Modeling and Analysis of Location Service Management in Vehicular Ad Hoc Networks

by

Hanan Saleet

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2010

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Hanan Saleet

Abstract

Recent technological advances in wireless communication and the pervasiveness of various wireless communication devices have offered novel and promising solutions to enable vehicles to communicate with each other, establishing a decentralized communication system. An emerging solution in this area is the Vehicular Ad Hoc Networks (VANETs), in which vehicles cooperate in receiving and delivering messages to each other. VANETs can provide a viable alternative in situations where existing infrastructure communication systems become overloaded, fail (due for instance to natural disaster), or inconvenient to use. Nevertheless, the success of VANETs revolves around a number of key elements, an important one of which is the way messages are routed between sources and destinations. Without an effective message routing strategy VANETs' success will continue to be limited.

In order for messages to be routed to a destination effectively, the location of the destination must be determined. Since vehicles move in relatively fast and in a random manner, determining the location (hence the optimal message routing path) of (to) the destination vehicle constitutes a major challenge. Recent approaches for tackling this challenge have resulted in a number of Location Service Management Protocols. Though these protocols have demonstrated good potential, they still suffer from a number of impediments, including, signaling volume (particularly in large scale VANETs), inability to deal with network voids and inability to leverage locality for communication between the network nodes.

In this thesis, a Region-based Location Service Management Protocol (RLSMP) is proposed. The protocol is a self-organizing framework that uses message aggregation and geographical clustering to minimize the volume of signalling overhead. To the best of my knowledge, RLSMP is the first protocol that uses message aggregation in both updating and querying, and as such it promises scalability, locality awareness, and fault tolerance.

Location service management further addresses the issue of routing location updating and querying messages. Updating and querying messages should be exchanged between the network nodes and the location servers with minimum delay. This necessity introduces a persuasive need to support Quality of Service (QoS) routing in VANETs. To mitigate the QoS routing challenge in VANETs, the thesis proposes an Adaptive Message Routing (AMR) protocol that utilizes the network's local topology information in order to find the route with minimum end-to-end delay, while maintaining the required thresholds for connectivity probability and hop count. The QoS routing problem is formulated as a constrained optimization problem for which a genetic algorithm is proposed. The thesis presents experiments to validate the proposed protocol and test its performance under various network conditions.

Acknowledgments

I would like to express my deepest gratitude to my supervisor Prof. Otman Basir for his great support and for giving me the opportunity to pursue this research work and write this dissertation under his supervision. His guidance, invaluable suggestions, and fruitful input have contributed greatly to this dissertation. I learned important lessons from his career experience that he shared with me. Without his vision, encouragement, and understanding my achievements could not have been possible.

I would like to thank my PhD committee, Prof. Eihab Abdel-Rahman, Prof. Sagar Naik, Prof. Kumaraswamy Ponnambalam and Prof. Baher Abdulhai, for their invaluable feedback and comments. I highly appreciate their effort in reading my thesis. Furthermore, my deep thanks to Prof. Keith Hipel for serving as the delegated dissertation committee member.

I would like to express my deepest appreciation to Prof. Raouf Boutaba for giving me the chance to be part of his wireless communication group, where I gained excellent knowledge. I am very grateful to him for his understanding, constructive advices, encouragement and personal guidance.

I would like to thank Prof. Rami Langar for his fruitful input and professional collaboration and for sharing with me his invaluable experience in wireless networking.

I would like to thank Prof. Sherman Shen for giving me the chance to participate in the Broadband Communication Research (BBCR) group. Special thanks to Dr. Atef Abdrabou and Albert Wasef for their support and encouragement.

Special thanks, also, for my friends and colleagues Dr. Rasha Kashif, Tarek Khalifa, Aziza Khalifa, Nabil Drawil, Sayyid Anas Vaqar and Khalid Almahrog for their support and interesting discussions.

Finally, I wish I can express my deepest gratitude to my husband and best friend, Abdulghafar, for his endless love, patience and encouragement.

Dedication

To Abdulghafar whose dreams shaped my life,

To my dear parents, brothers and sisters,

To my lovely daughters Rana, Dalia and Jana

Contents

List of Tables

List of Figures

Nomenclature

AMR: Adaptive Message Routing

CL: Cell Leader

GA: Genetic Algorithm

ID: Identification Number

ITS: Intelligent Transportation Systems

LS: Location server

LSC: Location Service Cell

MANETs: Mobile Ad hoc NETworks

QoS: Quality of Service

RSU : Road Side Unit

RLSMP: Region-based Location Service Management Protocol

WLAN: Wireless Local Area Networks

VANETs: Vehicular Ad hoc NETworks

Chapter 1

Introduction

1.1 Introduction

Wireless communication technologies are becoming increasingly available and inexpensive. Users are becoming connected nearly everywhere: at work, at home, and even on roads. In addition to the cellular networks and wireless local area networks (WLANs), vehicular ad hoc networks (VANETs) promise Intelligent Transportation Systems (ITS) new attractive and cost effective services that can definitely benefit users (drivers and passengers).

Vehicular ad hoc networks are non-traditional intelligent alternatives that play an important role in ITS because they provide promising solutions to enable infrastructure-free communication between vehicles. In VANETs, vehicles or nodes, are equipped with wireless communication devices that create wireless links between these nodes. A node can send data directly to another node which is located within its transmission range, without depending on an expensive fixed infrastructure. A node can also send data to another node that is not located within its transmission range with the help of intermediate nodes, forming a process of multihop message routing.

Message routing is a challenging problem in VANETs due to the inherent high degree of mobility of a large number of nodes. To enable message routing, the source node should be able to locate the destination node (node localization), and subsequently, it can build a reliable route towards the destination node. This thesis studies inter-vehicular communication in VANETs, as it relates to both challenges: node localization and message routing.

1.2 Vehicular Ad Hoc Networks

In many cases, existing fixed infrastructure-based communication systems may become overloaded, absent, or inconvenient to use. In such cases, allowing vehicles to communicate with each other without using the expensive infrastructure offers a viable alternative [1] – [10]. If a communication path is established between two end-users, then it is possible for them to communicate directly without resorting to a central management unit. Such a wireless communication network is built in an ad hoc and self-organizing manner.

1.2.1 Characteristics and requirements

VANETs represent a rapidly emerging and challenging class of Mobile Ad Hoc Networks (MANETs). Each node of the network operates not only as a host but also as a router, forwarding packets for other mobile nodes [11, 12]. When there is no direct path between the end users (node S and node D in Figure 1.1), the route is established through multihop routing such that users in the middle, such as node R, act as routers.

VANETs resemble MANETs in the sense of node mobility. However, unlike MANETs, the mobile nodes in VANETs tend to move at much higher velocity, resulting in rapidly and dynamically changing network topology. However, a

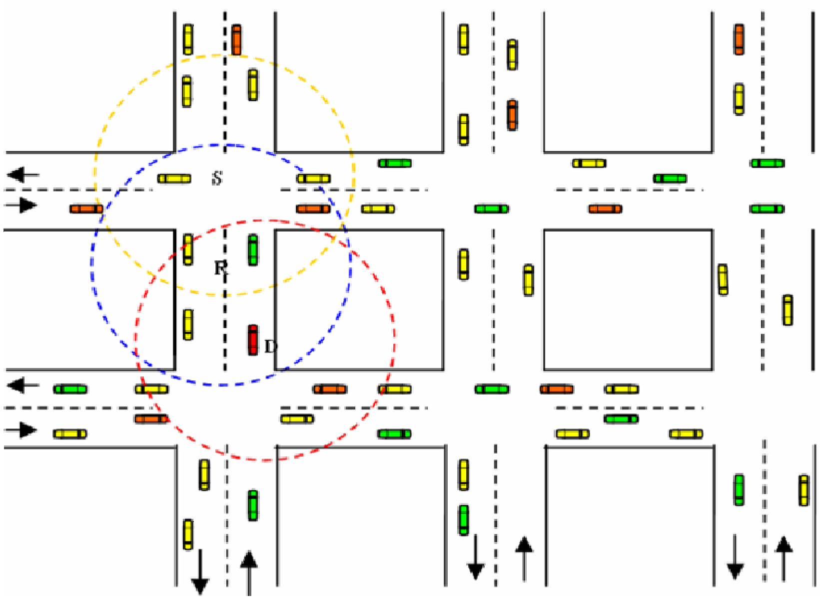


Figure 1.1: Forwarding messages in VANETs

vehicle's velocity is regulated by speed limits, stop signs and traffic lights. Furthermore, vehicles move on a road network, and hence their movement occurs on a limited number of pre-specified trajectories rather than a totally random movement as is the case in MANETs. Power scarcity and storage capacity are more of an issue in MANETs than in VANETs because VANETs vehicles are built to have communication devices with rechargeable sources of energy, extensive storage and processing capacities, and potentially long transmission ranges. Moreover, it is expected that all VANET equipped vehicles know their locations using pervasive technologies such as Global Positioning System (GPS) and integrated electronic road maps [13].

1.2.2 Applications

In VANETs, vehicles form a decentralized communication network by means of wireless multihop communication; thus, VANETs can enable a wide range of applications.

Recent research work on VANETs has resulted in the development of many valuable safety applications [14] – [16]. One such application is on-board active safety systems that enable a vehicle to disseminate information intelligently in order to assist drivers in avoiding collisions and road hazards. For example, in Figure 1.2 due to slippery road conditions, the yellow car derailed from its lane, and as a result it reports this incident to cars in its vicinity so as to allow them to take alternative paths, if possible. Similarly, real-time relevant road and traffic information such as road construction areas, localized fog conditions, icy conditions and traffic jams can be routed using VANETs.

VANETs not only enhance traffic safety but also enable applications such as traveler assistance applications or infotainment to the drivers and passengers via multihop communications between vehicles [1, 2]. For example, drivers and passengers may communicate to share files such as music or movies. Moreover, VANETs can be used to enable location-based content delivery applications such as traveler guidance, news, and weather information.

Minimal configuration and rapid deployment make VANETs suitable also for emergency applications. One vital scenario which may require an essential communication system between vehicles, is the case of a disaster in which all fixed infrastructure-based communication systems are destroyed. For example, in cases of disasters such as wars, tornados, earthquakes and fires, VANETs can enable parties like rescue workers, hospital staff and police to communicate with each other. Moreover, police or military platoons can be formed easily in any region that is not covered by infrastructure-based networks. In addition, in rural areas that lack infrastructure-based communication , and in cases of emergency, VANETs can be established temporarily to enable vehicles to cooperate in receiving and delivering messages. Also, in cases where existing infrastructure-based communication systems are overloaded, a hospital, for example, can use a VANET to locate and forward messages to an emergency car while it is in mission.

This research work proposes a system model to enable search and rescue forces to communicate using VANETs in case of a disaster emergency. The proposed VANET environment takes advantage of the presence of low cost road side units (RSUs) to achieve efficient message delivery. Nevertheless, there could be emergency situations where fixed infrastructure such as RSUs get destroyed or infeasible to setup; hence the proposed system aims to function in situations like this where the mobile nodes should cooperate to establish a VANET without relying on the RSUs.

1.3 Research Motivations and Challenges

Governments, car manufacturers, and academic institutions are increasingly interested in the research and development of VANETs [18, 19]. The promising applications and the cost effectiveness of VANETs constitutes a major motivation behind this level of interest. Nevertheless, the success of VANETs revolves around a number of key elements, an important one of which is the way messages are routed between sources and destinations. Without an effective routing strategy VANETs' success will continue to be limited.

In order for messages to be routed to a destination effectively, the location of the destination must be determined. Since vehicles move in relatively fast and in a random manner, determining the location (hence the optimal message routing path) of (to) the destination vehicle constitutes a major challenge. Re-

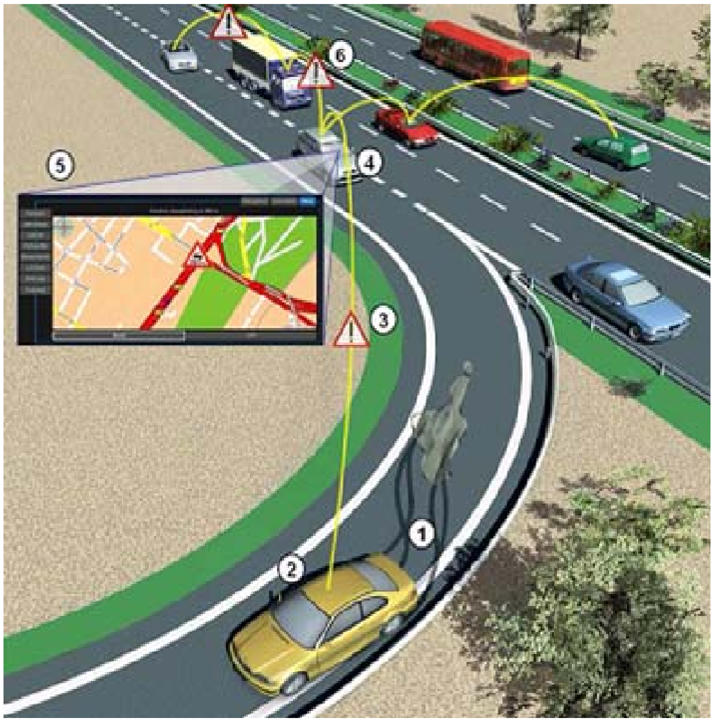


Figure 1.2: Active safety application [17]

cent approaches for tackling this challenge have resulted in a number of Location Service Management Protocols [20]. Though these protocols have demonstrated good potential, they still suffer from a number of impediments, including signaling volume (particularly in large scale VANETs), inability to deal with network voids and inability to leverage locality for communication between the network nodes.

A central challenge in the deployment of VANETs has been the design of an efficient location service management protocol that can track the locations of mobile nodes and reply to location queries with minimum overhead. Generally speaking, a location service management protocol consists of the following two main components [21]:

1. Location Updating: while the nodes are moving, the location of each

node is continuously updated and stored at other nodes called location servers of the network; the geographical positions of the location servers are called home regions (or pointers).

2. Location Retrieval (Query): when a node wants to communicate with another node, it sends a querying message to the location servers to retrieve location information about the destination node.

Location service management in VANETs is a challenging task for two reasons. Firstly, the large number of nodes in VANETs puts a communication burden on the limited channel bandwidth. Nodes not only send data messages but also send control messages such as location updating and location querying messages. Each node in the network needs to send location updating messages to inform other nodes about its current position which changes with time. On the other hand, when a node wants to communicate with another node, it sends a location querying message through the network to retrieve the location of that destination node. With a large number of nodes in the network, it is obvious that a large volume of control messages ensue to the extent that it may consume the already limited network bandwidth.

Secondly, even though VANETs have the potential to provide a more practical alternative than fixed infrastructure wireless networks which may be prohibitively expensive to construct especially in harsh environments, but unfortunately, in VANETs, random mobility of the nodes which have a limited transmission range, coupled with the lack of adequate fixed infrastructure, renders the rapidly and dynamically changing VANETs an unpredictable environment; this increases the complexity of the location service management protocols designed for VANETs.

There has been many proposals to deal with the location service management challenges in mobile ad hoc networks. These proposals can be divided into two categories: flooding-based and quorum-based protocols. Flooding-based protocols involve global network flooding, in which broadcast messages are sent through the whole network [22] – [25]. Such flooding results in severe performance degradation and scalability reduction, especially in urban areas with high vehicle density and with fast mobility [26]. These protocols are often suitable for small size networks with slow mobility [3]. On the other hand, in quorum-based protocols, the location servers, or quorums, result from mapping the nodes' identifiers or geographical information to quorums in either a static or a random way [27, 28]. To design a quorum-based location service management protocol, three main questions should be answered [29]: How location server(s) are chosen? How should a node update these location servers as it moves? How does a source node discover the appropriate location server(s) to retrieve the destination location information? Quorum-based protocols are more scalable than flooding-based protocols, especially in highly dynamic ad hoc networks [30]. However, their performance is dependent on the existence of an efficient location service management scheme. This thesis capitalizes on quorum based approaches to achieve efficient and self organizing location service management.

Since it is important to have up to date information about destinations' locations, mobile nodes should be able to route the updating and querying messages to the location servers in a timely manner. Nevertheless, routing updating and querying messages in VANETs is challenging for three reasons.

Firstly, nodes in a VANET change their location frequently, as such they need to send location updating messages to the location servers with minimum delay. Doing so allows the location servers to have the correct and up to date location information when replying to the location queries. Thus, it is important to select routes between nodes and the location servers that guarantee minimum delay.

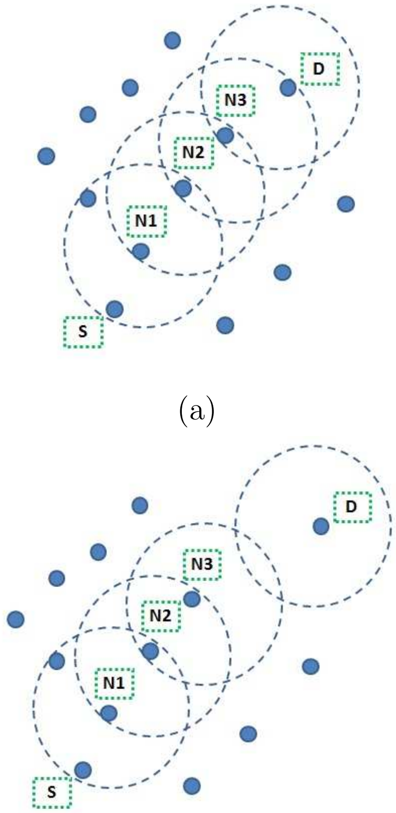
Secondly, vehicles move with high velocity and they often change their speed and direction. Node random mobility affects route stability. When a route is established between two nodes, mobility of the source, destination, or intermediate nodes may cause the route to break. If two nodes on the route move out of transmission range, the link between these two nodes vanishes, and then the route fails. Figure 1.3 (a) shows a route from the source node S to the destination node D through N1, N2 and N3. Node D moves out of the transmission range of node N3 (see Figure 1.3 (b)), and hence D can not receive massages. In this case, the route from S to D breaks because of mobility of node D, affecting the connectivity probability of the established route. Thus, it is critical to select routes that have a high connectivity probability.

Thirdly, the route length measured in terms of number of hops affects quality of the route. To save bandwidth, which is precious in VANETs, it is important to select a route with a minimum number of nodes.

One of the challenging tasks in the deployment of VANETs, therefore, is to design a routing protocol that minimizes the end to end delay while satisfying quality of service (QoS) constraints with respect to the probability of connectivity and hop count. In other words, it should be capable of handling frequent path disruptions caused by high mobility of vehicles and, at the same time, makes efficient use of the network bandwidth. Traditional routing protocols used in MANETs need to be modified so as to accommodate such unique characteristics of VANETs.

The routing protocols in MANETs fall into two categories: topology-based protocols and geographical based protocols [31] – [41]. In topology based protocols, it is assumed that each node has information about the entire network topology before the node begins forwarding messages. In general, this process causes deterioration in network performance as it introduces congestion in the communication channel. Studies in [30] have shown that geographical based routing protocols have better scalability than topology-based protocols in highly dynamic ad hoc networks. Geographical based routing protocols assume that each node knows its own location using, for example, GPS.

Figure 1.4 shows the relation between geographical based routing protocols



(a)

(b)

Figure 1.3: Connectivity in ad hoc networks. (a) Route from S to D, (b) Route Break

and location service management protocols. The routing protocols require input from the location service management protocols. Thus, when a source node knows the location of a destination node, it uses message routing to deliver the messages. The aim of this research is to design a location service management protocol that is suitable for the VANETs environment and to build routes that are used to forward the updating and querying messages in a timely manner.

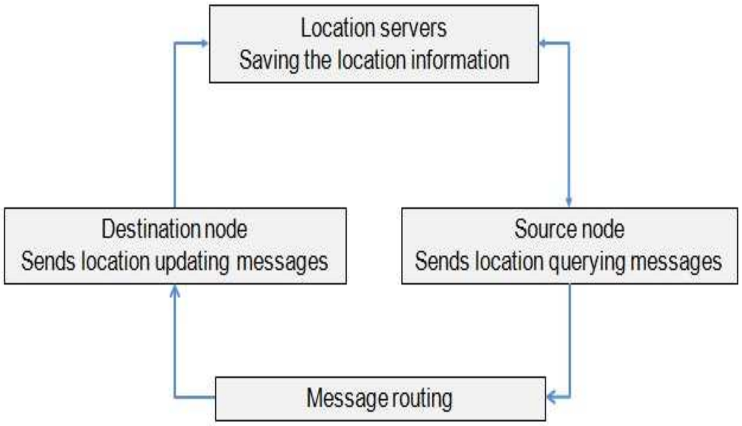


Figure 1.4: Message routing in ad hoc network

1.4 Research Objectives and Contributions

To support a wide range of applications, VANETs must be able to efficiently locate destination vehicles or nodes (for example, emergency vehicles, or a police or even a friend's vehicles). The proposed VANET environment takes advantage of the presence of low cost road side units. RSUs can be used as location servers that can manage location information. Nevertheless, there could be emergency situations where fixed infrastructure such as RSUs get destroyed or infeasible to setup; hence the proposed system aims to function in situations like this by using mobile nodes to perform location service management.

