently, the load of service requests processing is balanced among the different RRs inside RI.

Lemma 3. *Load balancing on RRs in the routing path between the service requester and the service provider: Our QoSLocVSDP permits to balance the load of RRs in the routing path between a service provider and a service requester.*

Proof: In the proposed QoSLocVSDP, the aggregated service reply message is geocast to the requesting vehicle. Thus, there are many paths that would be returned to the requester. An intermediate RR that receives the aggregated reply message more than once, forwards it only if the cumulated load of the routing path is less than a previously cumulated load through the current RR. The service requester receives the service reply message through multiple paths with indication of the cumulated load on each path. It chooses the less loaded path while taking into account other QoS requirements.

Lemma 4. *QoS of a service provider: a driver requester can specify some QoS requirements in the eventual service provider that are included in the service request. The elected leader takes into account these requirements while generating the aggregated service reply.*

Proof: In the proposed QoSLocVSDP only service providers with the appropriate QoS requirements are included in the aggregated reply message. Thus, the service requester will receive the service providers that satisfy the predetermined QoS requirements in the returned service reply message.

Lemma 5. *QoS guarantee of road components between the service requester and the service provider: Our proposed QoSLocVSDP guarantees some QoS requirements in the routing path between the service provider and the service requester.*

Proof: In our proposed QoSLocVSDP, and during the propagation of the aggregated reply message to the service requester, only the routing paths that satisfy the requester requirements are returned.

Theorem 1. *(QoSLocVSDP Correctness) Our QoSLocVSDP finds the less loaded service provider with the QoS requirements that satisfy the requesting vehicle in the desired RI in a finite time if the requested service is provided in the RI in the VANet.*

Proof: In our QoSLocVSDP, a QoS-based request message sent by a requester is propagated until the RI. In the RI, all RRs participate in the election of a leader and the construction of a spanning tree. Then, every RR sends a local reply to its parent in the spanning tree with its knowledge about the service providers that satisfy the QoS required by the requester as proven in Lemma 4. The leader RR collects all the information from the RRs inside RI and generates a unique service reply message. It includes in the reply message all service providers with the requested QoS requirements. Service providers are sorted from the less loaded one to most loaded one in terms of number of handled service requests. Thus, all service providers inside RI with the requested QoS requirements are returned to the service requester in a finite time. The service requester will be able to chose the appropriate service provider.

Theorem 2. *(QoSLocVSDP Completeness) Assume that there are many service providers of the requested service in the desired RI, our proposed QoSLocVSDP permits to the requester to receive the information of all the service providers in the RI ranked from the less loaded service provider to the most loaded, and with the specified QoS requirements information in finite time.*

Proof: In order to prove the completeness of our proposed QoSLocVSDP, we need to prove that there are no deadlocks in our scheme.

Assuming that the vehicular network is connected, the QoS-based request message is propagated towards the RI. When it reaches the RI, QoS-based election messages are generated and exchanged between the RRs inside RI. For each service request, a connected spanning tree is constructed inside RI and the root of the spanning tree is elected as leader to process the current

service request. The idea of spanning tree construction is based on timers. Thus, a leader is elected in the RI after a predetermined period of time. The elected leader will wait a predetermined period of time to receive local reply messages from its neighbors, and then generates the unique reply message to be sent to the requester. Again, the idea of generation of the unique service reply is based on timers and will not lead to a deadlock. The QoS-based service reply message will be propagated using geographical information until it reaches the destination. Thus, assuming that the VANet is connected, the QoS-based reply message will reach its destination with the required information of service providers.

6. Complexities computations and bandwidth usage

In the following, we compute the complexity of our proposed QoSLocVSDP, and its bandwidth usage. We use the abbreviation described in Table 1.

6.1. Message complexity of QoSLocVSDP

The message complexity of our QoSLocVSDP is the summation of the number of QoS-based service providers advertisement messages, the number of QoS-based request messages, the number of QoS-based election messages, the number of QoS-based local reply messages, and the number of QoS-based reply messages.

Lemma 6. *The number of QoS-based advertisement messages (nTotal_QoSAdv) during the simulation time T is:*

$$nTotal\_QoSAdv = \sum_{l=1}^{nQoSSP} \left(\frac{T}{\lambda_{QoSadv\_l}} \right) \qquad (2)$$

Proof: In our QoSLocVSDP, service providers advertise themselves by sending QoS-based service advertisement messages every advertisement period (λ_{QoSAdv}) to the neighboring RRs and vehicles. Thus, the number of QoS-based service advertisement messages n_QoSAdv sent by 1 service provider during the simulation period T is:

$$n\_QoSAdv = \frac{T}{\lambda_{QoSAdv}} \qquad (3)$$

Consequently, the total number of QoS-based advertisement messages sent by all the service providers $nQoSSP$ in the Vanet during the simulation period T is:

$$nTotal\_QoSAdv = \sum_{l=1}^{l=nQoSSP} \left(\frac{T}{\lambda_{QoSAdv\_l}} \right) \qquad (4)$$

Lemma 7. *The total number of QoS-based request messages (nTotal_QoSReq) for nQoSReqs QoS-based service requests during the simulation period T is:*

$$nTotal\_QoSReq = \sum_{i=1}^{i=nQoSReqs} Area\_QoSReq\_i \times (\rho + \rho v)$$