A critical issue in VANETs is to achieve an efficient location service management that minimizes the signalling traffic overhead encountered in updating and querying the location information. Signalling traffic overhead depends on traffic pattern, scalability, and the degree of fault tolerance provided by the protocols. Therefore, the first objective of this research is to model and design a location service management protocol that copes well with the demands of highly dense urban VANETs; it is characterized by scalability, locality awareness and fault tolerance. The second objective of this research is to enable the efficient routing of updating and querying messages in the network without exhausting the network limited bandwidth, while satisfying QoS constraints. In order to realize the two objectives, the following three main research stages are conducted.

1. An architectural framework for VANETs that supports inter-vehicular communication in large-scale urban scenarios where large numbers of vehicles are moving with high velocity is formulated. To achieve a stable structure for the VANETs' framework, a novel location service management protocol is developed that achieves location querying and location updating using the concept of geographical clustering. In addition, the framework employs message aggregation in both querying and updating, where location information is aggregated in an attempt to improve capacity utilization of VANETs. In addition, since the communication pattern in VANETs is considered, to a great extent, to be local (i.e., vehicles inside one geographical area are more likely to communicate with each other), the framework incorporates locality awareness as a building block in the design process. The framework also achieves fault tolerance whereby location servers should never fail to answer the queries.

2. The effectiveness of the proposed architecture is investigated by deriving analytical expressions for both location updating and querying cost. Furthermore, the total control overhead problem is formulated as an optimization problem of minimizing the communication overhead, allowing VANETs' designers the ability to consider alternative values for the network design parameters. Also, the performance of location information updating and querying is investigated based on simulation experiments using both Random Walk mobility model and real mobility traces. Analytical

and simulation results show that the proposed framework can significantly reduce location updating cost and provide low querying overhead under various scenarios.

3. The use of RSUs as location servers to maintain current location information about the mobile nodes in their vicinity is proposed. Since it is important to have up to date information about destinations' locations, mobile nodes should be able to route the updating and querying messages to the local RSU in a timely manner. Therefore, this thesis propose a new approach for routing delay-sensitive location updating and querying messages in vehicular ad hoc networks. The proposed approach to this problem uses paths that minimize the end-to-end delay while maintaining a threshold on the probability of connectivity (connectivity degree) and maintaining a threshold on the hop count along the selected path. To achieve this, analytical expressions for the delay, hop count, and connectivity probability for a two-way road scenario are derived. The QoS routing problem is formulated as a constrained optimization problem for which a genetic algorithm is proposed. Using both analytical and simulation studies, it is concluded that the proposed protocol achieves substantial end-to-end delay reduction, especially in sparse networks. The proposed solution, therefore, stands out as a promising candidate for large scale ad hoc networks such as VANETs.

1.5 Thesis Outline

This thesis is organized as follows: Chapter 1 introduces VANETs and describes the concepts of location service management and message routing, as well as the motivation for and objectives of this research. Chapter 2 presents the background material for this research. Chapter 3 describes the problem and the solution approach. Chapter 4 illustrates the proposed location service management protocol and its application in solving the location service management problem in VANETs. Chapter 5 presents a message routing protocol that accounts for quality of service to route the updating and querying messages in a timely manner. Chapter 6 summarizes the thesis and provides suggestions for interesting and challenging future research.

1.6 Summary

This chapter provides an overview of VANETs and discusses two of the main challenges encountered in their deployment, namely, location service management and message routing. It also identifies the research gaps and lists the research objectives. The next chapter provides an extensive survey of relevant literature.

Chapter 2

Literature Survey

2.1 Introduction

Vehicular ad hoc networks are emerging as an important requirement for modern communication. Many projects that avail of the cooperation between universities and car manufacturers have evolved to bring VANETs to life. The cost effectiveness of VANETs constitutes a major motivation behind this level of interest. Examples of initiatives that have developed in this area include: the Car2Car Communication Consortium which is dedicated to improving road safety by means of inter-vehicle communications [42]; the CarNet project which resulted in a geographic routing protocol to route packets from car to car without flooding the network [43]; FleetNet project which was launched with the aim to develop position-based forwarding strategies for VANETs, where location awareness plays a crucial role [44]; and the Car-to-Car Cooperation C3, a traffic advisory system that informs drivers in real time of current conditions on roads of concern [45]. Therefore, these projects envision VANETs as decentralized communication networks that are intended to support different types of applications [1, 2, 14, 16]. Nevertheless, to support such applications a number of challenges should be addressed. This thesis deals with two challenges which are location service management and message routing.

This chapter presents a number of the existing location service management protocols and message routing protocols and discusses the drawbacks that make these protocols unsuitable for VANETs.

2.2 Location Service Management Protocols

Location service management is designed to solve problems of location tracking and retrieval. In location service management, some or all nodes in the network act as location servers for other nodes. When a node moves to a new location, it updates its location servers with its new location; and when another node wants to communicate with this specific node, it queries the location servers to retrieve the location information.

Numerous protocols has been proposed for location service management. In the following section, the taxonomy of location service management is discussed. This taxonomy is shown in Figure 2.1, where location service management protocols are classified into flooding-based and quorum-based protocols.

2.3 Flooding-Based Location Service Management Protocols

Flooding-based protocols involve global network flooding, where broadcast messages are sent through the whole network [63]. Such flooding results in severe performance degradation and scalability reduction, especially in urban environments that exhibit high vehicle density and fast vehicle mobility. Therefore, these protocols are more suitable for small networks with slow mobility [3]. The

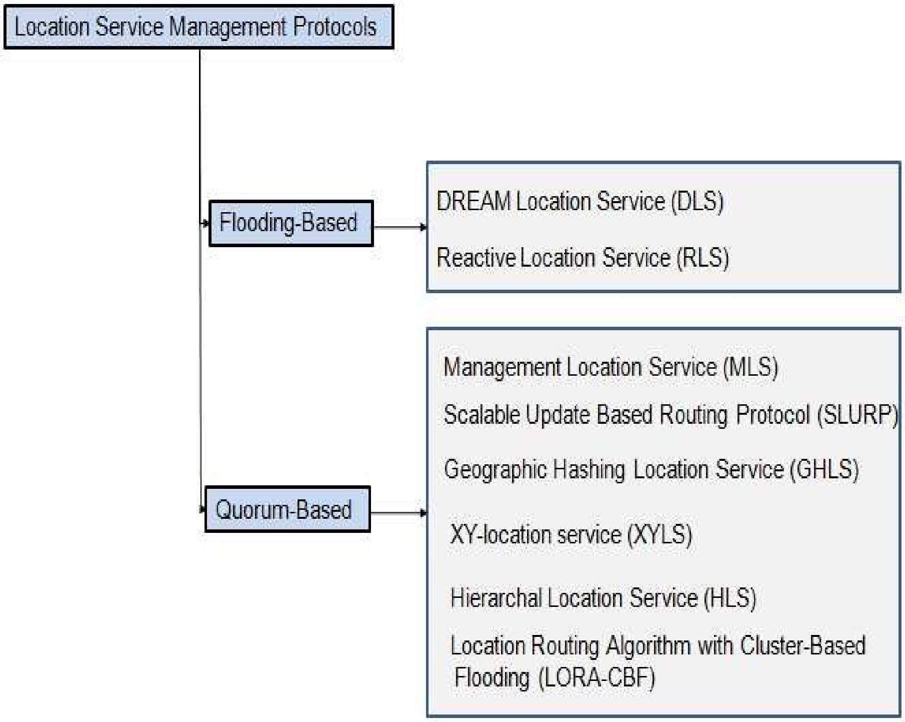


Figure 2.1: Taxonomy of the location service management Protocols

following sections reviews two of the existing flooding-based location service management protocols.

2.3.1 DREAM Location Service (DLS)

In DLS [46], each node in the network sends a location updating message which updates the location tables in all the other nodes. This message contains the coordinates of the source node, its speed and its time of transmission; the updating message is sent to nearby nodes more frequently than to faraway nodes. The frequency of transmission adapts to the distance the node has moved since its last known location. It is not necessary to keep up-to-date information about the location of faraway nodes since their relative movement is slower than that of nearby nodes. Although the transmission overhead encountered in DLS is less than that for blind flooding, it is still considered high, especially for dense networks. Because of the limited bandwidth, the scalability of this protocol is less favourable when the number of vehicles, the velocity of the vehicles, or both increase; which is typically the case in VANETs.

2.3.2 Reactive Location Service (RLS)

In RLS [47], when a node wants to communicate with a destination, it first checks its location table. If the destination location information is not available or if it has expired, then the source node floods the network with a location request packet. The location request is forwarded until it reaches the destination node. The location reply packet is returned via greedy forwarding. Since the source node waits until the location request reaches the destination, the delay in receiving the reply increases as the number of nodes in the network increases. This variation affects the scalability of the protocol, making it less suitable for vehicle-dense urban areas.

2.4 Quorum-Based Location Service

In quorum-based protocols, the location servers, or quorums, result from mapping the nodes' identifiers or geographical information to quorums in either a static or a random way [27, 28]. To design a quorum-based location service management protocol, three questions should be answered [29]: How location server(s) are chosen? How should a node update these location servers as it moves? How does a source node discover the appropriate location server(s) to retrieve the destination location information? The following section identifies some of the most popular protocols for quorum-based location service.

2.4.1 Scalable Update Based Routing Protocol (SLURP)

SLURP [48] divides the area covered by the network into rectangular regions as seen in Figure 2.2. For a given node *D*, a specific region, called home region, is selected by means of a hash function.A hash function is a well-defined mathematical function that maps the node ID into a specific region in the network. The hash function is many-to-one mapping and is known to all the nodes. All nodes that are located in the home region, such as node A in Figure 2.2, are responsible for keeping the location information of all nodes that are mapped to this region; such as node *D*. As *D* changes its position, it transmits position updates to node *A*. If another node, such as node *S* in Figure 2.2, wants to determine the position of *D*, it uses the same hash function to determine the region that may hold information about the position of *D*. *S* then sends a query to *A*, the node in the home region of *D*. Upon receiving a reply, *S* begins to forward the data packets to *D* using a geographical routing protocol. A drawback of SLURP is that it lacks the locality awareness property, since it assumes that any two nodes are equally likely to communicate with each other. In SLURP, the home region can be far away from both the source node and the destination node, for example nodes *D* and *S* are far away from node *A* in Figure 2.2, causing the path length of the updating and querying message to increase with the physical size of the network.

2.4.2 Management Location Service (MLS)

MLS assumes that any source node in the ad hoc network can start a communication session with any other node without knowing the current location of the destination node [49]. The nodes are densely distributed such that they form a connected network. Each mobile node uses a hash function that is known just for that mobile node and that maps its ID to certain positions or pointers. Each node

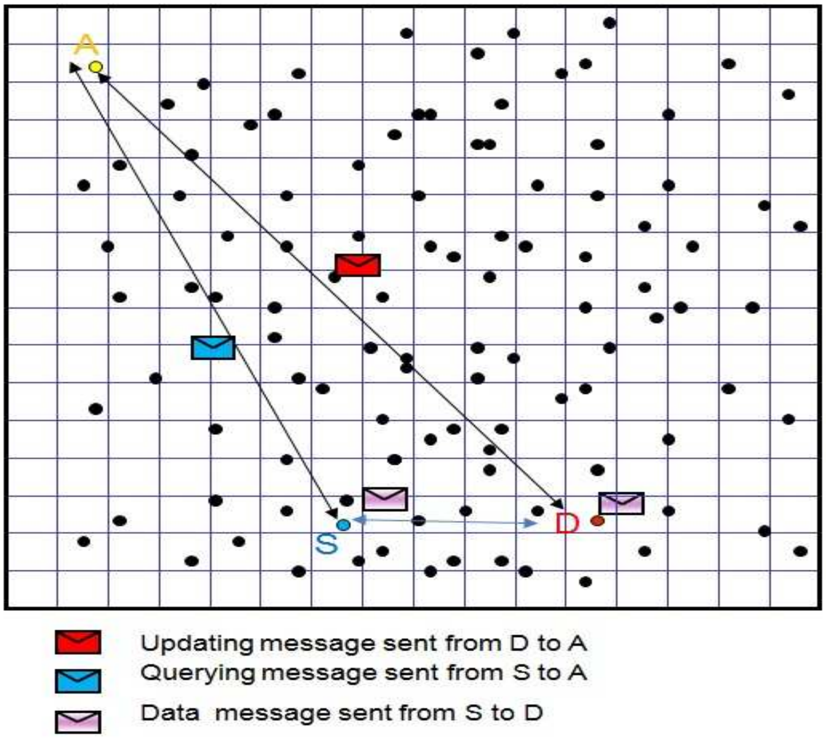


Figure 2.2: Location service management in SLURP

updates its location information on a hierarchical data structure that consists of these pointers. In MLS, the location querying algorithm works hand in hand with the location updating algorithm such that querying messages are routed to some of the destination's pointers to retrieve the location information. Since the location information is updated to a hierarchy of pointers, the updating cost is proportional to node mobility. In addition, pointers are reassigned each time the vehicle leaves one level of the hierarchical network structure. Rapidly moving vehicles can incur additional overhead by issuing frequent pointer reassignment. In a bandwidth restricted environment such as VANETs, this approach which assigns pointers that changes with the movement of nodes is often a daunting task because of the increase in communication overhead.

2.4.3 Geographic Hashing Location Service (GHLS)

GHLS is a quorum-based protocol that assumes the location server is a single node [50]. Like SLURP, GHLS uses a mapping function that takes the node ID as input, and outputs the home region location. It assumes that the node closest to this home region is the location server. Given this assumption, GHLS avoids the complexity of maintaining a hierarchy of pointers and the frequent pointer reassignment due to node mobility. Therefore, each node is required to update only one location server. GHLS tries to solve the locality awareness problem of SLURP by generating a location service region near the center of the whole network. Additionally, it deals with the problem of location server failure by asking the location server to piggy-back the locations stored in its location table in the beacon packets. Using this mechanism, the locations will be stored at all the neighbors of that location server. Even though the updating and querying overhead is lower than that of SLURP because of the central location servers, it is still considered high, given that the mapping function is used to assign the location server to a single point in the network with uniform probability. Thus, the location server for a node in the north of the network has to update its location to a location server located in, perhaps, the far south of the network, depending on the randomness of the output of the hash function.

2.4.4 XY- location service (XYLS)

XYLS assumes the network space consists of vertical and horizontal strips [51] as shown in Figure 2.3. For each destination node, such as node D in Figure 2.3, the nodes located along the north-south direction form the location servers, called the updating quorum, whereas for each source node, such as node S, the nodes that are located along the east-west direction form the querying quorum. Thus, the location updating information is disseminated in a direction such that a query can intersect an updating quorum. In other words, each node propagates the location updating packets along its column (north-south direction) in the network area, while the query is propagated along its row (east-west direction). Hence, the updating packets sent by D and the querying packets sent by S will intersect at node A. Node A has information about the location of D and then it replays to S with the location of D; after which D and S begin to communicate by sending the data packets. XYLS assumes that the communication traffic pattern in the network may be random, that is, any node may randomly initiate a communication session with any other node. Also, it assumes that the same node will traverse the entire network frequently, which rarely occurs in VANET. The traffic pattern in VANETs is mainly local, with vehicles often communicating within a local zone. Since the updating cost is proportional to the diameter of the network, the updating of location information to nodes located at the edges of the network inordinately consumes network bandwidth.

2.4.5 Hierarchal Location Service (HLS)

HLS [52] divides the network into a hierarchy of regions. The top level region, which covers the whole network, is subdivided into several regions of the next lower level until the lowest level, called cell, is reached. For a given node D, a hash function is used to assign one specific cell (home region) on each level of the hierarchy to be responsible for keeping the location information of D. As D changes its position, it updates its current position to these responsible cells or home regions. If a source node wants to communicate with D, it uses the same hash function to determine the responsible cells of D. It then sends a query to the nodes in these cells in the order of the hierarchy until it finds the current position of D. HLS requires each node to update its current position on a hierarchy of responsible cells each time it crosses the boundary of one level. In some cases, that target node may be oscillating between two points located

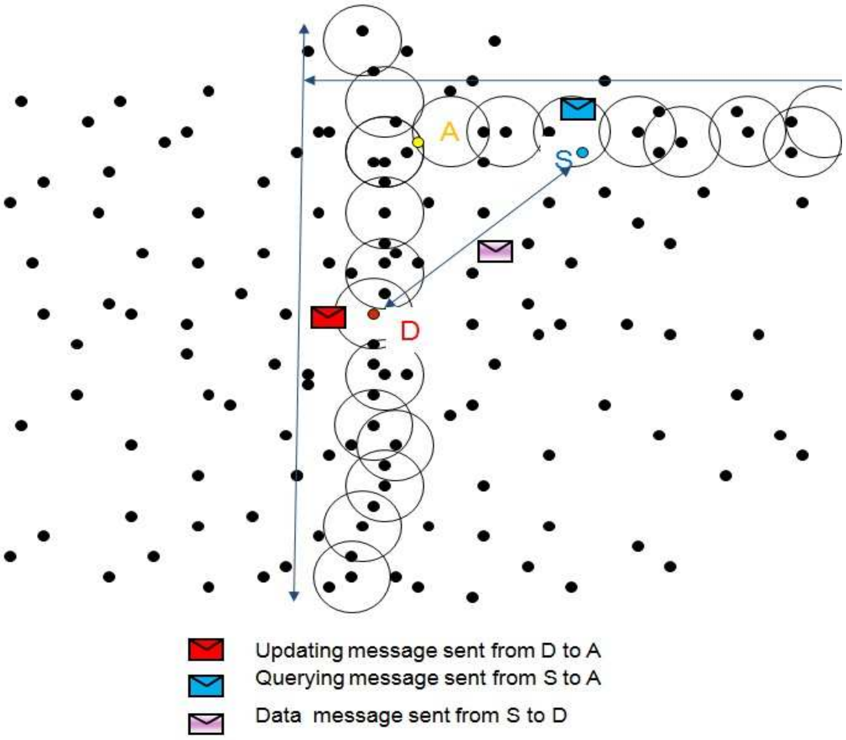


Figure 2.3: Location service management in XYLS

in two different levels which causes excessive amount of communication traffic. In other cases, all the responsible cells in the different levels of the hierarchy may need to be changed depending on the position of the target node, which introduces problems in the assignment of the hash function.

2.4.6 Location Routing Algorithm with Cluster-Based Flooding (LORA-CBF)

LORA-CBF tries to ensure the scalability by limiting the number of re-transmitters [53, 54]. The main two steps in LORA-CBF are as follows:

1- Location servers selection: The first step of the algorithm is the forming of clusters using hello messages. Each node begins as undecided, starts a timer, and sends a hello message. If it hears from a cluster head, it announces itself as a member of that cluster; otherwise it becomes a cluster head. The cluster head saves the addresses and location information of all the member nodes in its cluster. If a node hears from more than one cluster head, it changes its status to a gateway.

2- Location lookup: When a node wants to start a communication session with another node, it first checks its local table to see if it has information about the location of the destination; if yes, it forwards the data - otherwise it sends a location request packet (LREQ). Only gateways and cluster heads can retransmit the LREQ packet. Each cluster head looks in its local table to see if the destination is a member of its cluster. Upon success, a location reply (LREP) packet is sent back to the source node using geographic routing.

The main drawback of cluster-based algorithms is the additional overhead required for re-computation of membership among many clusters, meaning the re-clustering cost. This cost increases with the network dynamics, which makes clusters less and less stable. As a result, network performance degrades since it is coupled to the frequency of cluster reorganization. In a VANET, the location service management protocol should not rely entirely on clusters that are built based on the nodes' IDs. Therefore new solutions are required to deal with the highly dynamic networks.

2.5 Message Routing in VANETs

In a typical VANET, each node acts both as a host and a router where it not only is a source or a final destination but also is responsible for forwarding messages. The movement of nodes may cause the network topology to change frequently, making message routing a challenging problem. Consequently, VANETs require robust message routing protocols. Existing routing protocols fall into two categories: topology-based and geographical based routing protocols. In the topology based routing protocols, it is assumed that there is no information available about the locations of the mobile nodes. They often employ global flooding to announce location information and store this information in routing tables at each node [31,41,55,56]. In these protocols, the nodes should discover and maintain a global state of the whole network in order to route messages. Thus, these protocols are suitable for a network made of a small number of nodes moving slowly. These limitations of the topology based protocols make them unsuitable for VANETs.

In geographical based routing protocols, messages are routed based on knowledge of the geographical location of the source, intermediate nodes, and final destination. One advantage of geographical based routing protocols is that they can find a sub-optimal route from source to destination without the use of routing tables; therefore, there is no need to flood the network and store routing information at each node. The following sections review a number of the prominent existing routing protocols and discuss the drawbacks that make these protocols unsuitable for VANETs.