Proof: In our proposed QoSLocVSDP, a service requester generates a service request message with its QoS requirements *(QoSReq_msg)* and sends it to its neighboring RRs and vehicles toward the desired RI. The RRs and vehicles, that are closer to the RI than the sending road component and that receive the message for the first time, forward the *QoSReq_msg*. This way, the propagation of the QoS-based request message is controlled, not redundant, and not flooded in the VANet. Consequently, the number of QoS-based request messages is equal to the number of RRs and vehicles inside the forwarding zone *(QoS_Fzone)*. Knowing the densities ρ and ρ_V in terms of RRs and vehicles, respectively, and knowing the area of the forwarding zone *Area_QoSReq*, we can determine the number of QoS-based service request messages *(nQoSReq)* forwarded during one service request as:

$$n\_QoSReq = Area\_QoSReq \times (\rho + \rho v)$$

Thus for a total number of nQoSReqs requests during the simulation period T, the total number of QoS-based service request messages is:

$$nTotal\_QoSReq = \sum_{i=1}^{i=nQoSReqs} Area\_QoSReq\_i \times (\rho + \rho v)$$

Table 1 Variables description

Variable name	Variable description
nQoSSP	number of QoS-based service providers
λ_{QoSAdv}	advertisement period
T	the simulation duration
nQoSReqs	number of QoS-based request messages
Area QoSReq	The surface of the request forwarding zone
ρ	the density of the VANet in terms of number of RRs per meter square.
ρV	The density of the VANet in terms of vehicles per meter square.
Area RI	The surface of the RI.
Area_QoSRep	The surface of the reply zone.

Lemma 8. *The total number of QoS-based election messages nTotal_QoSElec for nQoSReqs service requests is:*

$$nTotal\_QoSElec = \sum_{i=1}^{i=nQoSReqs} (Area\_RI\_i \times \rho) \qquad (5)$$

Proof: In the proposed QoSLocVSDP, the election messages are exchanged between the RRs inside the RI specified by the service requester. Every election round, the less loaded RR is elected as leader to process the current service request. The load on RRs is expressed in terms of number of processed requests. Every RR inside the RI sends an election message to inform its neighbors about its load. Thus, the number of QoS-based election messages in every round is equal to the number of RRs inside the RI. Thus, knowing the area of RI *Area_RI* and the density in terms of RRs ρ, we can determine the number of QoS-based election messages as:

$$n\_QoSElection = Area\_RI \times \rho$$

During the simulation period T, the total number of QoS-based election messages (*nTotal_QoSElec*) for a number *nQoSReqs* of requests is:

$$nTotal\_QoSElec = \sum_{i=1}^{i=nQoSReqs} (Area\_RI\_i \times \rho) \qquad (6)$$

Lemma 9. *The number of QoS-based local reply messages nTotal_QoSlocalRep for nQoSReqs QoS-based service requests is:*

$$nTotal\_QoSLocalRep = \sum_{i=1}^{i=nQoSReqs} Area\_RI\_i \times \rho \qquad (7)$$

Proof: In our proposed QoSLocVSDP, a QoS-based local reply message is generated and sent by every RR inside RI to its selected parent. Thus, the number of QoS-based local reply messages for one service request is equal to the number of RRs inside RI. Knowing the area of RI (*Area_RI*) and its density in terms of RRs (ρ), we can determine the number of QoS-based local reply messages for one service request as:

$$n\_QoSLocalRep = Area\_RI \times \rho \qquad (8)$$

Consequently, the total number of QoS-based local reply messages (*nTotal_QoSLocalRep*) for a number of *nQoSReqs* requests is:

$$nTotal\_QoSLocalRep = \sum_{i=1}^{i=nQoSReqs} Area\_RI\_i \times \rho \qquad (9)$$

Lemma 10. *The total number of QoS-based reply messages nTotal_QoSRep for nQoSReqs QoS-based service requests during the simulation period T is:*

$$nTotal\_QoSRep = \sum_{i=1}^{i=nQoSReqs} Area\_QoSRep\_i \times (\rho + \rho v)$$

Proof: In our proposed QoSLocVSDP, QoS-based reply messages are sent from the elected leader or from the requested service providers to the requesting vehicle. Only the service providers or the routing paths that satisfy the QoS requirements specified by the service requesters are returned in the QoS-based service reply messages. The number of QoS-based service reply messages corresponding to a service request is the number of forwarding road components in the QoS-based reply forwarding zone. In the worst case, all the road components inside the forwarding zone will forward the QoS-based reply messages. Thus, knowing the area of the QoS-based reply forwarding zone (*Area_QoSRep*), and the densities ρ and ρ_V of the VANet in terms of RRs and vehicles, respectively, we can determine the number of QoS-based reply messages for one service request as:

$$n\_QoSRep = Area\_QoSRep \times (\rho + \rho v))$$

As a result, for a number *nQoSReqs* of requests, the total number (*nTotal_QoSRep*) of QoS-based reply messages is:

$$nTotal\_QoSRep = \sum_{i=1}^{i=nQoSReqs} Area\_QoSRep\_i \times (\rho + \rho v)$$

Theorem 3. *The total number of messages in our QoSLocVSDP during the simulation period T is:*

$$MC\_QoSLocVSDP = \sum_{l=1}^{nQoSSP} \left(\frac{T}{\lambda_{QoSAdv}}\right) + \sum_{i=1}^{i=nQoSReqs} (Area\_QoSReq\_i + Area\_QoSRep\_i) \times$$
$$(\rho + \rho v) + 2 \times (Area\_RI\_i \times \rho)$$

Proof: The total number of messages *MC_QoSLocV SDP* of our QoSLocVSDP during the simulation period T comprises: the total number of QoS-based advertisement messages, the total number of QoS-based request messages, the total number of QoS-based election messages, the total number of QoS-based local reply messages and the total number of QoS-based reply messages. We deduce our theorem from Lemmas 6-10.

6.2. Bandwidth usage computation of QoSLocVSDP

In the following, we present the computation of the bandwidth usage of QoSLocVSDP. The table in Figure 1 presents the different message types in our proposed QoSLocVSDP, their main fields descriptions and their sizes.

Lemma 11. *The bandwidth usage required for the advertisement of the QoS-based service advertisement messages during the simulation period is:*

$$nTotal\_QoSAdvBand = nTotal\_QoSAdv \times size\_QoSAdv$$

Message abbreviation	Message description	Fields	Size (bits)
QoSAdv	**QoS based advertisement message**	**Source @, destination @, packet ID, packet ID, service description, lifetime, service location, QoS attributes.**	**40**
QoSReq	**QoS based service request message**	**Source @, destination location information, sender @, TTL, packet ID, packet Type, service requested, RI, distance to RI, QoS requirements.**	**40**
QoSElec	**QoS based election message**	**Source @, destination @, Packet ID, packet type, distance to RI, fake distance to RI, requested service, QoS requirements.**	**20**
QoSLocalRep	**QoS based local reply message**	**Source @, destination @, Packet ID, packet type, service table, corresponding QoS values.**	**20**
QoSRep	**QoS based reply message**	**Source @, destination location, Packet ID, packet type, service table, corresponding QoS values.**	**60**

Figure 1 QoSLocVSDP messages description.

Proof: The bandwidth required for sending the QoS-based advertisement messages ($nTotal\_QoSAdvBand$) be the product of the total number of QoS-based advertisement messages ($nTotal\_QoSAdv$) and the size of a QoS-based advertisement message ($size\_QoSAdv$) as computed below:

$$nTotal\_QoSAdvBand = nTotal\_QoSAdv \times size\_QoSAdv$$

Lemma 12. *The total bandwidth nTotal_QoSReqBand required for the propagation of the QoS-based request messages during the simulation period T is:*

$$nTotal\_QoSReqBand = nTotal\_QoSReq \times size\_QoSReq$$

Proof: The total bandwidth required for all the request messages ($nTotal\_QoSReqBand$) during the simulation period T is the product of the total number of QoS-based request messages ($nTotal\_QoSReq$) and the size of a QoS-based request message ($size\_QoSReq$).

$$nTotal\_QoSReqBand = nTotal\_QoSReq \times size\_QoSReq$$

Lemma 13. *The bandwidth nTotal_QoSElecBand required for the propagation of the QoS-based election messages during the simulation period T is:*

$$nTotal\_QoSElecBand = nTotal\_QoSElec \times size\_QoSElec.$$

Proof: The bandwidth required for the propagation of the QoS-based election messages is the product of the total number of the QoS-based election messages ($nTotal\_QoSElec$) and the size of a QoS-based election message ($size\_QoSElec$):

$$nTotal\_QoSElecBand = nTotal\_QoSElec \times size\_QoSElec.$$

Lemma 14. *The bandwidth nTotal_QoSLocalRepBand required for the propagation of the QoS-based local reply messages for nQoSReqs requests is:*

$$nTotal\_QoSLocalRepBand = nTotal\_QoSLocalRep \times size\_QoSLocalRep.$$

Proof: The total bandwidth ($nTotal\_QoSLocalRepBand$) required for the propagation of the QoS-based local reply messages is the product of the total number

of QoS-based local reply messages ($nTotal\_QoSLocalR$) and the size of a QoS-based local reply message ($size\_QoSLocalRep$) as computed below:

$$nTotal\_QoSLocalRepBand = nTotal\_QoSLocalRep \times size\_QoSLocalRep.$$

Lemma 15. *The total bandwidth nTotal_QoSRepBand required for the propagation of the QoS-based reply messages during the simulation period T is the product of the total number of QoS-based reply messages nTotal_QoSRep and the size of a QoS-based reply message size_QoSRep:*

$$nTotal\_QoSRepBand = nTotal\_QoSRep \times size\_QoSRep$$

Proof: The total bandwidth $nTotal\_QoSRepBand$ required for the propagation of the QoS-based reply messages during the simulation period T is the product of the total number of QoS-based reply messages $nTotal\_QoSRep$ and the size of a QoS-based reply message $size\_QoSRep$:

$$nTotal\_QoSRepBand = nTotal\_QoSRep \times size\_QoSRep$$

Theorem 4. *The total bandwidth usage Band_QoSLocV SDP in our QoSLocVSDP during the simulation period T is:*

$$Band\_QoSLocVSDP = nTotal\_QoSAdv \times size\_QoSAdv + nTotal\_QoSReq \times size\_QoSReq + nTotal\_QoSElec \times size\_QoSElec + nTotal\_QoSLocalRep \times size\_QoSLocalRep + nTotal\_QoSRep \times size\_QoSRep$$

Proof: The total bandwidth usage of our QoSLocVSDP during the simulation period T comprises: the total bandwidth usage of the QoS-based advertisement messages, the total bandwidth usage of the QoS-based request messages, the total bandwidth usage of the QoS-based election messages, the total bandwidth usage of the QoS-based local reply messages and the total bandwidth usage of the QoS-based reply messages. We deduce our theorem from Lemmas 11-15.

7. Performance evaluation

In this section, we present the simulation we have conducted to evaluate the performance of our load balancing and QoS-aware location-based service discovery protocol in infrastructure-based vehicular network. We performed our simulation on the network simulator ns-2 [13]. In our experiments, we simulated an infrastructure-based vehicular network comprising 16 RRs. We set the average number of vehicles circulating along the network road sections to 130 vehicles. We used realistic mobility traces [14] in order to evaluate the basic and the load balancing and QoS aware LocVSDP's performance.

As illustrated in our simulation parameters table (Table 2), we use the 802.11 as wireless medium with a data transmission rate of 11 Mbps and a transmission

Table 2 Simulation parameters

Parameter name	Parameter value
Wireless medium	802.11
Data transmission rate	11 Mbps
Transmission range (m)	200
Average vehicle's Speed (m/s)	20
Simulation time (s)	1200
Simulation area (m^2)	$600 \times 600 = 120{,}000$
number of vehicles	130
RR number	16
Clients' number	40
servers' number	9
RI surface (m^2)	$\Pi \times 100^2$

range of 200 m. We use a modified version of the manual routing protocol proposed in [15] to route communication packets between RRs in the wireless backbone and RRs and vehicles. In [15], routing paths are generated in hop by hop basis and avoiding routing loops.