2.5.1 Greedy Perimeter Stateless Routing

GPSR [30] assumes that each node in the network has a local table in which all neighbouring nodes are listed by name (ID) and position. The entry of the local table is soft stated and updated after the related timer expires, where beacons broadcast information of the new neighbour or neighbours. GPSR also assumes that each source node knows the location of the destination with the aid of a location service. GPSR has two working modes: a greedy forwarding mode and a perimeter mode.

Greedy forwarding is the default mode, where the packet is forwarded to the

Algorithm 1 Greedy forwarding

```
In one Cluster do:
if (The node is the destination) then
    The node receives the data packet;
end if
if (The node is the destination neighbour) then
    The node looks into its local table to retrieve the destination location
    information;
    The node forwards the data packet to the destination;
end if
if (The node is an intermediate node but not the destination neighbour)
then
    The node forwards the packet to a neighbour that is geographically closer
    to the destination;
end if
```

node that is geographically closer to the destination. Algorithm 1 illustrates the greedy forwarding mode. The node that receives the packet checks whether it is itself the destination node; if it is not, then it forwards the packet. In order to forward the packet, the forwarding node looks into its local table to see whether it has information about the location of the destination node. A location entry of the destination means the destination is reachable within its transmission range. Consequently, the packet is directly delivered to the destination. As depicted in Figure 2.4, S is the source node. If node T is the destination node, then S relays the message directly to T, T being within the transmission range of S. Otherwise, if the destination is not a neighbouring node such as D in Figure 2.4, then S forwards the packet to a neighbour that is geographically closer to D, such as node T in this example. This process continues until the packet finally reaches its destination. It can be seen that nodes forward the packets without referring to any information about the topology of the network. Given this type of delivery, there is no need for routing tables at all.

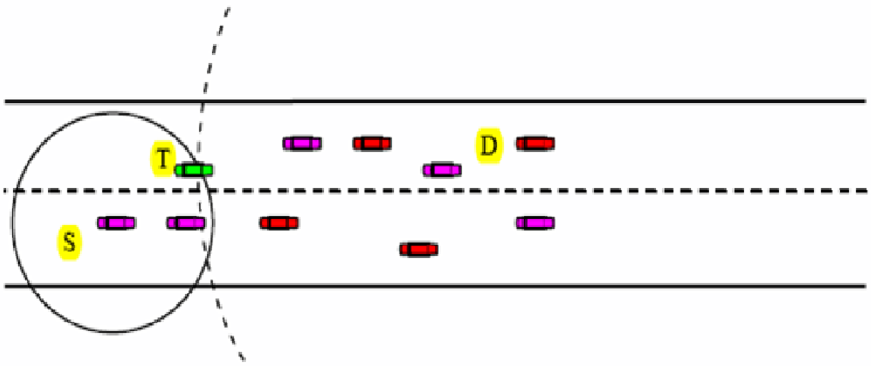


Figure 2.4: Greedy forwarding in GPSR

Greedy forwarding works well if there are no holes, meaning voids, in the network. Voids may be caused by physical obstacles like mountains or large buildings. If there is a void between the forwarding node and destination node, then the greedy forwarding may get deadlocked at the perimeter of the void. Thus, the forwarding node may not find a neighbour that is geographically closer to the destination than itself.

In such a scenario, the forwarding node switches to perimeter mode, where it chooses the neighbour as the next forwarder based on the right hand rule. As soon as that neighbour finds a node that is closer to the destination than itself is, it returns to greedy forwarding mode. However if such neighbour is not available, then the packet continues in the perimeter mode moving along the perimeter of the voids.

Because GPSR lacks information about the network topology, it can potentially go through loops. This occurs in the case of perimeter routing when the protocol routes the message in the wrong direction, resulting in performance degradation.

2.5.2 Greedy Perimeter Coordinator Routing

GPCR [57] assigns the routing decision to the nodes located at the street intersections and at the same time it uses the greedy forwarding strategy to route the message between the street intersections. Like GPSR, GPCR does not make use of road maps for routing the messages which may result in loops and introduce many hops in the route. In addition, GPCR does not take into consideration the quality of the routes nor does it have a method to select the best path.

2.5.3 Multi-Hop Routing Protocol for Urban VANETs

The Multi-Hop Routing Protocol for Urban VANETs (MURU) [58] assumes that each node has a static street map and that there is a location service that gives the source node information about the location of destinations. To find a route, therefore, the source node calculates the shortest path to the destination based on a static street map and the location of both the source and the destination. MURU provides routes that minimize the hop count. At the same time, it proposes the "expected disconnection degree" (EDD) to estimate the quality of the routes. EDD of a given route represents the probability that this route will fail during a given time period. MURU uses EDD to construct an optimal path based on predicted speed, location, and road geometry. Each node broadcasts route request packets, which are routed on paths that are constrained by node movement trajectory. MURU uses EDD to determine the next hop in the path. Since MURU uses the local information available to the forwarding node, it is susceptible to local optimum [34], which would significantly decrease the scalability of the routing protocol.

2.5.4 Delay-bounded Routing in VANETs

The authors in [59] propose a carry-and-forward algorithm to enable the vehicles to deliver messages during a limited time period which is specified by the VANET's application. It is assumed that each vehicle has access to a digital map that is preloaded with historical statistical data about the traffic on the roads. This traffic information is utilized to form the routes. One drawback of this scheme is that it assumes each node can update the statistical data about traffic conditions once it comes into contact with an access point. But given the fact that the access points can not be densely distributed in the network, they may not be found at all times. In addition, the traffic pattern changes throughout the day, resulting in frequent obsolete information that leads to wrong routing decisions.

2.6 Summary

This chapter presents the taxonomy of the location service management protocols, and it presents a number of the existing routing protocols, and discusses the drawbacks that make these protocols unsuitable for VANETs. The next chapter discusses in detail the problem statement and solution approach.

Chapter 3

Problem Description and Solution Approach

As discussed in Chapter 2, existing protocols are unsuitable for VANETs environments. This thesis proposes a location service management protocol and an adaptive message routing protocol that alleviate some of the shortcomings of existing protocols. In this chapter, the assumptions, the problem description, and the solution approach of the proposed protocols are presented.

3.1 Assumptions

This thesis considers a vehicular ad hoc network that consists of a number of vehicles moving in an urban area. The vehicles (nodes) are equipped with wireless communication devices, are powered by a chargeable battery, and have extensive storage and processing capacities. The nodes are identified by fixed identification numbers, IDs, and all the nodes have the capabilities in sending and receiving information. In addition, they can forward the information to the destination in case of multihop communication. Each node is equipped with a GPS that enables the node to know its current location. In addition, the node is aware of the position of the neighbouring nodes - those within its transmission range. Each node uses beaconing to broadcast its current location information to the neighbouring nodes. Furthermore, the nodes are equipped with an integrated electronic road map of the VANET network that has global coordinates represented by longitude and latitude. The road map includes the area covered by the mobility space of the vehicles in the network. Since cars move on the roads, it is assumed that communication routes between source and destination nodes follow road trajectories.

It is assumed that the communication pattern in the network is local, where the source and destination nodes are usually located within the same geographical area. This is a reasonable assumption since people who are located in the same geographical region tend to communicate more often with one another than with those in other regions.

3.2 Major Issues in Location Service Management

The efficiency of location service management protocols is measured by the associated querying and updating costs. One goal of this research is to design a location service management protocol that achieves minimum querying and updating cost in a large scale urban environment. Location management protocols should be efficient with respect to network size, mobility and traffic. The signalling overhead due to location management must be kept low so that the performance of the network is minimally affected. In general, there are three key factors deciding whether large scale ad hoc networks are feasible: scalability, locality awareness and fault tolerance.

3.2.1 Scalability

The VANETs resource with strong limits is the available bandwidth. When the range of the radio bandwidth is fixed, there is no way to increase the link capacity [30, 60]. In order to make the protocol scalable, the communication overhead for a single task must be reduced [61, 63]. Fast mobility of a large number of nodes induces a high volume of location querying and updating messages, causing network congestion and performance deterioration. Such network congestion results not only in dropped data packets but also in lost routing information.

Li et.al. [64] report that, in N nodes network, the capacity available per-node decreases as N increases irrespective of the protocol used for message routing. This constraint makes the feasibility of large scale ad hoc networks questionable. To remedy this scalability problem, location service protocols should be designed to scale with the increase in number of nodes. Therefore, decreasing the number of querying and updating messages, which may exacerbate the network capacity, can be a solution for large-scale networks with limited bandwidth. Otherwise, the available capacity per node will quickly approach zero as the number of nodes increases. In this research work, it is argued that employing message aggregation where the location information is aggregated is a promising strategy to ensure scalability. In other words, even if the number of nodes in the network increases significantly, the scalability of the protocol should not decrease dramatically. Consequently, better utilization of the capacity of the network can be achieved.

3.2.2 Locality Awareness

The second key factor determining the feasibility and effectiveness of large scale VANETs is the locality of traffic. In particular, the communication between any two nodes should remain local, and therefore, the average distance between source and destination nodes should remain approximately constant. Non-local traffic patterns cause a rapid decrease in the per-node capacity. Therefore, the question about the feasibility of large-scale VANET becomes a question about the locality of the traffic pattern [65]. For locality awareness, it is assumed that nodes within a fixed radius - for example, within the same cluster - are more likely to communicate with each other. Accordingly, the nodes seldom communicate with nodes located far from their local geographical cluster. Therefore, a VANET should be configured as a large network consisting of sub-networks or clusters. While connectivity across the clusters exists, the traffic that may use these inter-regional connections is kept at minimum.

3.2.3 Fault Tolerance

Most of the proposed location management protocols delegate the location service task to unit cells in the mobility space, called home regions or pointers. They assume that there is at least one node in the home region which is accessible to the network. However, the home region may become empty, a condition referred to as a void. Nevertheless, many of the existing location management protocols do not tolerate faults due to voids. The existence of voids in the network causes the throughput of the network to degrade dramatically. Consequently, the nodes should be able to make decisions to avoid the problems that arise when a home region is empty.

3.3 Location Service Management Problem in VANETs

There are many location service management protocols designed for the mobile ad hoc networks environment [66] – [68]. What distinguishes each protocol from the others is the way it selects home regions. The location of home regions is critical since it determines both the updating cost, measured in number of transmissions needed to update the location servers, and the querying cost, measured in number of transmissions needed to retrieve the location information by asking those location servers.

In quorum-based approaches, some protocols use fixed home regions (also called pointers). Fixed pointer assignment is used when the home regions are fixed irrespective of movement of the mobile nodes. Furthermore, these protocols assume that there is a hashing function that maps a mobile node's ID to its pointers in the network [48]. Such a hashing function has a random output as it maps the mobile nodes's ID to a randomly chosen region in the network. This type of random assignment assumes random communication traffic patterns and, therefore, it causes the control overhead to increase with the diameter of the network and ignores locality awareness.

To overcome the lack of locality awareness in the fixed pointer assignment, other approaches are proposed whereby pointers are reassigned according to the mobile node movement [49, 52]. In this case, the hashing function depends on both the mobile node's ID and its geographic location. Although this approach has lower control overhead, it still suffers from the problem of pointer reassignment, which sometimes leads to unbounded control overhead. On the other hand, some of the quorum based approaches select a set of home regions for a mobile node according to its geographic position only [51]. This approach assumes that any mobile node can traverse the network frequently, which is a rare case in VANETs.

In summary, existing protocols still cause congestion in the network and, therefore, do not scale well with the increase in the number of nodes and the mobility of nodes in the network. As the number and mobility of nodes in the network increase, the communication overhead encountered in the localization exacerbates the utilization of the communication channels. Location service is likely to be one of the most important necessities for VANET because of its potential in providing timely destination localization [21]. Accordingly, one objective of this research is to design a location service management protocol that improves the network performance by reducing communication traffic overhead. A quorum-based protocol, called Region-based Location Service Management Protocol (RLSMP) is proposed. RLSMP aims to build a self-organized framework for managing the network's location information. To the best of my knowledge, RLSMP is the first protocol that achieves minimum control overhead, locality awareness, and fault tolerance. RLSMP alleviates the shortcomings of existing quorum-based protocols and, at the same time, benefits from message aggregation and geographical clustering.

3.3.1 Message aggregation

Control message aggregation is a promising solution for large-scale ad hoc networks like VANETs, where the updating and querying messages of nearby nodes are aggregated in one control message. As the size of the network increases with the increase in the number of nodes, it becomes unfeasible to let each node send its updating and querying messages individually. This increases the control overhead that consumes the limited network bandwidth and results in minimal capacity per node [64].

To incorporate control message aggregation in the location service scheme, there should be a means to build distinct groups of nodes. In other words, there should be a criterion for building and maintaining clusters with minimum or no additional overhead. Nevertheless, given the problems arising from cluster management, building clusters based on the nodes' ID is a daunting task in VANETs that have a dense and highly dynamic network. Therefore, a better solution for VANETs is to build geographical clusters where nodes located in one geographic location are grouped in one cluster inside the mobility space.

3.3.2 Geographical clustering

In the proposed protocol, the location information about the nodes moving inside one geographical cluster is saved by the location server in that cluster. Position of a node inside the geographical cluster automatically implies its specific role in the network. Therefore, using the location of the nodes reduces the traffic overhead required to initialize and maintain clusters across the network. This is one of the motivations for the proposed framework, which is shown in Figure 3.1. The geographical clustering suggests the possibility of message aggregation, an essential component in reducing the communication overhead. Figure 3.2 depicts the proposed location service management protocol, where LSC refers to the location service cell (home region) that contains the location server. This system is discussed in details in Chapter 4.

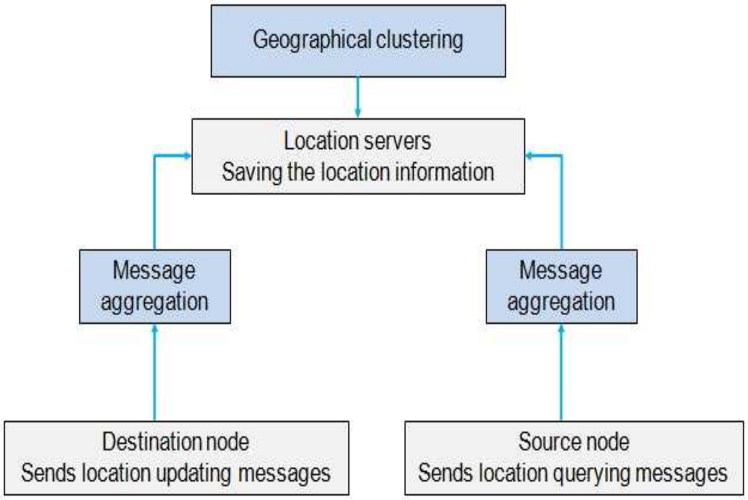
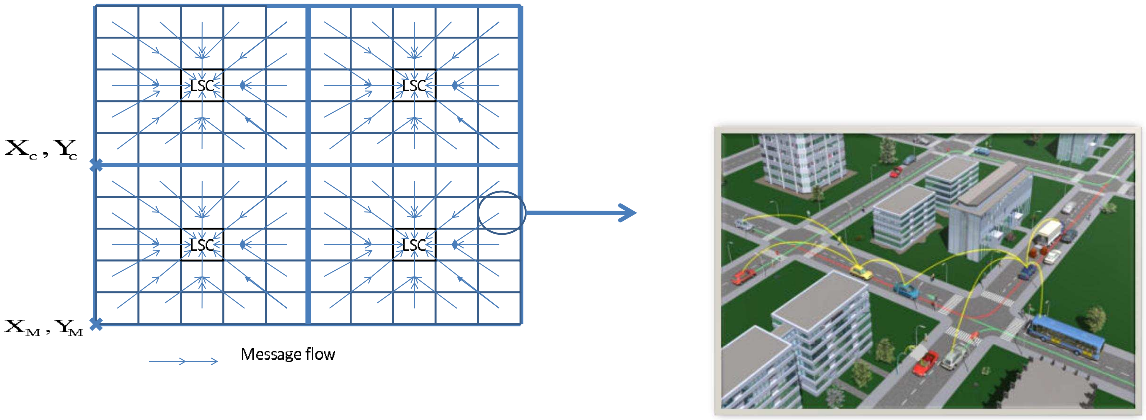


Figure 3.1: Framework of the proposed location service management protocol

In addition, the communication pattern in VANETs is predominantly local, where vehicles inside one geographical cluster are more likely to communicate with each other. Therefore, any protocol designed for VANETs should be locality aware so that it assigns the home regions while considering the local traffic pattern.



(a) (b)

Figure 3.2: System description: (a) Location information updating process, (b) Vehicle to Vehicle communication

To this end, it is assumed that the VANET is a high-density network where there is always at least one mobile node that acts as a location server. Nevertheless, there should be a mechanism to deal with the event in which the location server moves away from its location and another mobile node becomes the location server. This is a fault tolerance scheme, where the protocol tries to prevent failure due to lack of a location server. The process of electing a new location server should be achieved with minimum overhead as discussed in Chapter 4.

In conclusion, even though the nodes may move rapidly and their number may increase in the network, RLSMP is still efficient. To study the effectiveness of RLSMP, analytical expressions for costs due to location updating and querying in the context of a general two-dimensional (2-D) Random Walk mobility model are derived. Furthermore, simulations are conducted using real mobility patterns in order to evaluate the performance of the proposed scheme in real mobility situations. In addition, the total control overhead is considered as an optimization problem. Analytical and simulation results show that the proposed location service management protocol can significantly reduce the location updating cost and can yield low querying overhead compared with the performance of existing prominent schemes (SLURP [48], XYLS [51] and HLS [52]) under various scenarios.

3.4 Updating and Querying Message Routing Problem in VANETs

The second part of this work investigates the problem of routing the delay-sensitive updating and querying message in VANETs. This part investigates the use of low-cost infrastructure, namely Road Side Units (RSUs), as location servers. RSUs are inexpensive fixed infrastructure which are equipped with wireless communication devices. RSUs behave like stationary nodes and they can send and receive information by passing vehicles. RSUs store current location information about the mobile nodes in their vicinity [70, 71].

The delay-sensitive location updating and querying requirement can be mitigated using quality of service (QoS) routing between mobile nodes and the RSU. Therefore, for a given network, a QoS routing protocol that provides a high quality route under the mobility of vehicles is needed.

Several routing protocols have been proposed for VANETs. For example, GPSR [30], which is a geographical based routing protocol, has proved to be well suited to highly dynamic environments such as inter-vehicle communication on highways. However, radio obstacles, like those found in urban areas, will have a significant negative impact on the performance of this protocol [57]. GPCR [57], on the other hand, selects mobile nodes at intersections as hops to enable

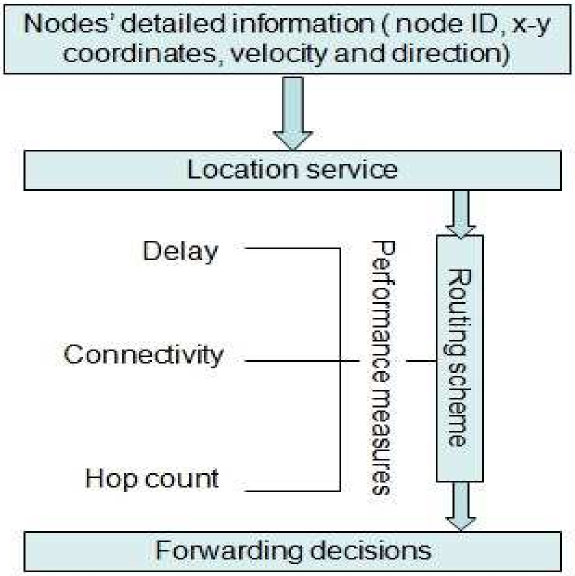


Figure 3.3: Block diagram of message routing

messages to be delivered along roads. However, this protocol does not consider the characteristics of a vehicle's movement and, thus, does not maintain stable routes - that is, it does not take into account the connectivity probability of the selected path. Many other protocols for routing information dissemination have been proposed [72] – [74]. However, their target applications are mainly delay-tolerant data dissemination and, thus, are inappropriate for routing the location updating and querying messages.

To overcome these limitations, an Adaptive Message Routing (AMR) protocol (see Figure 3.3) that adapts to the changes in the up-to-date information about the local topology is proposed. AMR is described in Chapter 5. AMR aims to find optimal or near-optimal routes that minimize end-to-end delay while satisfying QoS constraints such as connectivity probability and hop count thresholds. For example, referring to Figure 3.4, if node S wants to send location updates to the RSU located at node D, the selected route (route1, route2, or route 3) should satisfy the QoS constraints.

The QoS routing in a two-way road scenario is formulated as a constrained optimization problem, where a genetic algorithm is proposed to solve this optimization problem. analytical and simulation results show that the proposed algorithm gives an optimal or near-optimal solution and provides an interactive and effective design environment, leading to improved protocol performance compared with the well-known GPCR protocol [57].