We considered 9 service providers located in the RI and providing the same service requested by the 40 service clients circulating in the vehicular network. The arrival time of service requests follows an exponential distribution.

The evaluation of our load balancing and QoS aware discovery protocol required the investigation of various test cases. In our chosen test cases, we assume that service clients' queries target the same service type in the same RI. Besides, we assume that the 9 service providers are located in the same RI specified by the service clients and that they provide the same service type. Service clients can generate one or many service requests during the simulation time.

We assumed that 50% of the service requests specify some performance attributes on service providers and/or routing path.

In the course of our experiments, the request number is defined as the number of service queries sent by vehicle clients. The following metrics have been used for the performance evaluation of the original LocVSDP and the load balancing and QoS-based LocVSDP. They are success rate, connection rate, bandwidth usage, average response time, and average load on vehicular components.

- *success rate* which indicates the average fraction of successful service transactions;
- *connection rate* which indicates the average percentage of successful service connections; i.e., the service requester is able to connect to the service provider through the returned routing path;
- *average response time* which indicates the average time of successful request transactions. It measures the elapsed time for getting a valid service reply in

response to a service request sent by a vehicle. This metric takes into account several factors such as transmission and message processing delay, just to mention few; and

• *bandwidth usage* which measures the bandwidth usage of driver's service queries during the simulation time.

• *average load* which measures the average load on vehicular components such that service providers and RRs.

The results of our extensive simulation experiments are illustrated in Figures 2, 3, 4, 5, 6, 7, and 8. They are obtained after we averaged several runs with an interval of confidence between 90 and 95%. Let us now turn to our results.

Figure 2 illustrates the average load in terms of average number of requests handled by each service provider located in the RI. In this scenario, we set the number of requests to 100. In the basic LocVSDP protocol 97% of the requests are handled by the service provider 116, whereas in the QoS-based LocVSDP every service provider located in *RI* handles at most 11% of the total requests generated during the simulation time. This proves that our QoSLocVSDP succeeds in balancing the load among the different service providers in the RI. In the basic LocVSDP protocol, the service reply contains information about the closest service provider to the center of the RI. Since in our simulation scenario we assume that all the service requests specify the same RI, the closest service provider 116 located closer to the

center of the RI than the other service providers handled 97% of the total requests and 3% of the total requests were dropped. However, in the QoSLocVSDP, service replies contain information about the closest less loaded service provider to the center of the RI. Thus 99% of the total number of requests are handled by service providers inside RI such that every service provider handles at most 11% of the total requests.

Figure 3 depicts the load balancing on RRs inside the RI of the basic LocVSDP and the QoSLocVSDP respectively for 100 requests. Four RRs are located inside the RI, their IDs are 0, 1, 2, and 5, respectively. The figure shows that for the basic LocVSDP, almost 100% of the requests are handled by the elected leader RR ID 1. Whereas for the QoSLocVSDP, requests are handled equitably by all the RRs inside the *RI*. In fact, every RR inside *RI* is elected as leader to process requests of at most 25% of the total requests. Thus, our proposed load balancing mechanism succeeds in balancing the processing load of service requests by RRs inside the *RI*.

Figure 4 plots the load balancing on RRs in the vehicular network using the basic LocVSDP and the QoSLocVSDP respectively for 100 service requests. It shows mainly the used capacity of each RR in the simulated vehicular network. In the basic LocVSDP, almost 100% of the capacity of RR 1 is used. The used capacity of the other RRs is less that 50%. The reason behind this is that in the basic LocVSDP, the service reply returned to the service requester contains information about the fastest path to the service provider, i.e., the path from which the elected leader received the first request
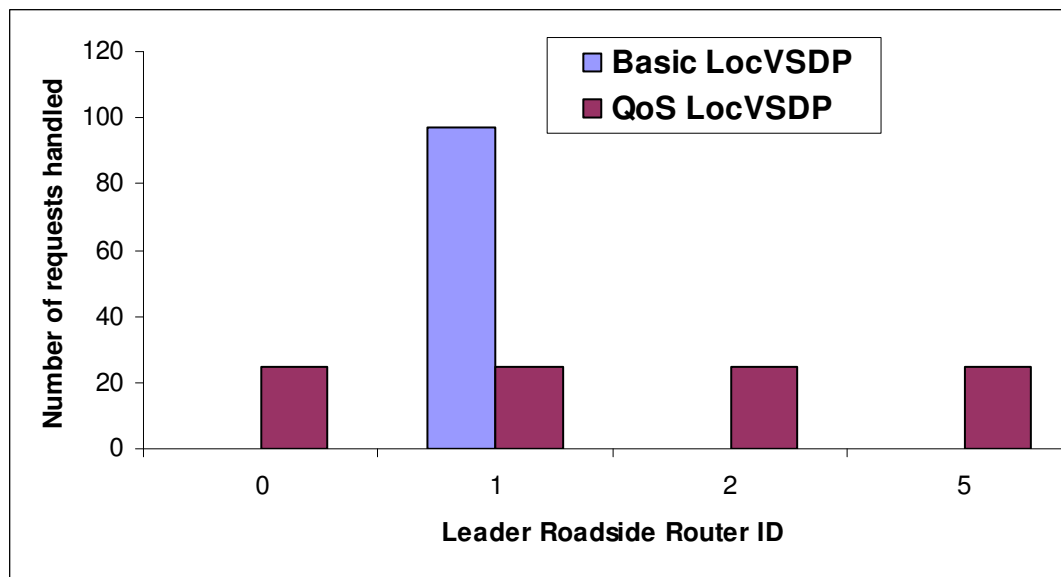

Figure 2 Comparison of load balancing on service providers of the basic LocVSDP and the QoS LocVSDP protocols using a city mobility model.
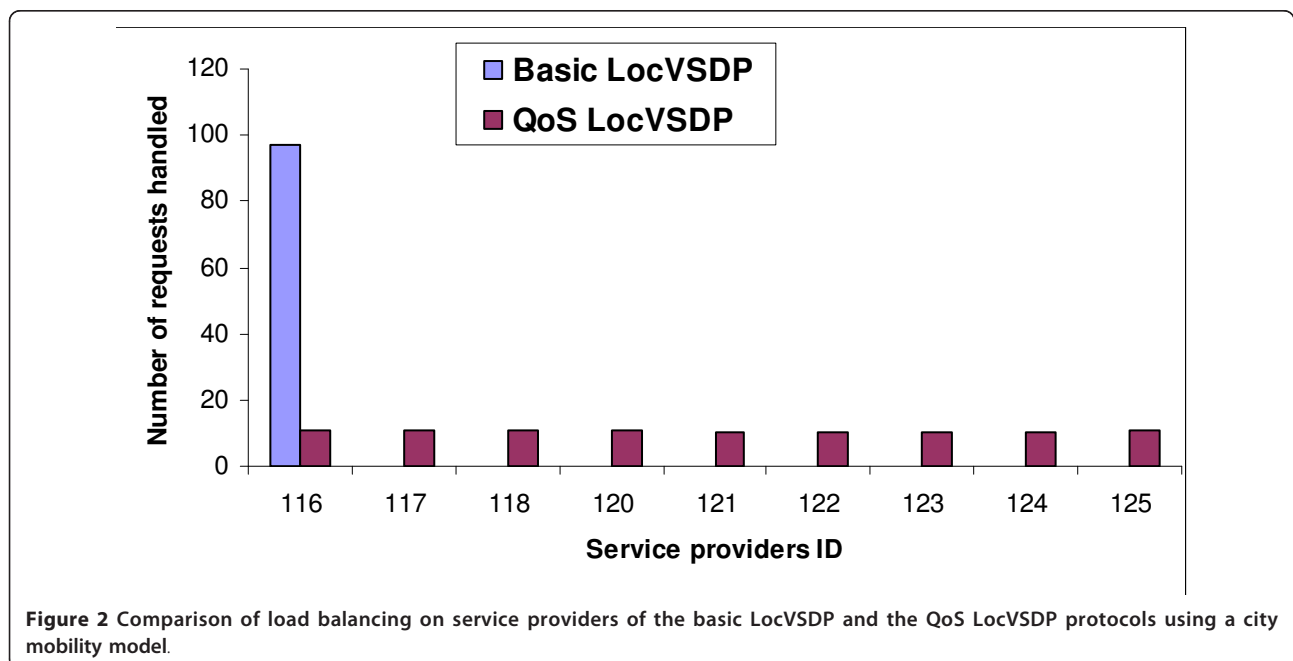
Figure 3 Comparison of load balancing on RRs inside the RI of the basic LocVSDP and the QoS LocVSDP protocols using a city mobility model.

message or parent message as we explained earlier. Thus the basic LocVSDP, does not consider the load and the capacity of RRs in the choice of the routing path between the service provider and the service requester. On the other hand, in the QoSLocVSDP, the used capacity of all the RRs is less than 40%. Thus, our proposed load balancing mechanism succeeds again in guaranteeing a balanced load on the RRs. In fact, in the QoSLocVSDP, a