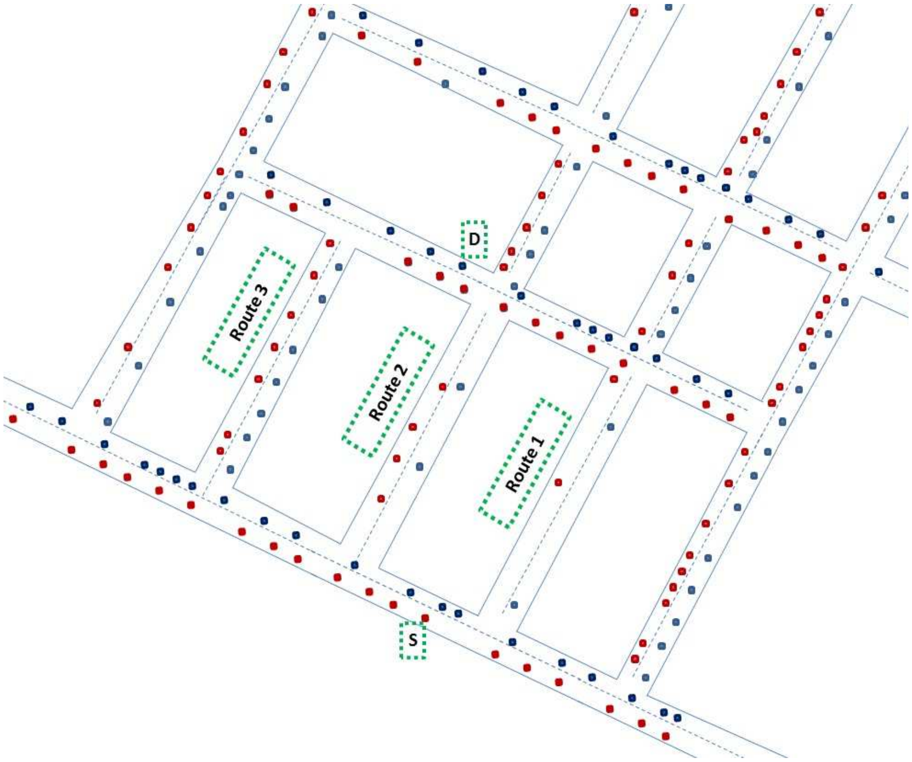


Figure 3.4: Connectivity in vehicular ad hoc network

3.5 Summary

This chapter presents the problem statement. It identifies the main shortcomings in the existing location service management and routing protocols and suggests a framework to develop efficient protocols that may be integrated in VANETs.

The next chapter presents the proposed solution for location service management in VANETs.

Chapter 4

Region-based Location Service Management Protocol (RLSMP)

4.1 Introduction

This chapter introduces the proposed Region-based Location Service Management Protocol (RLSMP) [75, 76]. The aim of RLSMP is to handle the location service management in vehicular ad hoc networks.

In existing location service management protocols, location updating and querying signalling exceeds the rest of message payload, a fact that hinders the transmission of data messages as a result of the limited channel capacity. To increase the channel utilization, RLSMP uses geographical clusterings and message aggregation. RLSMP uses the location of nodes as a criterion for building geographical clusters. That is, each vehicle automatically determines its geographical cluster while moving (as will be explained below), without any additional communication or delay. Therefore, RLSMP is robust to mobility in the sense that no overhead cost is acquired in initializing and maintaining the clusters across the network even when a vehicle leaves or joins the cluster. Furthermore, the geographical clustering suggests the likelihood of message aggregation, an essential component to reduce the number of control signals in the network.

4.2 System model and Hierarchal structure of RLSMP

The studied environment consists of roads with intersections as typically seen in urban areas. To better understand the functionalities of the protocol, the network structure depicted in Figure 4.1 is considered. The vehicles in this network are considered nodes in an ad hoc network, partitioned into virtual cells that form a virtual infrastructure. The mobility space of the nodes is viewed as a grid. Each node is aware of the location of the grid origin (X_M, Y_M), represented by longitude and latitude. Also, each cell has an origin (X_c, Y_c) with respect to the grid origin. The origin of each cell gives the cell a unique identifier (ID), which identifies its location with respect to the grid origin. Each cell has a particular node called cell leader (CL), which is responsible for aggregating the location information about all nodes within the cell. The CL acts as a location server of that particular cell. Furthermore, the grid is divided into segments, each segment containing a number of cells. The nodes inside one segment construct one geographical cluster.

Figure 4.1 shows a grid of four clusters and each cluster consists of 81 cells. Each node is aware of the size of the grid and the clusters, as well as the size of each cell. Therefore, by mapping its GPS location into coordinates in the grid, a node can determine which cell and cluster it currently resides in.

RLSMP uses geographical clustering, where nodes that reside in a geographical region are grouped into one cluster. the cell leader that is located in the central cell of one cluster is responsible for saving current location information

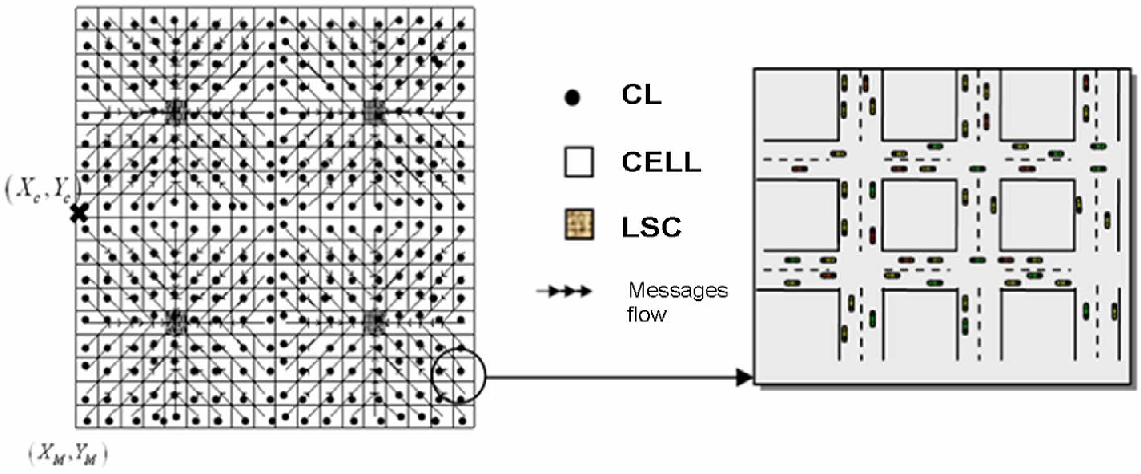


Figure 4.1: System description and LSC updating process

about all nodes that belong to that cluster. Consequently, the location information is kept locally inside each geographical cluster in the network. This arrangement explains the strategy of the protocol in denoting the central cell of each cluster as a home region or a location service management entity. This central cell is called the Location Service Cell (LSC).

Inside each cell, the cell leader CL aggregates the location information about all the nodes in its cell and forwards the aggregated control message to the LSC of its cluster. Specifically, each CL stores detailed information about the mobile nodes that it manages. As depicted in Figure 4.2, the CL in cell number 1 maintains information about the nodes in its cell. This information contains node ID, X-Y coordinates of the node location, time of the last update, and velocity and direction of the node movement. An aggregated location message is forwarded to the next CL which in turn forwards the aggregated message down stream to the LSC. The aggregated message contains summarized information about nodes IDs, cells IDs, and time stamp. It is important to mention that intelligently filtered or summarized information about the location of nodes in the network is sufficient. Therefore, it is not necessary to send the detailed information to the LSC, given that the desired level of details decreases with increase in distance and time.

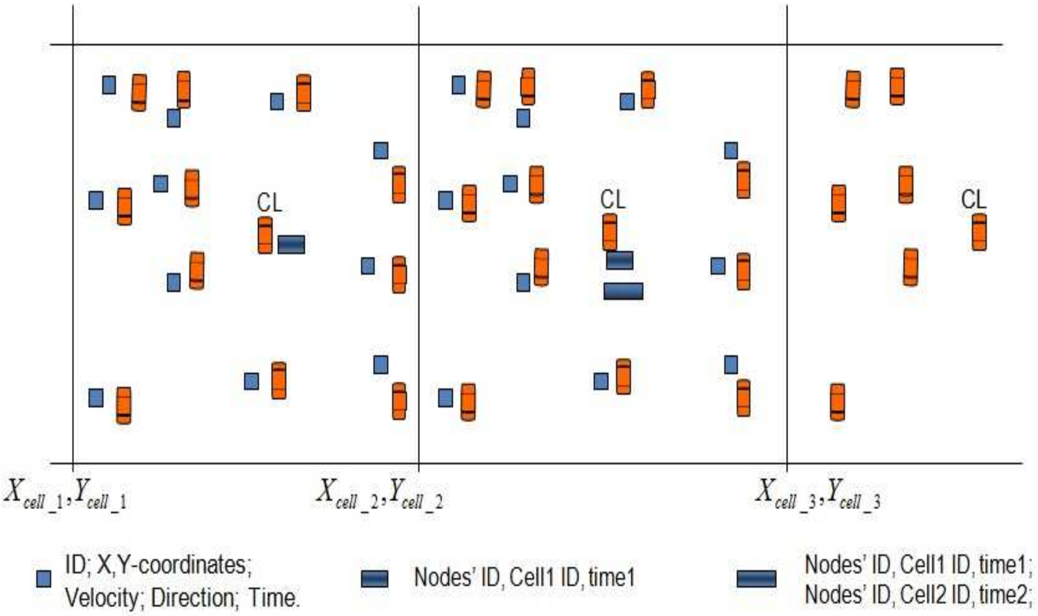


Figure 4.2: Message aggregation in RLSMP

In Figure 4.1, the arrows represent flow of message aggregation, each square representing a cell with one cell leader, and the shaded cells representing the location service cells. When a CL receives location information messages from another CL in its sub-trees, it collects and combines them into one aggregated message, which will be forwarded to its parent. Such a message aggregation step effectively reduces the number of messages in the whole network. RLSMP uses message aggregation to avoid the drawbacks of independent updating and querying messages.

In RLSMP, CL is responsible for keeping the information about the nodes in its cell. Nevertheless, there is a possibility that the CL may fail. To account for the probability of the CL failure, the CL is required to piggy-back the location information in the beacons -HELLO messages- that it sends to the one hop neighbouring nodes. It is worth mentioning here that each node in the network sends periodic beaconing messages to all one hop neighbouring nodes to inform them about its ID and its location. Therefore, each node knows its location using GPS and the location of its neighbours from the periodic beaconing messages.

RLSMP consists of two main steps: location updating, and location retrieval (querying). The cost of location updating involves two terms. The first term concerns the messages sent by the mobile nodes to the CL in order to update their locations within the cell. The second term is related to the messages sent by the CL to the LSC in order to update the location information of the nodes that reside in that cluster. The first term is referred to as the *CL updating cost* and to the second term as the *LSC updating cost*. On the other hand, the location querying cost consists of local querying and global querying. In local querying, the queries are answered by the local LSC located within the same cluster, whereas in global querying the queries are answered by a LSC other than the local one. These terms will be discussed thoroughly in the following sections.

4.2.1 Cell Leader Updating in RLSMP

As stated above, the CL tracks the mobility of nodes within its cell and keeps the mobile nodes location information up-to-date. For each movement, a mobile node updates its location to the CL as follows. The decision of CL updating is based on the mobile node's location relative to the center of the cell. By analyzing this information, the mobile node can make a decision without consulting other nodes. Specifically, using geographical based routing, the mobile node sends its current location information to the central cell element each time it crosses one cell element (see Figure 4.4(a)). This figure considers one cell with side length of l divided into 25 cell elements. Thus, the CL updating procedure is performed whenever the mobile node moves a distance d from its current location, where d depends on the transmission range T_r of the mobile node. The CL which resides in the central cell element will then receive these messages.

In the analysis, it is assumed that the VANET covers highly vehicle-dense urban area; thus it assumes that at least one node is available in the central cell element. In addition, it is assumed that the VANET has no dedicated location servers, all nodes are identical; in other words, the nodes themselves should cooperate to provide location services in a cooperative and self-organized manner. These assumptions can be relaxed when using road infrastructure, for example, the Road Side Unit (RSU) can act as a fixed CL which stores the detailed information about the mobile nodes in that cell - as discussed in Chapter 5.

The mobile nodes are aware of the current cell boundary. Therefore, each time the mobile node crosses the boundary of the current cell, it informs the old CL about its movement. At the same time, it announces itself to the new CL by sending an updating message to the center of the new cell. In the proposed protocol, the boundary of one cell is estimated by the number of rings (r) around the central cell element. r is translated into an upper bound, triggering the renewal process as shown in Section 4.3.

An important point worth mentioning is that when the CL is about to leave a central cell element, it looks into its local table and chooses the mobile node that is closest to the center of the cell and moves with minimum velocity V. This node is selected as the new CL and, hence, will receive from the old CL all the stored information about the mobile nodes located in the cell. This information is, in fact, piggybacked to the new CL by means of the Hello message sent from the old CL to its one-hop neighbours - that is, beaconing mechanism. This mechanism is used by all protocols - such as the case in SLURP, XYLS and HLS - since each node must transmit HELLO messages periodically to its one-hop neighbours to allow nodes to know the position of their neighbours. Hence, the proposed scheme does not introduce additional overhead due to the CL renewal process.

4.2.2 Location Service Cell (LSC) Updating

The Location Service Cell (LSC) is defined as the central cell of a cluster whose CL is responsible for keeping track of all the mobile nodes located in the cluster. The RLSMP protocol relies on aggregating and forwarding the location updating messages. This process is achieved by all CLs residing in the cluster and must be synchronized among them. Indeed, a time schedule, denoted by $Time\_to\_Send$, is used by each CL to know when to begin sending the aggregated message. Recall that in each cell, the CL is responsible for forwarding the aggregated packets of all mobile nodes residing in the cell. Each CL stores detailed information about the mobile nodes that it manages. This information contains the node ID, X-Y coordinates of the node location, time of the last update, and velocity and direction of the node movement. The CL forwards summarized information (node ID, cell ID and time stamp) about those nodes to the LSC of its cluster. The forwarding zone is defined by the tree structure (shown in Figure 4.1) which visits the CLs. In this figure, the arrows represent flows of message aggregation, and each square represents a cell with one CL. When a CL receives location information messages from another CL in its sub-trees, it collects and combines them into one aggregated message, which is forwarded to its parent until it reaches the LSC of the cluster. Such message aggregation process effectively reduces the number of itinerant messages in the whole network. The LSC updating algorithm is described by the pseudocode in Algorithm 2.

Message aggregation reduces signalling overhead given that the packet size is limited. As shown in Algorithm 2, whenever a packet is full, it is sent and subsequent location information received by the cell leader from other nodes is aggregated into a new packet. Putting a limit on packet size is important in order to enhance scalability since large packets may get corrupted which forces the system to send them again.

Algorithm 2 Location Information Updating Algorithm

```
In one Cluster do:
if (Cell Leader) then
    Save detailed information (nodes_ID, X, Y, V, Dir, time stamp) in local_table;
    Aggregate summarized information (nodes_ID, Cell_ID, time stamp);
    if (packet_size $\geq$ packet_size_limit) then
        start aggregation in a new packet;
    else
        Continue aggregation in the same packet;
    end if
    if (Time_to_Send) then
        Send the aggregated messages to the next Cell Leader in the downstream direction towards the LSC;
        if (next Cell Leader is the LSC) then
            Stop;
        end if
    end if
end if
```

Using this strategy, node located in the center of the LSC can act as a CL of that LSC. This “special” CL has detailed information about the mobile nodes residing in that LSC as well as summarized information about all nodes belonging to the corresponding cluster.

4.2.3 Local Querying in RLSMP

RLSMP uses message aggregation when there is a location querying. The steps involved in location querying are as follows. When a vehicle wants to communicate with another one, it forwards a query to the local CL. The CL aggregates the querying messages and forwards them to the CL located in the local LSC. If the queries are answered by the local LSC, the LSC of the cluster where the source and destination are registered, this query is called a *local query*.

Therefore, for local queries, when source node S wants to communicate with destination node D, one of the following three cases may occur:

Case one: If S and D are neighbors, node S retrieves the location of D from its local table. This information is accurate and S can forward the data directly to the location of D.

Case two: if S and D resides in the same cell. Then S sends a query to the local CL which has information about the location of both S and D and then it replies with the location information of D.

Case three: if S and D are located in the same cluster, but not in the same cell. S sends a query message to the local CL. Since the local CL does not have information about D, it forwards the query to the local LSC. As part of RLSMP location service set up, if S and D are in the same cluster the local LSC replies to the query sent by S.

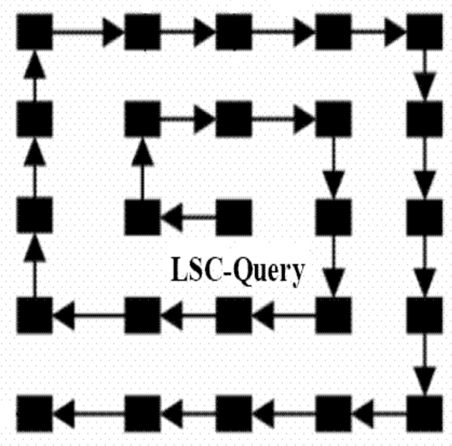


Figure 4.3: Spiral shape of the location information retrieval

4.2.4 Global Querying in RLSMP

The queries are called *global queries* if the source and destination are located in different clusters. In this case, the local LSC aggregates the queries which are sent by the vehicles residing in that cluster. The aggregated queries are then forwarded through the different LSCs as shown in Figure 4.3. They are forwarded in a spiral shape around the local LSC. The spiral visits all surrounding LSCs until it finds information about the destinations' location. The nodes inside the visited LSCs use the information stored in their own tables to determine the location of the destinations.

If it is found that one of the surrounding LSCs has information about the cell where destination node D resides, the source node S forwards data packet to the CL specified in the location reply packet received from that LSC. This location information is considered approximate. When the data packet reaches the specified CL, that CL first checks its table to see if D is co-located in the same cell. If so, the packet is successfully forwarded to the destination. By construction, that CL must have an entry for D in its location database. If it has *fresh* information about D, it further forwards the packet to D, otherwise, the packet is forwarded to the next expected CL based on D direction and velocity information. When the destination node receives the first message from the source node, it sends back an acknowledgment which contains the exact location of the destination. Therefore the first message may follow a longer bath, but the subsequent messages will be routed using a shorter path. This approach accounts for the high mobility inherent in VANET where the destination node may move while forwarding the message.

Algorithm 3: Querying Algorithm

```
 1: In the network do:
 2: if (Cell Leader) then
 3:    Save detailed information
       (srcs_ID, X, Y, V, Dir, Dst_ID, time stamp) in local_table;
 4:    Aggregate summarized information
       (srcs_ID, Cell_ID,Dst_ID, time stamp );
 5:    if (packet_size ≥ packet_size_limit) then
 6:       Start aggregation in a new packet;
 7:    else
 8:       Continue aggregation in the same packet;
 9:    end if
10:    if (Time_to_Send) then
11:       Send the aggregated message to the next Cell Leader
          in the downstream direction towards the LSC;
12:       if (Dst_ID is found) then
13:          Go to step 28;
14:       else
15:          if (Next Cell Leader is LSC) then
16:             Look up in the surrounding 8 LSCs;
17:             if (Dst_ID is found) then
```

```
18:                     Go to step 28;
19:                 else
20:                     Continue Look up by enlarging the spiral
                        around the source LSC;
21:                     if (Dst_ID is found) then
22:                         Go to step 28;
23:                     end if
24:                 end if
25:             end if
26:         end if
27:     end if
28:     Send a reply packet to the source node;
29:     if (source node receives the reply packet) then
30:         Stop;
31:     end if
32: end if
```

As mentioned earlier, communications between the nodes should be local so as to reduce the communication overhead. The use of the spiral shape in location service protocol accounts for the locality awareness property of VANETs. Indeed, the communication pattern in VANETs is almost local (i.e., within the same geographical area) [77, 78]. In the proposed scheme, this is achieved by exploring the source nodes' vicinity since the destinations search begins in clusters surrounding the local cluster of the source nodes. In addition as opposed to broadcasting or multicasting trees, the spiral shape minimizes the communication overhead and saves bandwidth. This querying algorithm of RLSMP is described by the pseudocode in Algorithm 3. Based on this fact, the communication is assumed to be more frequent inside one cluster (local queries) and less frequent between different clusters (global queries) which would enhance the efficiency of the network.

It is worth noting that, for both SLURP [48] and HLS [52], since the structure of these two protocols assumes that each mobile node in the network has its own home region, it is difficult to aggregate the messages and then send them to the different home regions. Thus, in this case, updating and querying messages are sent individually to the corresponding home regions. On the other hand, XYLS [51] can support aggregation, since aggregated updating and querying messages can be sent in the whole column or row. However, this process is not considered in XYLS [51]. Indeed, according to [51], each node will send its own updating and querying messages individually, since the mobile nodes are free to move and traverse the network frequently.