service reply is geocast to the service requester and the path retained for the communication between the service provider and the service requester is not necessarily the fastest path, but it is the less loaded path among the traversed routing paths in the geocast region between the service provider and the service requester. This explains the balanced capacity used in RRs 0, 1, 2, 3, 4, 5, 6, 7, 11 located inside the geocast region.
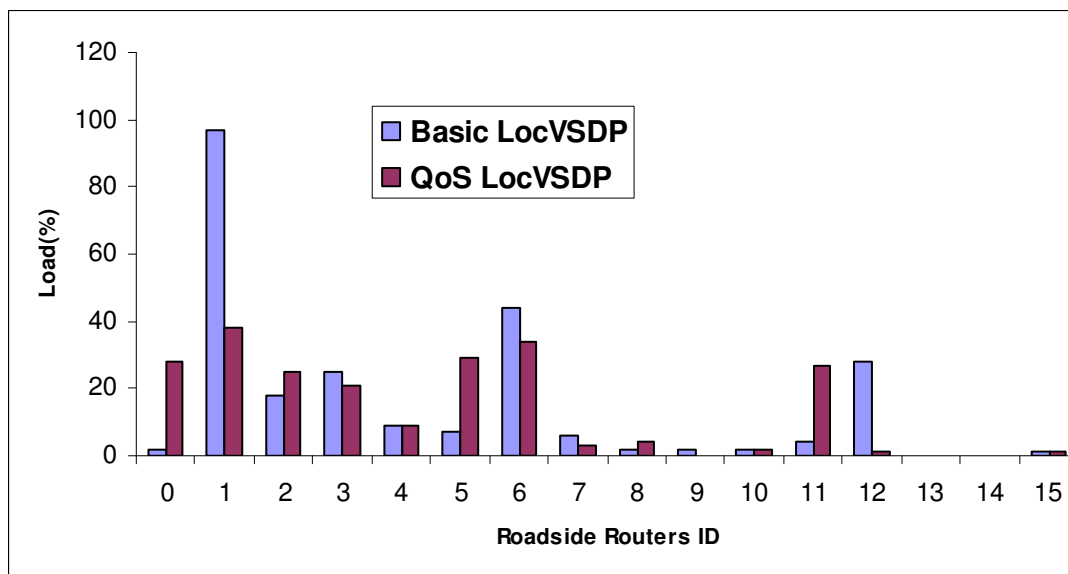
Figure 4 Comparison of load balancing on RRs in the geocast region of the vehicular network of the basic LocVSDP and the QoS LocVSDP protocols using a city mobility model.
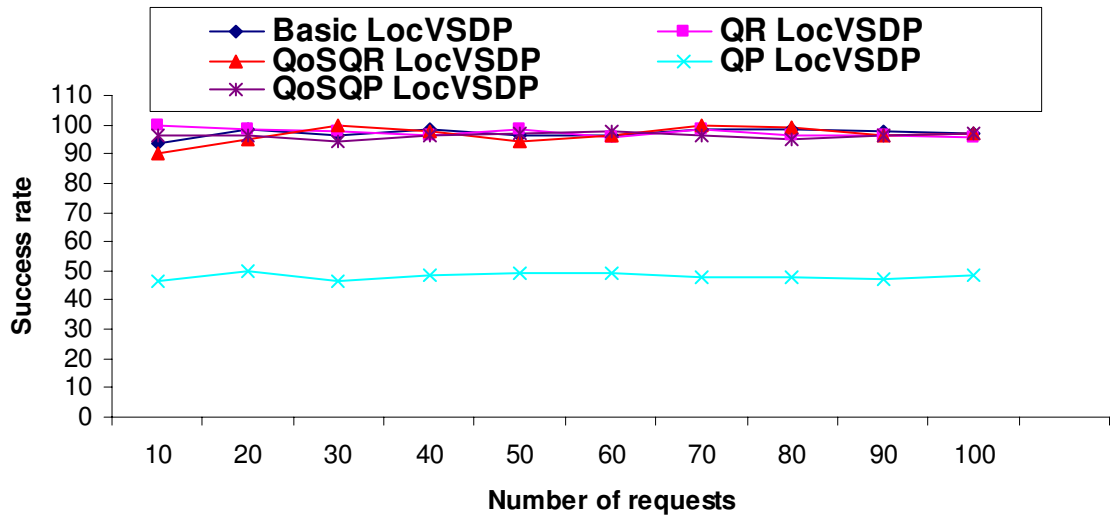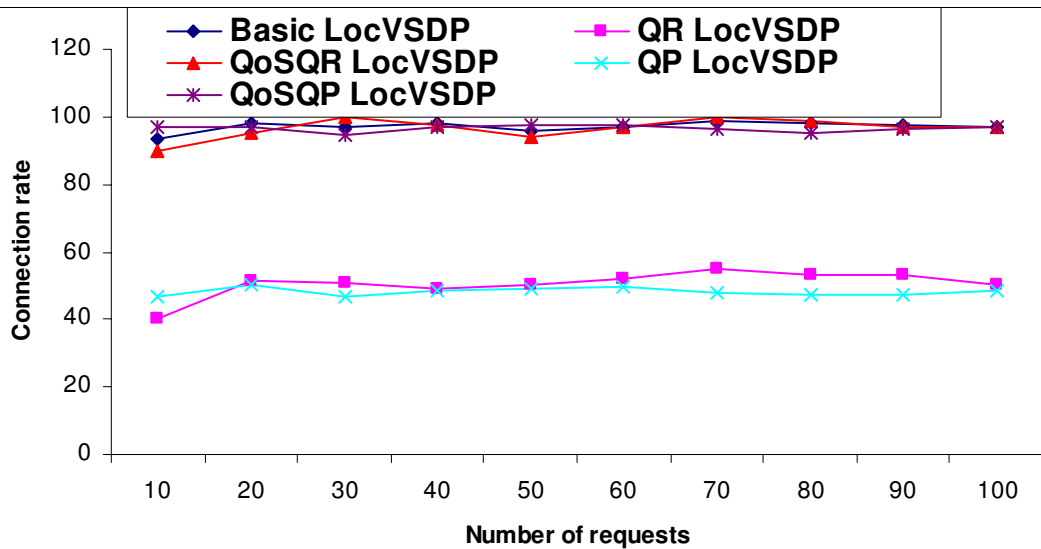
Figure 5 Success rate comparison of multiple scenarios related to the basic LocVSDP and the QoS LocVSDP protocols using a city mobility model.

In Figures 5, 6, 7, and 8, we plot the curves related to the success rate, the connection rate, the bandwidth usage and the average response time respectively, of different variants of the LocVSDP and the QoSLocVSDP when the number of requests varies between 10 and 100. The different variants are describes below:

• Basic LocVSDP: corresponds to the execution of the basic LocVSDP protocol when the service

requester does not specify any QoS attributes in its request.

• QR LocVSDP: corresponds to the execution of the basic LocVSDP when the service requester specifies a QoS requirement in the routing path used for communication to the service provider. Some RRs in the vehicular network between service requesters and service providers do not satisfy the requested QoS requirement.