The following section presents a framework for analyzing the behavior of the location management scheme of RLSMP. A mathematical model is developed to estimate the communication overhead. The obtained results will be used, at a latter stage, to derive the protocol performance metrics.

4.3 Framework for Location Management Analysis

This section presents the derivation of analytical expressions for both location updating and querying costs of RLSMP. The location updating and querying costs are defined as follows:

Location updating cost: the number of bit forwarding transmissions from hop to hop in order to update the location information on location servers in the network.

Location retrieval (query) cost: the number of bit forwarding transmissions from hop to hop to retrieve the location information of a certain destination.

As stated before, the location updating cost in RLSMP involves two terms.

The first term is pertaining to the messages sent by the mobile nodes to the CL in order to update their locations within the cell (i.e., CL updating cost). The second term is related to the updates sent from the CLs to the local LSC in order to update the location information of the nodes that reside in that cluster (i.e., LSC updating cost). Table 4.1 describes the parameters used in the analysis.

Table 4.1: List of parameters

Parameter	Description
A	network area
u_1	the number of bytes representing the content of the updating packet
u_2	the number of bytes representing the content of the query packet
d	side length of one cell element
l	cell side length
n	cell side length multiplicity
N	number of nodes in the network
p	probability of movement to the neighboring cells
p_b	probability of boundary crossing
p_l	probability of local querying
r	number of rings of cell elements in one cell
R	number of rings of cells in one cluster
R_c	number of rings of clusters in the network
T_q	Querying period
T_r	transmission range
T_u	updating period
z	average progress in one hop
γ	node density

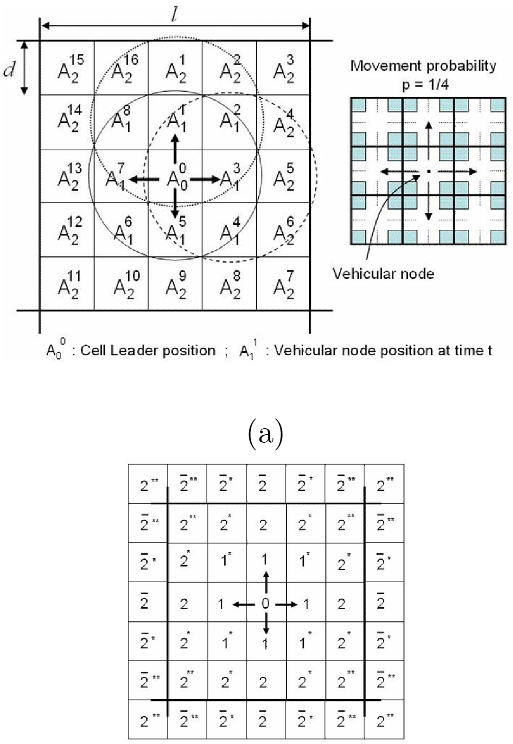
4.3.1 Cell Leader updating cost

This analysis considers the square-based 2-D model. As opposed to [79], where the focus is on n-subarea cells that form a diamond-shape cluster, the focus here is on the cell itself. Typically, a cell with side length of l is divided into $n \times n$ square cell elements that form a grid (see Figure 4.4). n is called the cell side length multiplicity. The side length d of each cell element is determined such that each node D_i in the cell element i can communicate directly (i.e., in one hop) with a node D_j located in the adjacent cell element j. In this case, $d = T_r/\sqrt{5}$.

This study considers first a general 2-D Random Walk model as in [79] – [82]. Consequently, under this model, the mobile node can move to one of the neighboring cell elements with equal probability p ($p = 1/4$). However, the general methodology presented applies to other mobility models. By using different values of the probability p_k such that $\sum_{k=1}^{4} p_k = 1$, different non-random mobility patterns can be generated. For example, for the Manhattan mobility model [83], the probability of moving right or left is 0.25; and the probability that the mobile node moves in the same direction is 0.5.

Figure 4.4(a) represents an example of the cell model used in this analysis when $n = 5$. The cell contains the CL's area surrounded by $r = \lfloor n/2 \rfloor = 2$ rings of cell elements. Each element is referenced by the ring label and its position inside that ring, which determines the mobile node's position with respect to the CL. For example, cell elements belonging to ring 1 are referenced by $A_1^j, 1 \leq j \leq 8$, those belonging to ring 2 are referenced by $A_2^j, 1 \leq j \leq 16$, and so on. To generalize, let $i = 0, 1, \ldots, r$ designate the ith ring away from the cell leader. The cell element containing CL is denoted by A_0^0. Cell elements belonging to ring i are referenced by $A_i^j, 1 \leq j \leq 8i$.

Let $X(t)$ be the mobile node's location within the cell at time t. The sojourn time of a mobile node in each element A_i^j is assumed to be exponentially distributed with a mean value of $1/\mu$. $\{X(t), t \geq 0\}$ is, therefore, a Markov process with continuous time and finite state space $E = \{A_i^j | 0 \leq i \leq r, 1 \leq j \leq 8i\}$. Recall that the main objective is to determine the mobile node's position within the cell in order to predict its evolution. According to its next location, the CL location updates cost as well as the cell boundary crossing rate corresponding to RLSMP can be calculated.



(a)

(b)

Figure 4.4: Square-cell modeling approach. (a) Cell model with $r = 2$ rings, (b) Corresponding states aggregation

The resolution of the Markovian chain $X(t)$, as defined above, is time-consuming. Moreover, this chain suffers from the state space explosion problem, mainly when the number of rings is high. To avoid this issue, a new chain $Y(t)$ is extracted from $X(t)$ by aggregating its states. In other words, all the states where the mobile node exhibits exactly the same behavior will be aggregated. Hence, the size of the state space E will be reduced. To achieve this, the symmetric property of the 2-D model is utilized. The algorithm to perform the state aggregation is described as follows.

1. Let A_i^j denote the cell element that contains the mobile node. As presented in figure 4.4(a), the state A_i^1 is chosen to be the one at the top of state A_{i-1}^1. Subsequently, each ring i consists of $8i$ elements labeled in a clockwise direction as $A_i^1, \ldots, A_i^{8i}$. Let A_i^{j*} denote the new aggregated state of the cell, where i always designates the ring reference, and $j*$ is the state label inside the ring. Since all cells of the cluster have the same size, the aggregated states A_{r+1}^{j*} located at the ring $r+1$ represent the boundary states of the cell under consideration. These states will be denoted by $\bar{A}_r^{j*}$.
2.

```
Start with i = 1;
until $(i = r)$
Repeat{
    set $A_i^{0*} = A_i = \bigcup_{0 \le j \le 3} A_i^{2ij+1}$
    For $m = 1$ to $m = i$
        set $A_i^{m*} = \bigcup_{0 \le j \le 3} A_i^{2ij+m+1} + \bigcup_{1 \le j \le 4} A_i^{2ij-m+1}$
    $i = i + 1$;
}
set $\bar{A}_r^{0*} = \bar{A}_r = \bigcup_{0 \le j \le 3} A_{r+1}^{2j(r+1)+1}$
For $m = 1$ to $m = r$
    set $\bar{A}_r^{m*} = \bigcup_{0 \le j \le 3} A_{r+1}^{2j(r+1)+m+1} + \bigcup_{1 \le j \le 4} A_{r+1}^{2j(r+1)-m+1}$
```

For instance, for $r = 2$, the following aggregated states are obtained.

$A_0^{0*} = A_0 = \{A_0^0\}$,
$A_1^{0*} = A_1 = \{A_1^1, A_1^3, A_1^5, A_1^7\}$,
$A_1^{1*} = \{A_1^2, A_1^4, A_1^6, A_1^8\}$,
$A_2^{0*} = A_2 = \{A_2^1, A_2^5, A_2^9, A_2^{13}\}$,
$A_2^{1*} = \{A_2^2, A_2^4, A_2^6, A_2^8, A_2^{10}, A_2^{12}, A_2^{14}, A_2^{16}\}$,
$A_2^{2*} = \{A_2^3, A_2^7, A_2^{11}, A_2^{15}\}$,
$\bar{A}_2^{0*} = \bar{A}_2 = \{A_3^1, A_3^7, A_3^{13}, A_3^{19}\}$,
$\bar{A}_2^{1*} = \{A_3^2, A_3^6, A_3^8, A_3^{12}, A_3^{14}, A_3^{18}, A_3^{20}, A_3^{24}\}$,
$\bar{A}_2^{2*} = \{A_3^3, A_3^5, A_3^9, A_3^{11}, A_3^{15}, A_3^{17}, A_3^{21}, A_3^{23}\}$.

For ease of use, the aggregate states were assigned numbers, as follows [see Figure 4.4(b)]:
$A_0 \to 0$, $A_1 \to 1$, $A_1^{1*} \to 1^*$, $A_2 \to 2$, $A_2^{1*} \to 2^*$, $A_2^{2*} \to 2^{**}$, $\bar{A}_2 \to \bar{2}$, $\bar{A}_2^{1*} \to \bar{2}^*$, $\bar{A}_2^{2*} \to \bar{2}^{**}$.

Theorem:

Let $F = \{A_0, A_1, A_1^{1*}, \ldots, A_i, A_i^{1*}, \ldots, A_i^{i*}, \ldots, A_r, A_r^{1*}, \ldots, A_r^{r*}, \bar{A}_r, \bar{A}_r^{1*}, \ldots, \bar{A}_r^{r*}\}$ designate the state space of the new chain $Y(t)$ obtained from the aggregation of the initial Markovian chain $X(t)$. The resulting aggregated process $Y(t)$ is also Markovian.

Proof:

For convenience, F_i represents each state of the set F and $M = (r+1) \times (r+4)/2$ represents the set size. Let Q_{init} designate the generator matrix of the initial Markov chain $X(t)$. The states of E are arranged according to the space

F partitions $\Big($ i.e., $E = \{ \underbrace{A_0^0}_{F_1},\ \underbrace{A_1^1, A_1^3, A_1^5, A_1^7}_{F_2},\ \underbrace{A_1^2, A_1^4, A_1^6, A_1^8}_{F_3},\ \underbrace{A_2^1, A_2^5, A_2^9, A_2^{13}}_{F_4},$ $\underbrace{A_2^2, A_2^4, A_2^6, A_2^8, A_2^{10}, A_2^{12}, A_2^{14}, A_2^{16}}_{F_5},\ \underbrace{A_2^3, A_2^7, A_2^{11}, A_2^{15}}_{F_6},\ \ldots \}\Big)$. In this case, the infinitesimal matrix Q_{init} can be written as a $S \times S$ matrix ($S = 1 + (\sum_{i=1}^{r+1} 8i) - 4$) and has the following form:

$$Q_{init} = \big(B_{ms}\big)_{1 \leq m,s \leq M}$$

Where B_{ms} is the block matrix corresponding to the transition probabilities between each element of the set F_m and the set F_s. In addition, these blocks verify the constant-row sum property [84]. In other words:

$$\forall\, i\ ,\ \sum_j (B_{ms})_{ij} \text{ is a constant, denoted by } c_{ms} (c_{ms} \geq 0).$$

Thus, according to [84], the resulting aggregated process is Markovian. To illustrate this result, let us revisit the example of Figure 4.4(b) where $r = 2$. As discussed before, and according to the state aggregation algorithm, the state space E of the initial Markov chain $X(t)$ can be arranged with respect to the partition F. The generator matrix Q_{init} of the process $X(t)$ can be written as in (4.1), where $a = p\mu$ and $b = -4p\mu$. Recall that the sojourn time of a mobile node in each element A_i^j is assumed to be exponentially distributed with a mean value of $1/\mu$. It is obvious to see that each block matrix justifies the constant-row sum property. The new generator matrix Q_{aggr} of the aggregated Markov chain $Y(t)$ is therefore given by (4.2). The associated state transition diagram (i.e., for $r = 2$) is depicted in Figure 4.5.

It is important to note that the proposed analytical model uses a new feature to reduce constraints on mobile movements and provide a more realistic roaming scenario with minimal assumptions, compared to that reported in [79]. Indeed, in [79], all boundary cells are aggregated into one state, called absorbing state. The model used in this work use rather a set of aggregate states (i.e., original

$$
Q_{init} = \left[\begin{array}{c|cccc|cccc|cccc|cccccccc|cccc|cccc|cccccccc|cccccccc}
A_0^0 & A_1^1 & {}^3_1 & {}^5_1 & {}^7_1 & A_1^2 & {}^4_1 & {}^6_1 & {}^8_1 & A_2^1 & {}^5_2 & {}^9_2 & {}^{13}_2 & A_2^2 & {}^4_2 & {}^6_2 & {}^8_2 & {}^{10}_2 & {}^{12}_2 & {}^{14}_2 & {}^{16}_2 & A_2^3 & {}^7_2 & {}^{11}_2 & {}^{15}_2 & A_3^1 & {}^7_3 & {}^{13}_3 & {}^{19}_3 & A_3^2 & {}^6_3 & {}^8_3 & {}^{12}_3 & {}^{14}_3 & {}^{18}_3 & {}^{20}_3 & {}^{24}_3 & A_3^3 & {}^5_3 & {}^9_3 & {}^{11}_3 & {}^{15}_3 & {}^{17}_3 & {}^{21}_3 & {}^{23}_3 \\
b & a&a&a&a & &&& & &&& & &&&&&&& & &&& & &&& & &&&&&&& & &&&&&&& \\
\hline
a & b&&& & a&&&a & a&&& & &&&&&&& & &&& & &&& & &&&&&&& & &&&&&&& \\
a & &b&& & a&a&& & &a&& & &&&&&&& & &&& & &&& & &&&&&&& & &&&&&&& \\
a & &&b& & &a&a& & &&a& & &&&&&&& & &&& & &&& & &&&&&&& & &&&&&&& \\
a & &&&b & &&a&a & &&&a & &&&&&&& & &&& & &&& & &&&&&&& & &&&&&&& \\
\hline
 & a&a&& & b&&& & &&& & a&a&&&&&& & &&& & &&& & &&&&&&& & &&&&&&& \\
 & &a&a& & &b&& & &&& & &&a&a&&&& & &&& & &&& & &&&&&&& & &&&&&&& \\
 & &&a&a & &&b& & &&& & &&&&a&a&& & &&& & &&& & &&&&&&& & &&&&&&& \\
 & a&&&a & &&&b & &&& & &&&&&&a&a & &&& & &&& & &&&&&&& & &&&&&&& \\
\hline
 & a&&& & &&& & b&&& & a&&&&&&&a & &&& & a&&& & &&&&&&& & &&&&&&& \\
 & &a&& & &&& & &b&& & &a&a&&&&& & &&& & &a&& & &&&&&&& & &&&&&&& \\
 & &&a& & &&& & &&b& & &&&a&a&&& & &&& & &&a& & &&&&&&& & &&&&&&& \\
 & &&&a & &&& & &&&b & &&&&&a&a& & &&& & &&&a & &&&&&&& & &&&&&&& \\
\hline
 & &&& & a&&& & a&&& & b&&&&&&& & a&&& & &&& & a&&&&&&& & &&&&&&& \\
 & &&& & a&&& & &a&& & &b&&&&&& & a&&& & &&& & &a&&&&&& & &&&&&&& \\
 & &&& & &a&& & &a&& & &&b&&&&& & &a&& & &&& & &&a&&&&& & &&&&&&& \\
 & &&& & &a&& & &&a& & &&&b&&&& & &a&& & &&& & &&&a&&&& & &&&&&&& \\
 & &&& & &&a& & &&a& & &&&&b&&& & &&a& & &&& & &&&&a&&& & &&&&&&& \\
 & &&& & &&a& & &&&a & &&&&&b&& & &&a& & &&& & &&&&&a&& & &&&&&&& \\
 & &&& & &&&a & &&&a & &&&&&&b& & &&&a & &&& & &&&&&&a& & &&&&&&& \\
 & &&& & &&&a & a&&& & &&&&&&&b & &&&a & &&& & &&&&&&&a & &&&&&&& \\
\hline
 & &&& & &&& & &&& & a&a&&&&&& & b&&& & &&& & &&&&&&& & a&a&&&&&& \\
 & &&& & &&& & &&& & &&a&a&&&& & &b&& & &&& & &&&&&&& & &&a&a&&&& \\
 & &&& & &&& & &&& & &&&&a&a&& & &&b& & &&& & &&&&&&& & &&&&a&a&& \\
 & &&& & &&& & &&& & &&&&&&a&a & &&&b & &&& & &&&&&&& & &&&&&&a&a \\
\hline
 & &&a& & &&& & &&& & &&&a&a&&& & &&& & b&&a& & &&&&&&& & &&&&&&& \\
 & &&&a & &&& & &&& & &&&&&a&a& & &&& & &b&&a & &&&&&&& & &&&&&&& \\
 & a&&& & &&& & &&& & a&&&&&&&a & &&& & a&&b& & &&&&&&& & &&&&&&& \\
 & &a&& & &&& & &&& & &a&a&&&&& & &&& & &a&&b & &&&&&&& & &&&&&&& \\
\hline
 & &&& & &a&& & &&a& & &&&&&&& & &a&& & &&& & b&&&a&&&& & &&&&&&& \\
 & &&& & &&&a & &&&a & &&&&&&& & &&&a & &&& & &b&&&&&a& & &&&&&&& \\
 & &&& & &&a& & &&&a & &&&&&&& & &&a& & &&& & &&b&&&a&& & &&&&&&& \\
 & &&& & a&&& & a&&& & &&&&&&& & a&&& & &&& & a&&&b&&&& & &&&&&&& \\
 & &&& & &&&a & a&&& & &&&&&&& & &&&a & &&& & &&&&b&&&a & &&&&&&& \\
 & &&& & &a&& & &a&& & &&&&&&& & &a&& & &&& & &&a&&&b&& & &&&&&&& \\
 & &&& & a&&& & &a&& & &&&&&&& & a&&& & &&& & &a&&&&&b& & &&&&&&& \\
 & &&& & &&a& & &&a& & &&&&&&& & &&a& & &&& & &&&&a&&&b & &&&&&&& \\
\hline
 & &&& & &&& & &&& & &&a&a&&&& & &&& & &&& & &&&&&&& & b&&a&a&&&& \\
 & &&& & &&& & &&& & &&&&&&a&a & &&& & &&& & &&&&&&& & &b&&&&&a&a \\
 & &&& & &&& & &&& & &&&&a&a&& & &&& & &&& & &&&&&&& & &&b&&a&a&& \\
 & &&& & &&& & &&& & a&a&&&&&& & &&& & &&& & &&&&&&& & a&a&&b&&&& \\
 & &&& & &&& & &&& & &&&&&&a&a & &&& & &&& & &&&&&&& & &&&&b&&a&a \\
 & &&& & &&& & &&& & &&a&a&&&& & &&& & &&& & &&&&&&& & &&a&a&&b&& \\
 & &&& & &&& & &&& & a&a&&&&&& & &&& & &&& & &&&&&&& & a&a&&&&&b& \\
 & &&& & &&& & &&& & &&&&a&a&& & &&& & &&& & &&&&&&& & &&&&a&a&&b
\end{array}\right] \tag{4.1}
$$

$$Q_{aggr} = \begin{bmatrix} A_0 & A_1 & A_1^{1*} & A_2 & A_2^{1*} & A_2^{2*} & \bar{A}_2 & \bar{A}_2^{1*} & \bar{A}_2^{2*} \\ b & 4a & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ a & b & 2a & a & 0 & 0 & 0 & 0 & 0 \\ 0 & 2a & b & 0 & 2a & 0 & 0 & 0 & 0 \\ 0 & a & 0 & b & 2a & 0 & a & 0 & 0 \\ 0 & 0 & a & a & b & a & 0 & a & 0 \\ 0 & 0 & 0 & 0 & 2a & b & 0 & 0 & 2a \\ 0 & a & 0 & 0 & 2a & 0 & a+b & 0 & 0 \\ 0 & 0 & a & a & 0 & a & 0 & a+b & 0 \\ 0 & 0 & 0 & 0 & 2a & 0 & 0 & 0 & 2a+b \end{bmatrix} \tag{4.2}$$

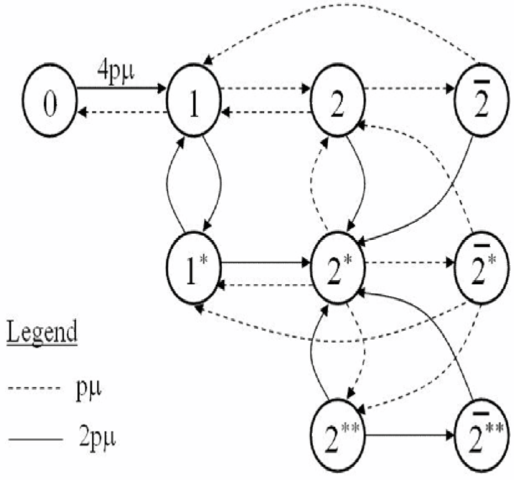


Figure 4.5: State transition diagram of the aggregated Markov chain for $r = 2$

states) to trace user movement within one cell and then use a set of special (bar) states when crossing the boundary of this cell (i.e., for the boundary cells). Consequently, the mobile node's movement is modeled to enter the original states again from the special states, once the mobile node starts moving within the new cell.