Figure 6 Connection rate comparison of multiple scenarios related to the basic LocVSDP and the QoS LocVSDP protocols using a city mobility model.
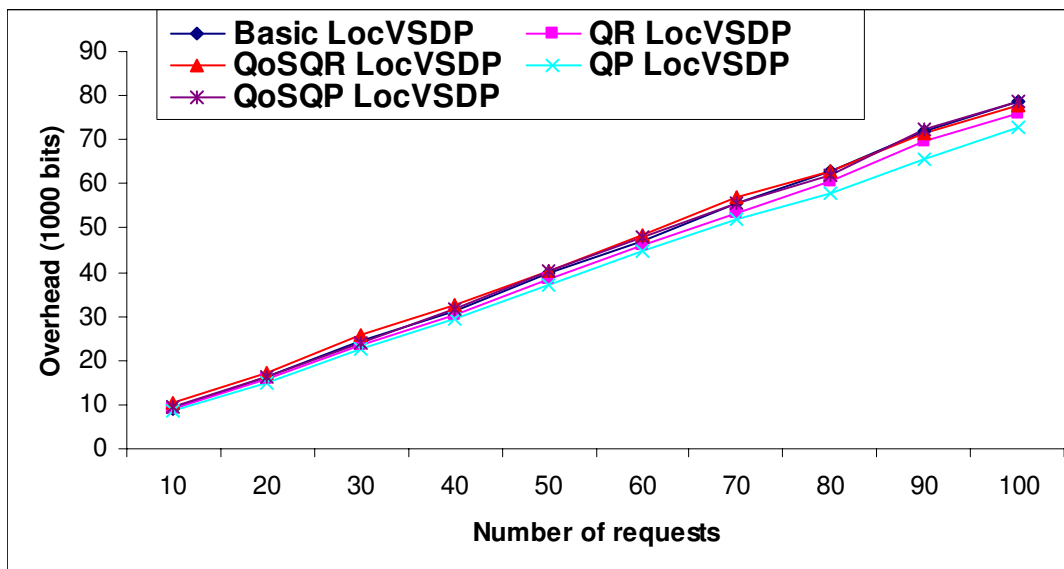
Figure 7 Bandwidth usage comparison of multiple scenarios related to the basic LocVSDP and the QoS LocVSDP protocols using a city mobility model.

• QoSQR LocVSDP: corresponds to the execution of the QoSLocVSDP when the service requester specifies a QoS requirement in the routing path used for communication to the service provider. Some RRs in the vehicular network between service requesters and service providers do not satisfy the requested QoS requirement.

• QP LocVSDP: corresponds to the execution of the basic LocVSDP when 50% of the total number of requests have QoS requirement in the requested service provider; and the service provider that have the ID 116 does not satisfy the QoS requirement.

• QoSQP LocVSDP: corresponds to the execution of the QoSLocVSDP when the service requester specifies a QoS requirement in the requested service provider.

In Figure 5, we depict that the success rate of the different scenarios is quite high (more than 90%) except
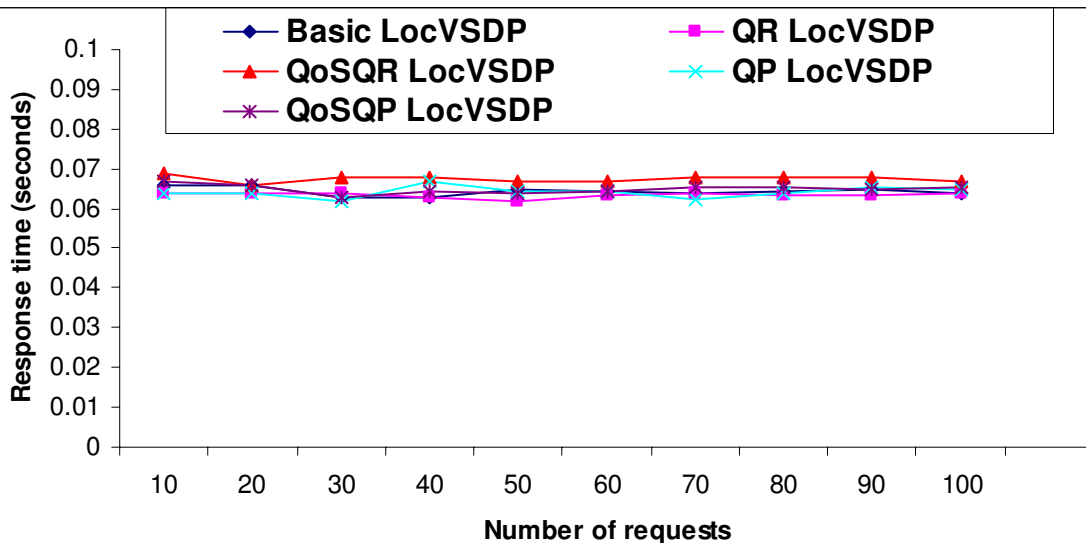

Figure 8 Average response time comparison of multiple scenarios related to the basic LocVSDP and the QoS LocVSDP protocols using a city mobility model.

for the *QP LocVSDP* Scenario. This latter have a success rate in the order of 50%. *QP LocVSDP* corresponds to the execution of the basic LocVSDP when 50% of the total number of requests have QoS requirement in the requested service provider. Since the basic LocVSDP does not take into consideration any QoS requirement and that the closest service provider to the origin of the *RI* is the provider with ID 116, almost 50% of the service requests in this scenario cannot be satisfied. This explains the low success rate of in the *QP LocVSDP* scenario. However for the *QoSQP LocVSDP*, which corresponds to the execution of the QoSLocVSDP when 50% of the total number of requests have QoS requirement in the requested service provider, have high success rate. This is because our proposed QoSLocVSDP mechanism takes into account the QoS requirements specified in the clients requests. The success rate in the other scenarios (*QR LocVSDP* and *QoSQR LocVSDP*) is not affected when the requester specifies a QoS requirement in the requested service provider.

Figure 6 plots the curves related to the connection rate of service requesters to their correspondent service providers. We notice that the connection rate of the curve that correspond to the scenario *QR LocVSDP* is less than 50%. This is mainly due to the fact that the basic LocVSDP returns a service reply to the service requester containing the fastest path to the service provider and does not consider any QoS requirements in the routing path. The fastest path could not have all its routing components satisfy the QoS requirement specified in the client request. Consequently, even if the service provider is found and the service reply is returned to the service requester (which explains the high success rate of *QR LocVSDP* in Figure 5), this latter cannot establish a communication connection through the returned routing path. That is why the connection rate in the *QR LocVSDP* scenario is not high. However, the connection rate in the *QoSQR LocVSDP* scenario is high because this latter takes into account the QoS requirements specified in a service request.