To illustrate this, let us consider again the example of Figure 4.4(b). In this figure, as long as the mobile node moves within the same cell, it is in one of the main aggregate states, i.e., $1, 1^*, 2, 2^*$, or 2^{**}. The transitions to the bar states indicate that the mobile node moves to a boundary state in an adjacent cell. When the mobile node starts moving in the new cell, for example when it moves from state $\bar{2}$ to $\bar{2}^*$ in the new cell, the proposed model binds to the normal states (i.e., original states), indicating that the mobile node has now started moving in cell elements belonging to the same new cell. This is shown by the transitions from the bar states (i.e., $\bar{2}, \bar{2}^*$ and $\bar{2}^{**}$) to the normal states in Figure 4.5.

Steady state probabilities:

Based on the state transition diagram of the aggregated Markov chain (see Figure 4.5 where $r = 2$), the steady state probability for state $F_i, (i = 1, \dots, M)$ can be obtained. Denote by Π_i and $\Pi_i^{(m)}$ $\Big(i = (0, 1, \dots, r)$ and $m = (1, \dots, i)\Big)$ the stationary probability of the system for the aggregated state A_i and A_i^{m*}, respectively. Denote also by $\Pi_{\bar{r}}$ and $\Pi_{\bar{r}}^{(m)}$ $m = (1, \dots, r)$ the stationary probability of the system for the boundary states. The balance equations for the aggregated Markov chain are obtained recursively $\forall r \geq 2$ as follows:

$$\begin{cases} \Pi_0 = p\Pi_1 \\ \Pi_1 = 4p\Pi_0 + 2p\Pi_1^{(1)} + p\Pi_2 + \alpha p\Pi_{\bar{2}} \\ \forall \ 2 \leq i \leq r-1 \\ \qquad \Pi_i = p\Pi_{i-1} + p\Pi_{i+1} + p\Pi_i^{(1)} + \beta p\Pi_{i\bar{+}1} \\ \Pi_r = p\Pi_{r-1} + p\Pi_r^{(1)} + p\Pi_{\bar{r}}^{(1)} \end{cases} \tag{4.3}$$

$$\begin{cases} \Pi_1^{(1)} = 2p\Pi_1 + p\Pi_2^{(1)} + \alpha p\Pi_{\bar{2}}^{(1)} \\ \forall \ 2 \leq i \leq r-1 \\ \qquad \Pi_i^{(i)} = p\Pi_i^{(i-1)} + p\Pi_{i+1}^{(i)} + \beta p\Pi_{i\bar{+}1}^{(i)} \\ \Pi_r^{(r)} = p\Pi_r^{(r-1)} + p\Pi_{\bar{r}}^{(r-1)} \end{cases} \tag{4.4}$$

$$\begin{cases} \forall \;\; 2 \leq i \leq r-1, \;\; j=1 \\ \qquad \Pi_i^{(1)} = 2p\Pi_i + p\delta\Pi_{i-1}^{(1)} + p\delta\Pi_i^{(2)} \\ \qquad\qquad\quad +p\Pi_{i+1}^{(1)} + \beta p\Pi_{\bar{i+1}}^{(1)} \\ \forall \;\; 2 \leq j \leq r-2 \;\text{ and }\; j+1 \leq i \leq r-1 \\ \qquad \Pi_i^{(j)} = p\Pi_i^{(j-1)} + p\delta\Pi_{i-1}^{(j)} + p\delta\Pi_i^{(j+1)} \\ \qquad\qquad\quad +p\Pi_{i+1}^{(j)} + \beta p\Pi_{\bar{i+1}}^{(j)} \end{cases} \tag{4.5}$$

$$\begin{cases} \Pi_r^{(1)} = 2p\Pi_r + p\Pi_{r-1}^{(1)} + p\Pi_r^{(2)} + 2p\Pi_{\bar{r}} + p\Pi_{\bar{r}}^{(2)} \\ \forall \;\; 2 \leq j \leq r-1 \\ \qquad \Pi_r^{(j)} = p\Pi_r^{(j-1)} + p\theta\Pi_{r-1}^{(j)} + p\theta\Pi_r^{(j+1)} \\ \qquad\qquad\quad +p\theta\Pi_{\bar{r}}^{(j+1)} + p\Pi_{\bar{r}}^{(j-1)} \end{cases} \tag{4.6}$$

$$\begin{cases} 3p\Pi_{\bar{r}} = p\Pi_r \\ \forall \;\; 1 \leq j \leq r-1, \qquad 3p\Pi_{\bar{r}}^{(j)} = p\Pi_r^{(j)} \\ 2p\Pi_{\bar{r}}^{(r)} = 2p\Pi_r^{(r)} \end{cases} \tag{4.7}$$

$$\text{where} \quad \begin{array}{ccc} \alpha = \begin{cases} 1 & \text{if } r=2 \\ 0 & \text{otherwise} \end{cases} & ; & \beta = \begin{cases} 1 & \text{if } i=r-1 \\ 0 & \text{otherwise} \end{cases} \\ \delta = \begin{cases} 2 & \text{if } i=j+1 \\ 1 & \text{otherwise} \end{cases} & ; & \theta = \begin{cases} 2 & \text{if } j=r-1 \\ 1 & \text{otherwise} \end{cases} \end{array} \quad \text{and}$$

$$\sum_{i=0}^{r} \Pi_i + \sum_{i=1}^{r} \sum_{j=1}^{i} \Pi_i^{(j)} + \Pi_{\bar{r}} + \sum_{j=1}^{r} \Pi_{\bar{r}}^{(j)} = 1 \tag{4.8}$$

Given the balance Equations ((4.3)-(4.7)) and the normalization Equation (4.8), the steady state probabilities of the aggregated Markov chain can be derived. In the following, the obtained results will be used to derive the cell boundary crossing rate and the CL location updating cost.

a) Cell Boundary Crossing Rate

Let P_b denote the probability that a mobile node crosses the cell boundary when moving within the cell. Such situation happens when the mobile node is located either at the ring r of the cell or at the boundary states, and it moves in the direction that increases the number of rings with respect to the current cell. Based on the above analysis, this probability can be given by:

$$P_b = p \times \left(\Pi_r + \Pi_{\bar{r}} + \Pi_r^{(r)} + \Pi_{\bar{r}}^{(r)} + \sum_{j=1}^{r}(\Pi_r^{(j)} + \Pi_{\bar{r}}^{(j)})\right) \tag{4.9}$$

b) CL Location updates cost

Let $CL_{updates}$ denote the cost of CL location updates when the mobile node moves within the cell. According to the node mobility, this cost can be written as follows:

$$C_1 = Cost_{intra} + Cost_{inter} \tag{4.10}$$

where $Cost_{intra}$ and $Cost_{inter}$ denote, respectively, the signaling cost of location updates when the mobile node moves within the same cell (i.e., intra-cell movement) and when the mobile node crosses the cell boundary (i.e., inter-cell movement). Using the results of Section 4.3, the expressions of $Cost_{intra}$ and $Cost_{inter}$ are given as

$$\begin{aligned} Cost_{intra} = & \sum_{i=1}^{r-1} 2p_i(2i+1)\Pi_i + \sum_{i=1}^{r-1}\sum_{j=1}^{i} 4p(i+j)\Pi_i^{(j)} \\ & + p(3r+1)\Pi_r + 2p(2r-1)\Pi_r^{(r)} \\ & + \sum_{j=1}^{r-1} p(3r+3j-1)\Pi_r^{(j)} \end{aligned} \tag{4.11}$$

and

$$
\begin{aligned}
Cost_{inter} = {} & p(2r+1)(\Pi_r + \Pi_{\bar{r}}) + p(4r+1)(\Pi_r^{(r)} + \Pi_{\bar{r}}^{(r)}) \\
& + \sum_{j=1}^{r} p(2r+2j+1)(\Pi_r^{(j)} + \Pi_{\bar{r}}^{(j)})
\end{aligned}
\tag{4.12}
$$

where

$$
r = 0.5(n-1) \tag{4.13}
$$

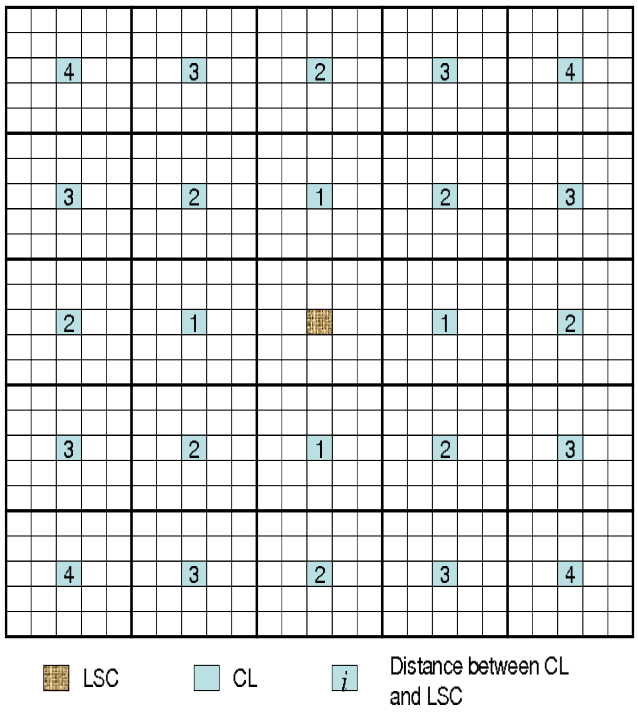


Figure 4.6: The distance between the CLs and their LSC in one cluster

4.3.2 Location Service Updating Cost

To evaluate the LSC updating cost, a new term: the number of cell rings (R) is defined. Figure 4.6 distinguishes between the number of rings of cell elements (r), and the number of cell rings (R). For example, the first 8 cells surrounding the local LSC are called the first cell ring and the next 16 surrounding cells are called the second cell ring, and so on. In this figure, the number of cell rings R is equal to 2 and the number of rings of cell elements r is also equal to 2.

To calculate the LSC updating cost, three main factors are considered as follows. Node density, cluster size and updating frequency.

Node density

The LSC updating cost is proportional to the number of nodes in one cell. Recall that in each cell, the CL is responsible of forwarding the aggregated packets sent by all the mobile nodes residing in that cell. Therefore, the LSC updating cost (in terms of bytes) is proportional to $\gamma(nd)^2u_1$, where u_1 is the number of bytes that represent the data for one specific node (see Table 4.2) and γ is the vehicles density measured in $nodes/km^2$.

Cluster size

The average distance of the different CLs from the local LSC depends on the cell side length multiplicity n and the number of cell rings R around the LSC. In Figure 4.6, the numbers assigned to the CL of each cell is the distance between that CL and the local LSC in terms of the cell side length l. Recall that $l = nd$. Therefore, to calculate the average distance, the power series is used as follows:

$$
\begin{aligned}
D_{ave} &= \frac{nd}{(2R+1)^2}\left(\sum_{i=1}^{R} 4i((2R+1)-i) + \sum_{i=1}^{R} 4i^2\right) \\
&= \frac{nd}{(2R+1)^2}\sum_{i=1}^{R} 4i(2R+1) \\
&= \frac{nd}{(2R+1)} 2R(R+1) \qquad (4.14)
\end{aligned}
$$

Updating frequency

Recall that, the aggregating and forwarding process in RLSMP are synchronized using the time schedule *Time_to_Send*. This time schedule represents the LSC updating frequency f_u.

By considering the three factors mentioned above, the LSC updating cost (in terms of bytes × hops / second) is formulated as follows.

$$C_2 = f_u \frac{nd}{(2R+1)} \frac{2R(R+1)}{z} \gamma (nd)^2 u_1 \qquad (4.15)$$

where f_u denotes the frequency of sending the aggregated updates and z denotes the average forward progress made towards a destination in the course of one transmission [48]. Note that z depends on the transmission range T_r and the average node density γ. This study assumes that both T_r and γ are constant, hence, z is constant.

4.3.3 Local querying cost

RLSMP distinguishes between local and global queries. Local queries correspond to those answered by the local LSC inside one cluster (i.e., both the source and the destination nodes are registered in the same cluster). Whereas in global queries, other LSCs will be involved in the location query process to find the destinations' IDs.

When a source node wants to communicate with a destination node, it forwards a query to the CL of the cell where it resides. The simple case happens when both source and destination nodes reside in the same cell. In this case, the CL informs directly the source with the destination location. Otherwise, a timer is triggered by the CL to begin aggregating the querying messages and forwarding them to the local LSC. Following the same steps in Section 4.3.2, the overhead introduced by this kind of queries can be determined as follows.

$$C_3 = p_l f_q \frac{nd}{(2R+1)} \frac{2R(R+1)}{z} \gamma (nd)^2 u_2 \tag{4.16}$$

where, f_q and u_2 denote the frequency of sending the queries and the average query packet size, respectively. Recall that in VANETs, the traffic pattern is assumed to be local. Therefore, in this analysis it is assumed that the probability of initiating a query is exponentially decaying, as shown in (4.17). This means that the probability of a local querying (i.e., p_l) is higher than that of the global querying (i.e., $1 - p_l$).

$$\sum_{i=1}^{2} p_i = p_l + \frac{p_l}{2} = 1 \tag{4.17}$$

4.3.4 Global querying cost

If the destination is not located in the local cluster where the source node resides, the query is called *global query*. For simplicity, let us refer to the first 8 clusters surrounding the local LSC, as the first cluster ring. The next 16 surrounding clusters correspond to the second cluster ring and so on. The global querying cost is affected by the following three factors: node density, network size, and global querying probability.

Node density

The global querying cost is proportional to the number of nodes in one cluster. Therefore, the global querying cost (in terms of bytes) is proportional to $\gamma(2R+1)^2(nd)^2u_2$.

Network size (A)

The distance traveled from the local LSC to the first LSC which is located in the first cluster ring is equal to $(nd)(2R+1)$. Thus, the distance traveled to visit the *ith* cluster ring ($i = 0, ..., Rc$) is $8i(nd)(2R+1)$, where R_c denote the total number of rings of clusters in the network; in Figure 4.7, R_c is 2. This parameter can be given as:

$$\begin{aligned}
\frac{A}{(2R+1)^2(nd)^2} &= 1 + 8\times 0 + 8\times 1 + 8\times 2 + ... + 8\times R_c \\
&= 1 + \sum_{i=1}^{R_c} 8i \\
&= 1 + 4R_c(R_c+1) \\
R_c &= \frac{-1}{2} + \sqrt{\frac{A}{4(2R+1)^2(nd)^2}} \qquad (4.18)
\end{aligned}$$

Querying frequency

Recall that f_q designates the frequency by which the nodes send the queries, and $1 - p_l$ corresponds to the probability of a global query. The cost of sending the queries from the local LSC to the surrounding LSCs can be given as follows.

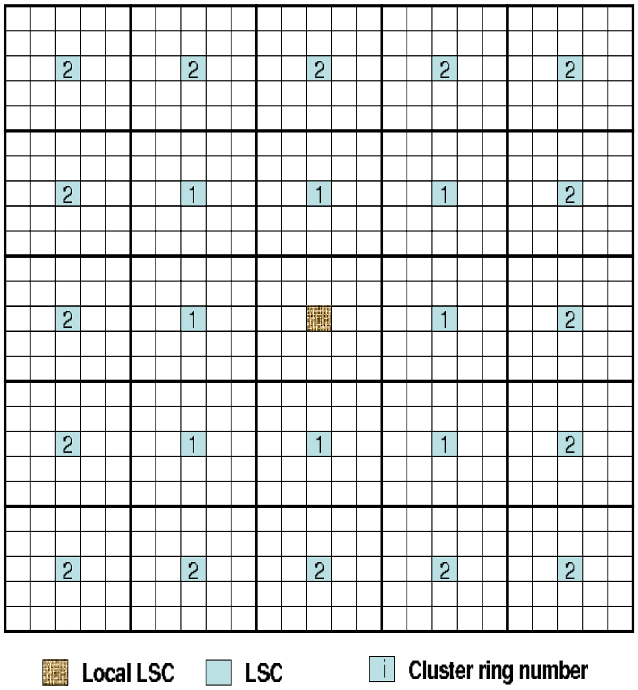


Figure 4.7: The distance between the different LSCs in the network

$$
\begin{aligned}
Cost_q &= (1-p_l) f_q \frac{8u_2\gamma(2R+1)^3(nd)^3}{z} \sum_{i=1}^{R_c} i \\
&= (1-p_l) f_q \frac{8u_2\gamma(2R+1)^3(nd)^3}{2z} R_c(R_c+1)
\end{aligned} \tag{4.19}
$$

Hence, the global querying cost C_4 is expressed as:

$$
C_4 = \frac{(1-p_l)C_3}{p_l\gamma(nd)^2} + \frac{cost_q}{\gamma(2R+1)^2(nd)^2} \tag{4.20}
$$

where the first term corresponds to the normalized cost related to the queries sent from the CLs to the local LSC, and the second term is the normalized cost of forwarding the queries from the local LSC to the surrounding ones.

Table 4.2: Details about u_1 and u_2 in number of bytes [63, 85]

Field type	u_1	u_2
$Node\ Identifier(ID)$	8	8
$Destination\ Identifier(ID)$	-	8
$Cell\ Identifier(ID)$	8	8
$Time\ stamp$	2	2

4.4 Formulating the total control overhead as an optimization problem

The cells form basic building block for the proposed location service management protocol. Therefore, their design (size) should be selected after careful studies. Side length of one cell is considered a network parameter which is critical for determining the communication traffic overhead. For example, increasing the side length of a cell increases the updating cost, whereas decreases the querying cost. Another network parameter which is critical for determining the overhead is the number of cells in one cluster. Increasing the number of cells in one cluster decreases the querying cost whereas increases the updating cost.

This section addresses the optimal values of the cell side length multiplicity n, and the number of rings of cells, R, in one cluster that minimize the total control overhead.

Based on the above analysis, the total communication cost is defined as the sum of the total updating cost, which includes both the CL updates (i.e., C_1) and the LSC updates (i.e., C_2), as well as the total querying cost that comprises the local queries (i.e., C_3) and the global queries (i.e., C_4).

Consequently, the problem is formulated as an optimization problem with the following objective function :

$$\min_{n,R} \sum_{i=1}^{4} C_i \tag{4.21}$$

subject to:

$$n \in \left[1, 2, ..., \left\lfloor \frac{\sqrt{A}}{d} \right\rfloor\right] \tag{4.22}$$

$$R \in \left[0, 1, 2, ..., \left\lfloor \sqrt{\frac{A}{(2R+1)^2 (nd)^2}} - 1 \right\rfloor\right] \tag{4.23}$$

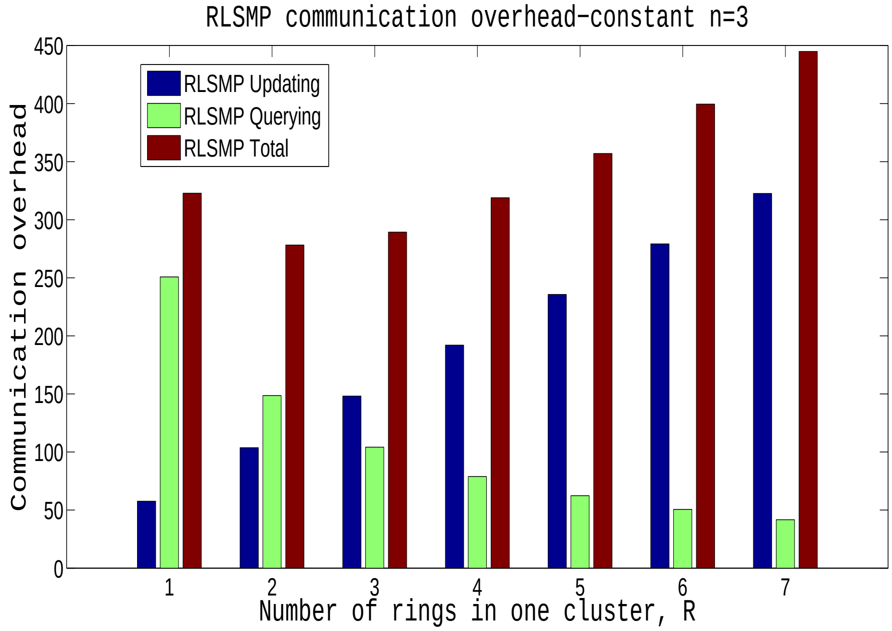


Figure 4.8: Total communication cost-constant n

The rational behind this is that, given a network of area A, the network designer can compute the optimal values of n and R by solving the above problem using total enumeration (see Figure 4.8 and 4.9). Thus, the total control overhead will be minimized and an optimal network structure will be formed.

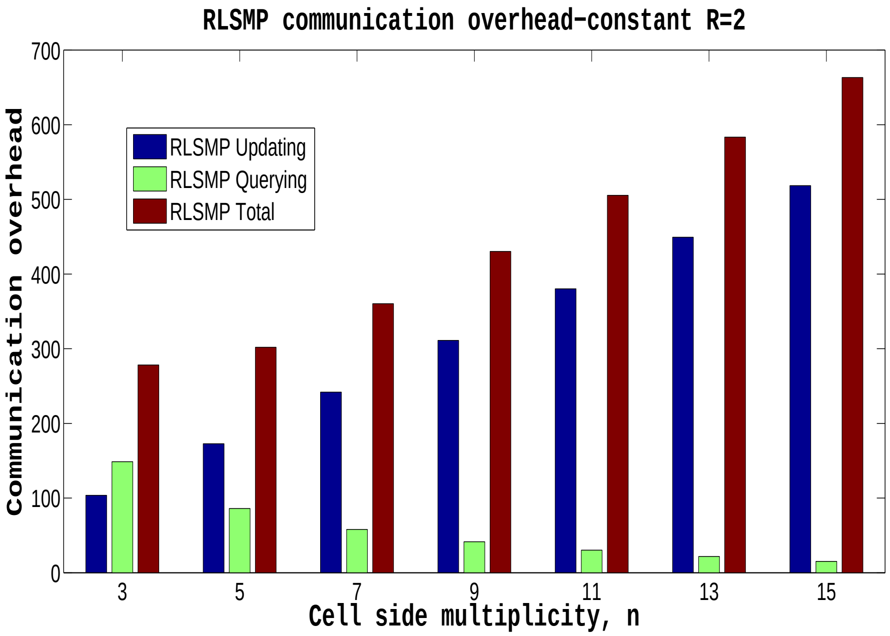


Figure 4.9: Total communication cost-constant R

The following section examines the obtained numerical and simulation results. Specifically, it compares the total signaling cost of RLSMP with that of SLURP [48], XYLS [51] and HLS [52].

4.5 Comparison between RLSMP and the existing location service management protocols

This section compares RLSMP with respect to SLURP [48], XYLS [51] and HLS [52] through both simulations and analytical approaches. To evaluate the cost of location updates and location queries by simulations, a discrete-event simulator is developed.

The simulation environment consists of a large scale wireless ad hoc network. A number of nodes N are randomly generated within the network.

In the experiments, the density γ of the nodes in the network is kept constant and equals to 8 nodes/km^2. In this case, N is varied between 160 and 4000 nodes. The mobility of nodes is simulated using either a Random Walk model or real mobility patterns. The latter case uses the mobility traces of taxi cabs in San Francisco, USA provided by the Dartmouth University. These traces record the identity, the physical location (i.e., X, Y locations) and the time that a specific car has sent its updates, for about 540 vehicles over a period of 30 days [86]. The parameter settings in the experiments are listed in table 4.3, where t_s denotes the simulation time and V is the mobile node's velocity. According to each location management policy, the location information updating and querying costs are measured.

Table 4.3: Parameter settings

Parameter	Value	Parameter	Value
t_s	1000 sec	$Time\_to\_Send$	30 $\sim$ 180 sec
T_r	250 m	γ	8 nodes/km^2
V	10 m/s	N	160 $\sim$ 4000
f_u	1/60	A	20 $\sim$ 590 km^2
f_q	1/60	r	1 $\sim$ 9

This analysis considers several scenarios by varying the cell size (Figures 4.10 – 4.13), and the network size (Figures 4.14 – 4.18). In all figures, the analytical and simulation curves with respect to the 2-D Random Walk model for RLSMP almost coincide, which illustrates the accuracy of the models.

Figure 4.10 – Figure 4.12 plot the CL updates cost as a function of the number of cell element rings, r, under the 2-D Random Walk model (Figure 4.10) and

real mobility patterns (Figure 4.11). In this case, r is varied between 1 and 9 to represent different cell sizes (i.e., 1 km × 1 km, ..., 6 km × 6 km). Recall that the cell side length multiplicity n is equal to $(2r + 1)$ and the cell side length is equal to $n\ d$. Figure 4.10 – Figure 4.12 shows that the CL updates for all protocols increase with r, since the cell size increases. The optimal cost of the CL updates depends on the CL position. Specifically, in SLURP and HLS, the CL is chosen randomly inside one cell whereas the CLs of XYLS are chosen to be all nodes located along the cell column. Hence, the CL updates are not optimal for these three protocols, as shown in Figure 4.12. RLSMP, on the other hand, reduces the updates cost since the CL is located in the center of the cell. It is worth noting that the cost of CL updates in RLSMP is equivalent to the cost of intra-cell movement when r is large as it is a dominant cost. This is shown clearly in Figure 4.11, when real mobility patterns are used. Indeed, the probability of crossing the cell boundary by a mobile node decreases with r as shown in Figure 4.13, which implies that the cost of inter-cell movements becomes non-dominant compared to that of intra-cell movements for both mobility traces.

In addition, the communication cost for all protocols (except for XYLS) using real mobility patterns is lower than that for the Random Walk model, as shown in Figure 4.12. Moreover Figure 4.13 shows that the probability of crossing the cell boundary for RLSMP is higher under the Random Walk model than under real mobility patterns. This means that the Random Walk represents the worst case scenario, which illustrates the accuracy of the proposed analytical model.

Figure 4.14 depicts the LSC updates cost of all underlying protocols as a function of the network area A under the 2-D Random Walk model. In this experiment, the network size is varied from 20 to 590 km^2. Likewise the CL updates cost, the LSC updates using SLURP, HLS and XYLS is higher than that of RLSMP. The reason is that, in SLURP, HLS and XYLS, the distance between the mobile node and its location servers grows dramatically with the

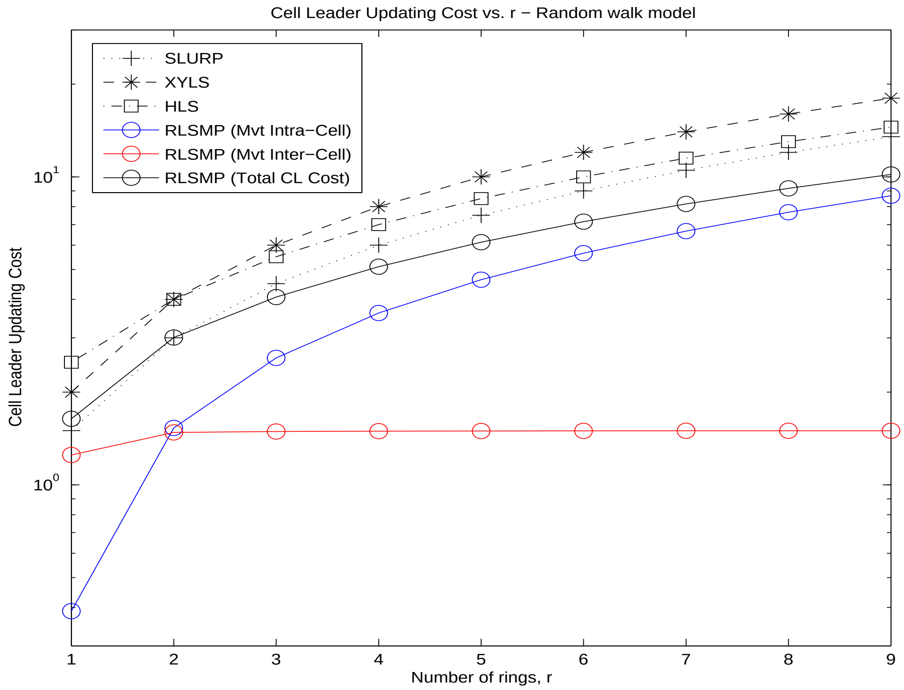


Figure 4.10: Cell Leader updating cost: random walk mobility model

network size since the location servers are randomly chosen in the network for the SLURP and HLS cases and along the network column for the XYLS case. This allows longer forwarding path to be formed and more byte transmissions. Note that for HLS, the updating cost is less than that of SLURP and XYLS, since it does not require the mobile node to update the faraway home regions unless it crosses the higher level boundary. On the other hand, when considering RLSMP, the optimal values of the side length multiplicity n and the number of cells per cluster R are used for each value of A. In this regard, the optimal solution for the optimization problem described in Section 5.4 and that minimizes the total control overhead.

Figure 4.15 presents the location information query of all protocols as a func-

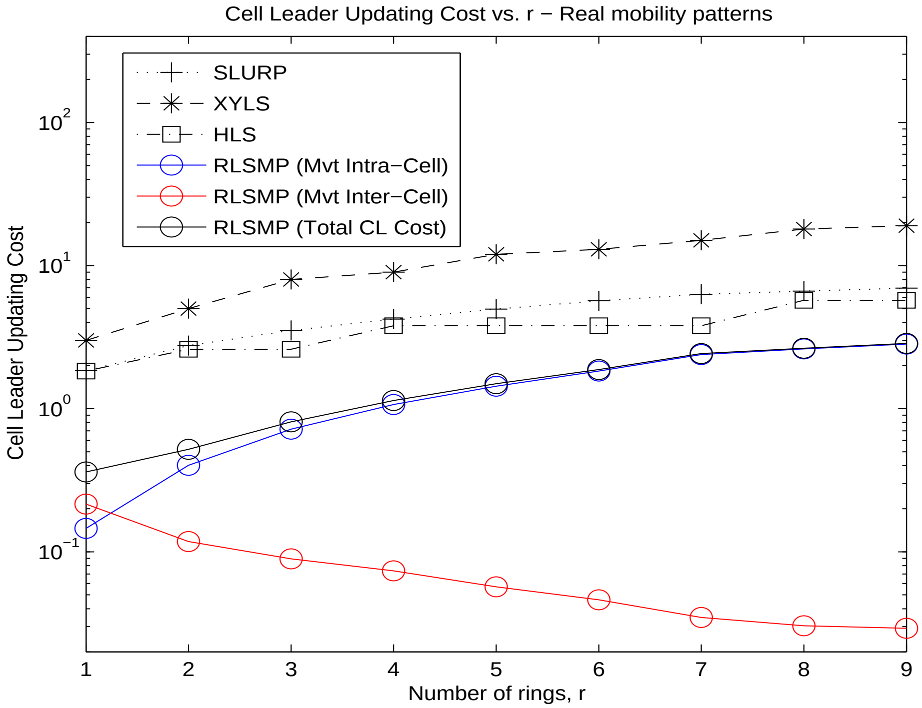


Figure 4.11: Cell Leader updating cost: real mobility patterns

tion of the network size under the 2-D Random Walk model. In this experiment, the network size A is varied from 20 to 590 km^2. For each value of A, the same values for n and R that minimizes the total overhead is used. Figure 4.15 shows that RLSMP reduces also the location queries cost since it incorporates the locality awareness. Indeed, the queries are first forwarded to the LSCs in the mobile node's vicinity. In SLURP, the distance between the source and the destination's location server increases with the network size since it is assumed that any two nodes in the network are equally likely to communicate with each other. Considering the XYLS scheme, the node can move in the whole network frequently. This results in an increase of the distance between the mobile node and the quorum (i.e., the cell where the updating and querying intersect) when the network size increases. Considering HLS, it is worth noting that even though the source

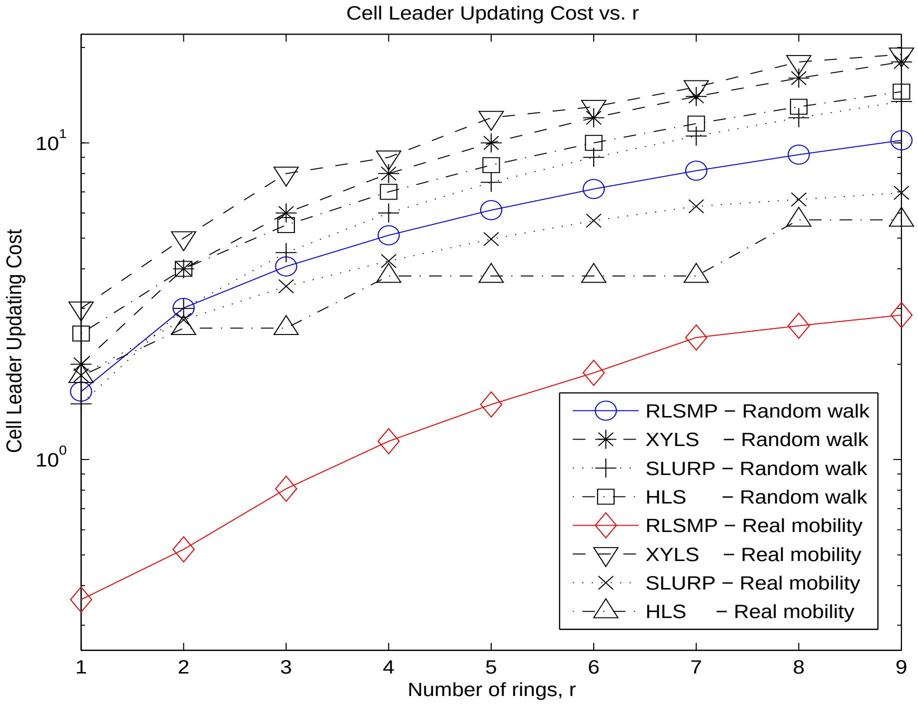


Figure 4.12: Cell Leader updating cost: comparison between both mobility models

and destination nodes are geographically close to each other, they are assumed to be located in different levels due to the virtual boundaries of the grid. This forces the query to travel to the destination pointers that are randomly located in the higher levels. As a result, the querying cost increases, as shown in Figure 4.15.

Figure 4.16 compares the total overhead of all protocols when using the optimal RLSMP configuration. It confirms that RLSMP stands out as the best solution from the communication overhead perspective. Figure 4.17 shows the impact of the nodes' velocity V on the total communication overhead; the increase of the nodes' velocity affects mainly the CL updating cost inside one cell,

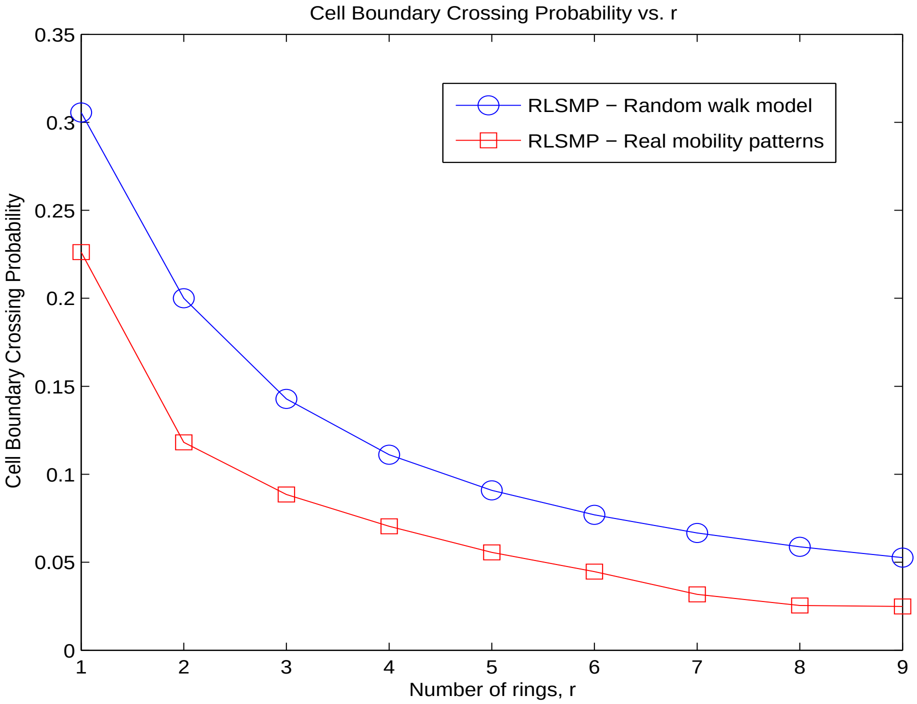


Figure 4.13: Cell Boundary crossing probability

since in this case the nodes will send more frequent updates to the cell leader. It is worth noting that, in the simulations, the nodes move with velocity that does not exceed 70 km/hr in the urban areas, and then the re-election of the CL occurs after 30 sec which is a reasonable time to transfer the location information to the new cell leader.

Figure 4.18 shows the average query response delay of all protocols as a function of the network size under the 2-D Random Walk model. In this experiment, $Time\_to\_Send$ is set equal to 60 sec. Each mobile node sends randomly 1 query per $Time\_to\_Send$ interval time. Figure 4.18 shows that the average response time per query for RLSMP is lower than that of the remaining schemes, since the CL will receive the response of all aggregated queries at the same time.

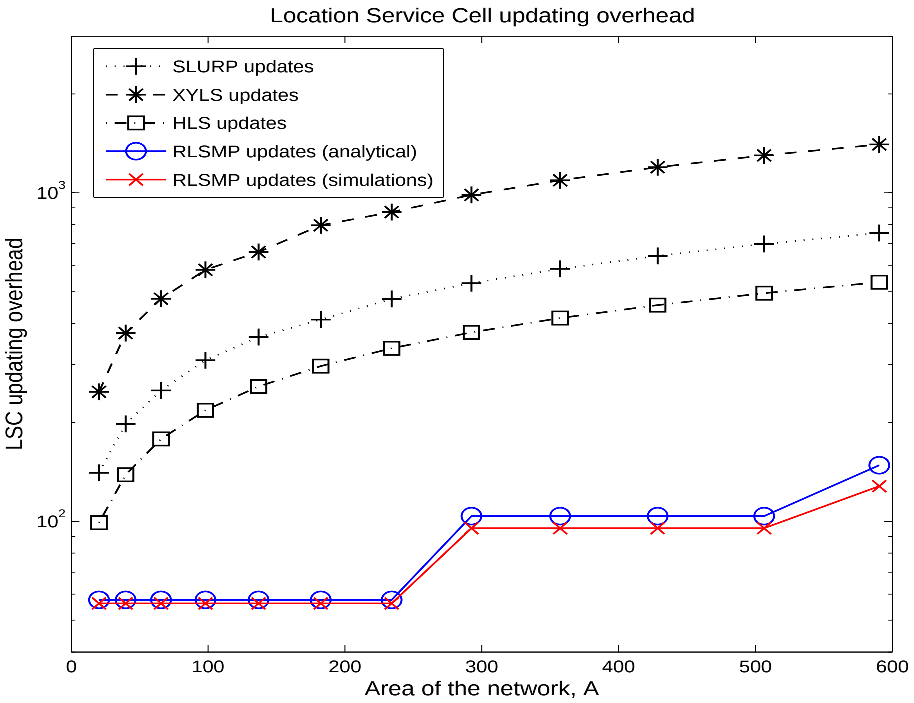


Figure 4.14: LSC updating cost

Figure 4.19 shows the impact of the time schedule $Time\_to\_Send$ on the average query response delay for RLSMP under both mobility models (i.e., 2-D Random Walk and real mobility patterns). In this experiment, $Time\_to\_Send$ is varied between 30 and 180 *sec*. From this figure, it can be noticed that the average query response delay increases with the increase of $Time\_to\_Send$ for both mobility models, since the queries will be more delayed at the level of the CL node to perform message aggregation. In addition, it is noticed again that the Random Walk mobility model represents the worst case scenario.

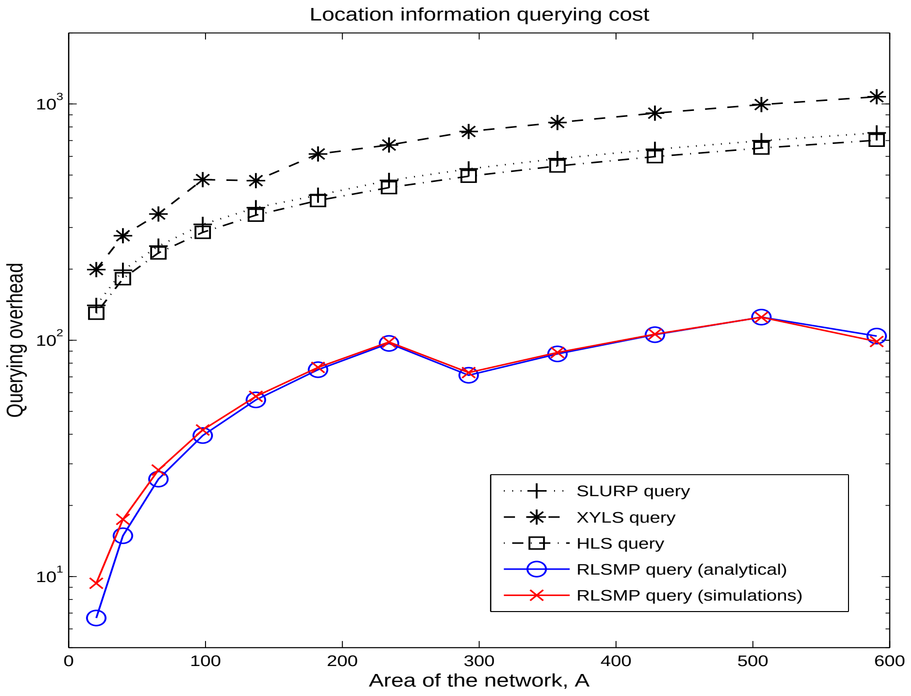


Figure 4.15: Location information query cost

4.6 Fault Tolerance

RLSMP delegate the location service task to a particular node, the cell leader, that is located in the central element of each cell. The cell leader or the location server LS is responsible for keeping the location information about the nodes in its cell. Many existing location service management protocols do not account for the case when the LS moves from the home region. Nevertheless, RLSMP propose a mechanism to deal with the event when the current location server moves away from its location and another mobile node becomes the location server, so that the location server should not fail to answer the queries.