The curve corresponding to *QP LocVSDP* scenario has its connection rate as good as its success rate. This is because only the requesters that receive service replies with the appropriate service provider that can establish a connection with their correspondent service providers.

Figure 7 plots the curves of the bandwidth usage of the different scenarios related to the basic LocVSDP and the QoSLocVSDP when the number of requests ranges between 10 and 100. We notice that all the curves have the same behavior. The bandwidth usage of the different scenarios is in the order of 10000 bits for 10 requests and increases to reach around 80000 bits for 100 requests. If we examine more closely the curves, we notice that the bandwidth usage in the scenarios related to the QoSLocVSDP is slightly more than the bandwidth usage in the scenarios related to the basic LocVSDP. This is mainly due to the fact that in the QoSLocVSDP, more messages are exchanged between vehicular components that are related to load balancing and QoS requirements satisfaction purposes.

Figure 8 depicts the average response time of the different scenarios related to the execution of the basic LocVSDP and the QoSLocVSDP when the number of requests ranges between 10 and 100. We notice that the average response time of the different scenarios is in the order of 65 milliseconds. The QoSLocVSDP protocol does not have an impact on the average response time.

We conclude that our proposed load balancing and QoSLocVSDP protocol succeeds in balancing the load between the different components of the vehicular networks (RRs and service providers) and satisfies clients requests with different QoS requirements, while guaranteeing a good response time and appropriate bandwidth usage.

8. Conclusion

In this article, we proposed an efficient load balancing and QoS-based location-based service discovery protocol for vehicular networks. We presented its proof of correctness and we computed its message complexity and its bandwidth usage. We presented our simulation experiments and performance evaluation in comparison to the original LocVSDP. Our proposed protocol succeeds in balancing the load between the different components of the vehicular network, including RRs and service providers. Moreover, it showed high performances, in terms of response time and bandwidth usage while satisfying clients requests with different QoS requirements.

Author details
[1]DIVA Research Center, University of Ottawa, Ottawa, ON K1N 6N5, Canada [2]Paradise Research Laboratory, University of Ottawa, Ottawa, ON K1N 6N5, Canada [3]College Of Computer and Information Sciences, King Saud University, Riyadh 11543, Kingdom Of Saudi Arabia

Competing interests
The authors declare that they have no competing interests.

References
1. MD Dikaiakos, A Florides, T Nadeem, L Iftode, Newblock location-aware services over vehicular ad-hoc networks using car-to-car communication. IEEE J Sel Areas Commun. **25**(8), 1590–1602 (2007)

2. N Klimin, W Enkelmann, H Karl, A Wolisz, A hybrid approach for location-based service discovery in vehicular ad hoc networks, in *Proc of 1st Intl Workshop on Intelligent Transportation (WIT)*, Citeseer, Hamburg, Germany, (Mar 2004)
3. A Boukerche, K Abrougui, RW Pazzi, Context-aware and location-based service discovery protocol for vehicular networks, in *Proceedings of the 6th ACM symposium on Performance evaluation of wireless ad hoc, sensor, and ubiquitous networks*, ACM, 93–100 (2009)
4. F Nidito, M Battelli, S Basagni, Fault-tolerant and load balancing localization of services in wireless sensor networks, in *2007 IEEE 66th Vehicular Technology Conference, 2007. VTC-2007 Fall*, Citeseer, 382–386 (2007)
5. M Albano, S Chessa, F Nidito, S Pelagatti, Q-NiGHT: Adding qos to data centric storage in non-uniform sensor networks, in *Mobile Data Management, 2007 International Conference on*, Citeseer, 167–173 (2007)
6. S Ratnasamy, B Karp, S Shenker, D Estrin, R Govindan, L Yin, F Yu, Data-centric storage in sensornets with GHT, a geographic hash table. Mob Netw Appl. **8**(4), 427–442 (2003). doi:10.1023/A:1024591915518
7. O Riva, T Nadeem, C Borcea, L Iftode, Context-aware migratory services in ad hoc networks. IEEE Trans Mob Comput. **6**(12), 1313–1328 (2007)
8. S Dolev, S Gilbert, NA Lynch, E Schiller, AA Shvartsman, JL Welch, Virtual Mobile Nodes for Mobile Ad Hoc Networks. Lecture Note Comput Sci. **3274**, 230–244 (2004). doi:10.1007/978-3-540-30186-8_17
9. R Handorean, R Sen, G Hackmann, GC Roman, Context aware session management for services in ad hoc networks, in *IEEE International Conference on Services Computing*, 113–120 (2005)
10. B Hull., *et al*, CarTel: a distributed mobile sensor computing system, in *4th ACM SenSys*, Boulder, CO, 125–138 (2006)
11. AT Campbell, SB Eisenman, ND Lane, E Miluzzo, RA Peterson, People-centric urban sensing, in *Proceedings of the 2nd annual international workshop on Wireless internet*, (ACM Press, New York, NY, USA, 2006), p. 18
12. Urban Sensing Project, http://research.cens.ucla.edu/projects/2006/systems/urban_sensing/
13. The network simulator 2, http://www.isi.edu/nsnam/ns/
14. J Haerri, F Filali, C Bonnet, in *Mobility models for vehicular ad hoc networks: a survey and taxonomy*, (Institut Eurecom, 2007). research rep. RR-06-168
15. A Raniwala, T Chiueh, Architecture and algorithms for an IEEE 802.11-based multi-channel wireless mesh network, in *Proceedings IEEE INFOCOM*. **3**, 2223–2234 (2005)