This mechanism is called fault tolerance where the protocol tries to prevent failure due to lack of a location server. It requires the current location server

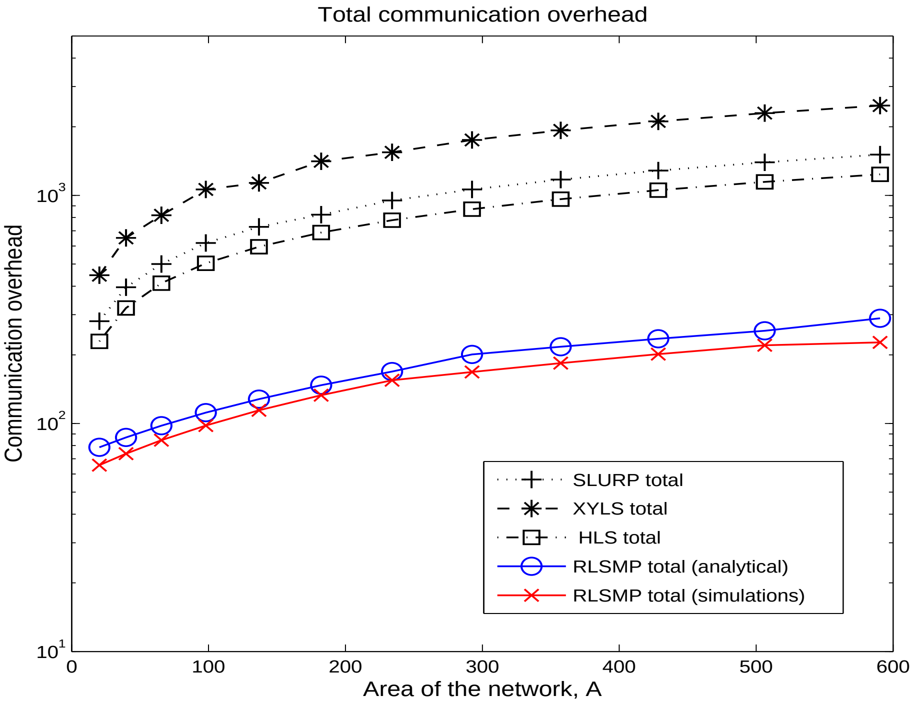


Figure 4.16: Total communication overhead with optimal values of n and R

to elect a new location server among the candidate nodes and to forward the location information to that new location server. The process of electing a new location server should be achieved with minimum overhead. In fact, since the location server has the location, speed and direction information about the nodes in its cell, it can elect the new location server without additional overhead.

This analysis uses two cost functions to select the new LS: the first cost function is related to the candidate LS_i's location information which is denoted as C_{1i}. The second cost function is related to the candidate LS_i mobility characteristics denoted as C_{2i}. Therefore, the candidate LS_i will be elected based on minimizing the summation of the two cost functions:

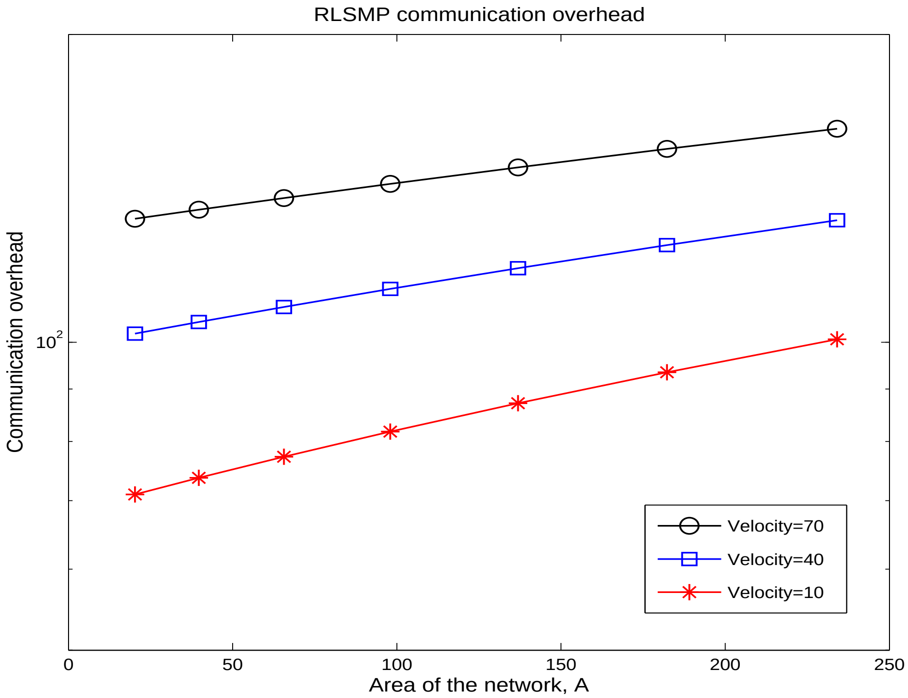


Figure 4.17: Communication overhead with different nodes' velocity

$$\min_{LS_i} \sum_{k=1}^{2} C_{ki} \tag{4.24}$$

The following section presents the analytical formulation of the two cost functions used to select the new location server.

4.6.1 Candidate LS location information

The location of the candidate LS_i determines the distance between LS_i and the mobile node N_j, which is given as:

$$D_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{4.25}$$

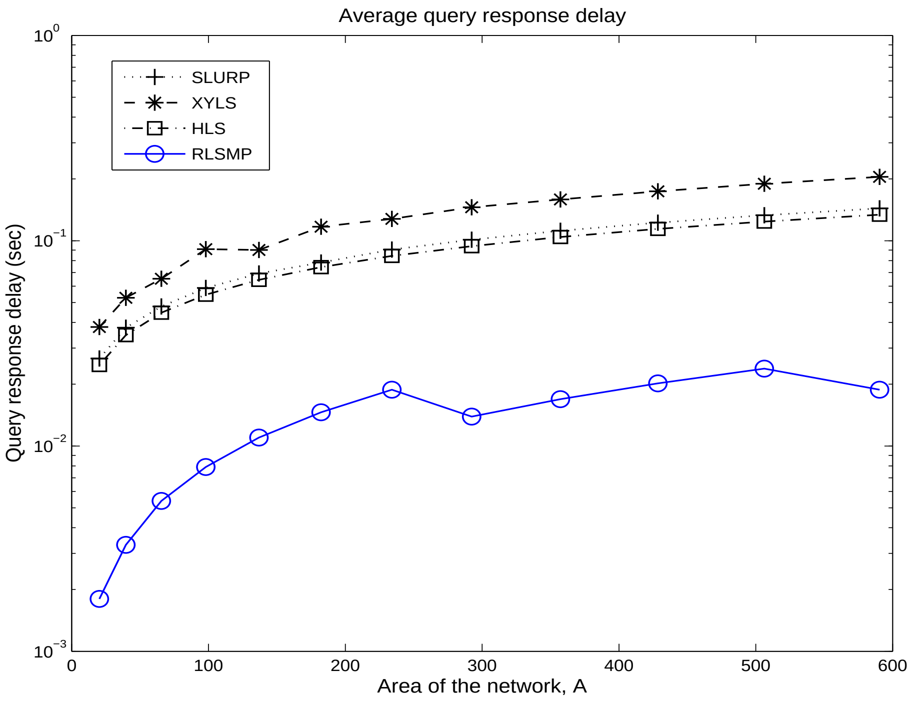


Figure 4.18: Average query response delay

where (x_i, y_i) are the coordinates of the location of candidate LS_i and (x_j, y_j) are the coordinates of the location of N_j where $j = 1, 2, .., N_c$. N_c is the total number of nodes that reside in that cell.

The total number of hops between the candidate LS_i and the rest of nodes in one cell is:

$$C_{1i} = \frac{1}{N_c} \sum_{j=1}^{Nc} \frac{D_{ij}}{z} \tag{4.26}$$

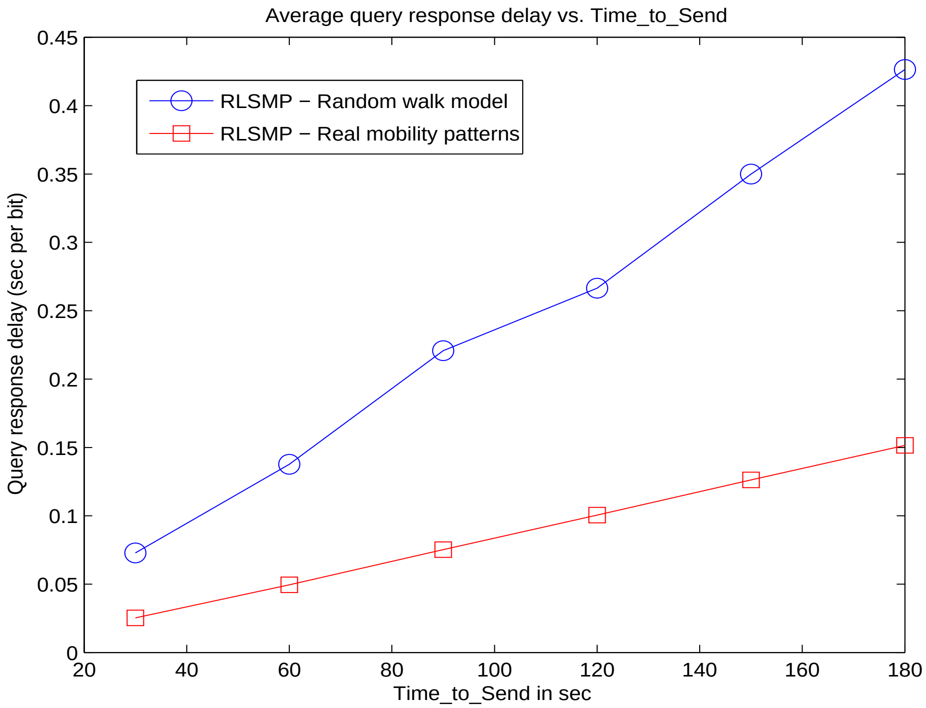


Figure 4.19: Impact of *Time_to_Send* on the average query response delay for RLSMP

4.6.2 Candidate LS mobility

The second cost function represents the effect of mobility of the node N_j on the number of hops $C2_{ij}$ between the mobile node N_j and the candidate LS_i. Therefore, if N_j moves with an average velocity $\widehat{v_j}$ and LS_i moves with an average velocity $\widehat{v_i}$, the number of hops between N_j and LS_i will change according to the following formula:

$$C_{2i} = \frac{1}{N_C} \sum_{j=1}^{Nc} \frac{D(t)_{ij}}{z} \tag{4.27}$$

where $D(t)_{i,j}$ is given by:

$$D(t)_{ij} = \sqrt{(x_i + (\widehat{v}_i cos\beta_i)t - x_j - (\widehat{v}_j cos\alpha_j)t)^2 - (y_i + (\widehat{v}_i sin\beta_i)t - y_j - (\widehat{v}_j sin\alpha_j)t)^2} \tag{4.28}$$

For example, Figure 4.20 illustrates how to specify the mobility characteristics of the mobile nodes with respect to LS_i.

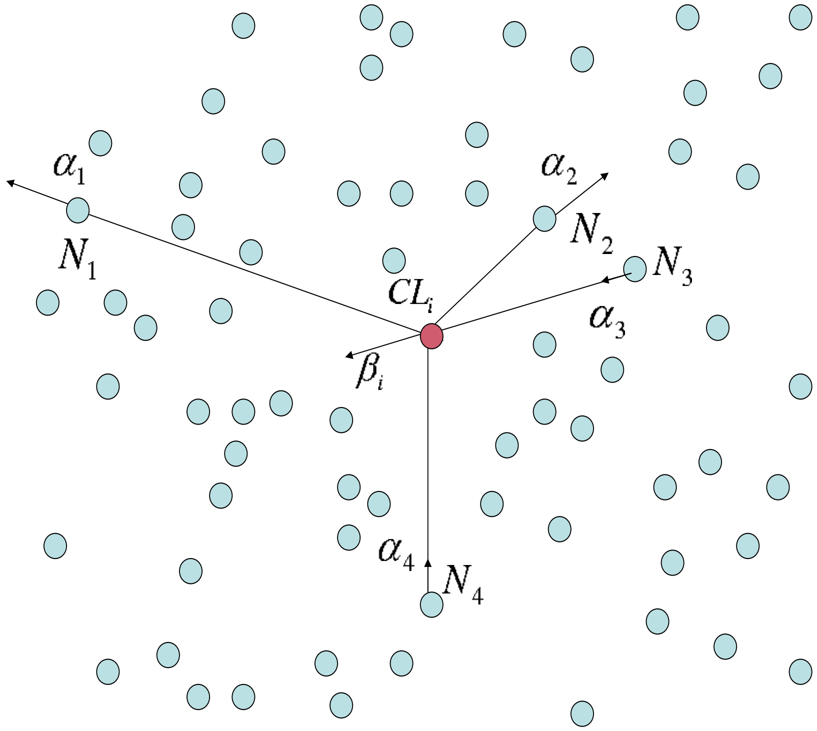


Figure 4.20: Illustration of the direction of nodes' movement with respect to the candidate LS

4.6.3 Experimental results

The simulations use a cell of size 2km x 2km and assume there is a location server in the center of the central cell element. The candidate LSs are identified,

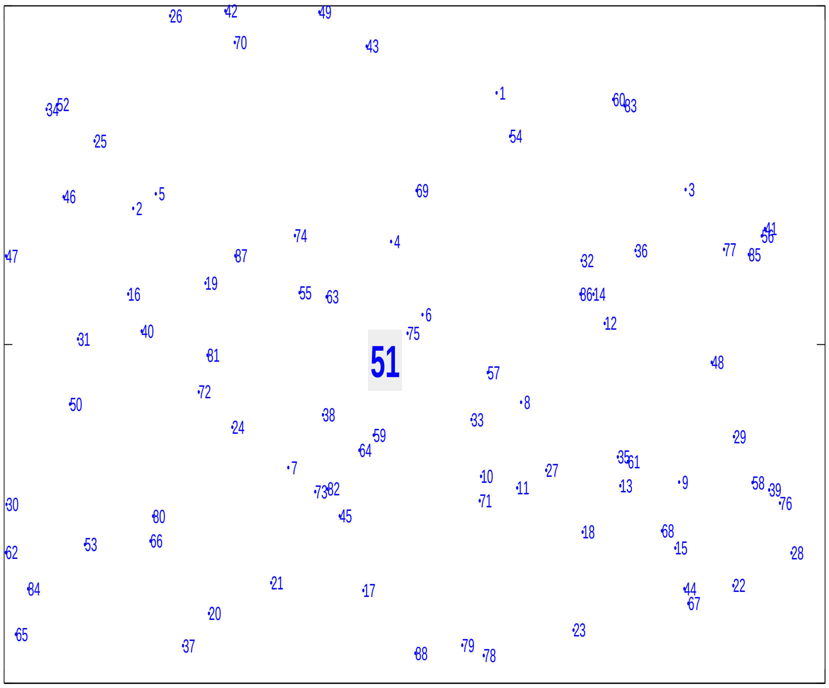


Figure 4.21: The new elected LS at time period 1

so that the new LS will be elected as soon as the current LS leaves the central cell element. The experiments use an average node density of 30 $nodes/km^2$. The nodes move with an average velocity of $50km/hr$. Therefore, the elected LS is shown in three different times (see Figure 4.21 - Figure 4.23). Based on location and mobility of the candidate location servers, the one that has a location that minimizes the total hop count between itself and the rest of nodes, and at the same time, moves towards a new location that minimizes the hop count is elected to be the new LS. For example in Figure 4.21 the elected LS is node 51, Figure 4.22 the elected LS is node 63, and for Figure 4.23 node 34 is the elected LS. As it is mentioned, this strategy is important to protect the location information

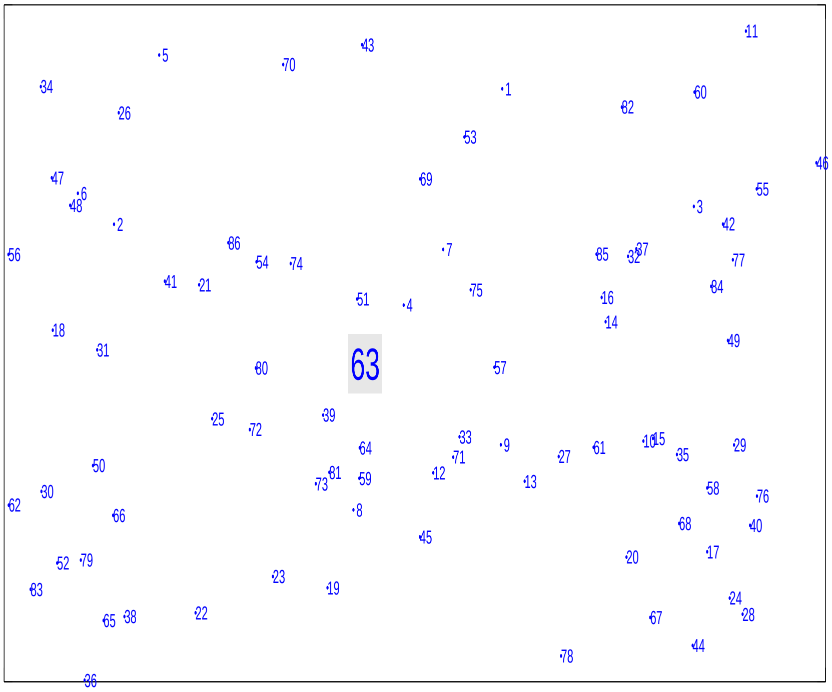


Figure 4.22: The new elected LS at time period 2

from being lost due to the mobility of the current location server.

4.7 Summary

This chapter described a new location service management protocol, called RLSMP that features minimum overhead, locality awareness and fault tolerance in VANETs. In addition, it presented both analytical and simulation approaches to compare the proposed scheme with existing solutions (SLURP, HLS and XYLS). The proposed scheme stands out as a promising candidate for large scale wireless ad hoc network such as VANETs. The following chapter addresses the challenging

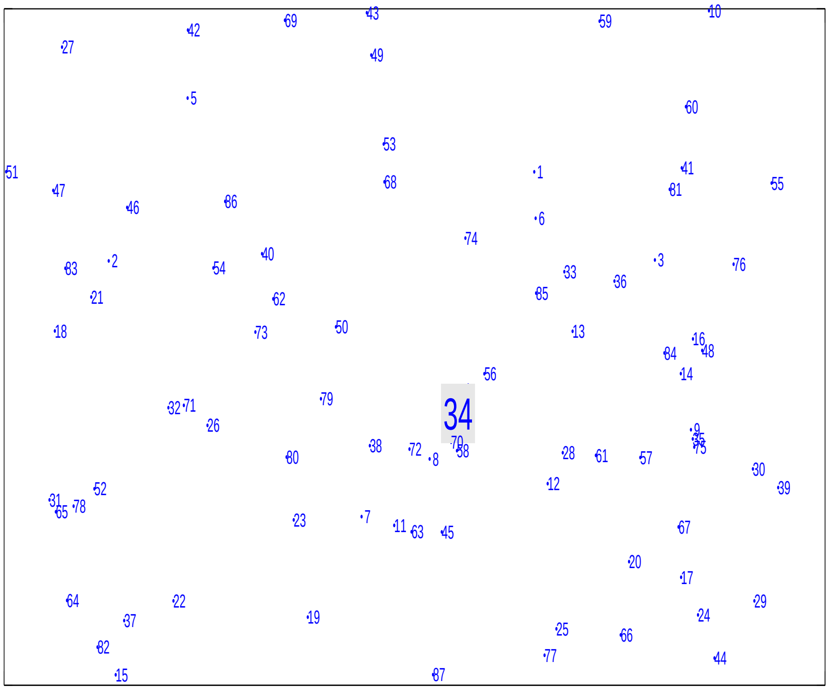


Figure 4.23: The new elected LS at time period 3

problem of QoS routing in VANETs.

Chapter 5

Adaptive Message Routing

5.1 Introduction

As VANETs continue to gain momentum for large scale adoption, there is a persuasive need to support Quality of Service (*QoS*) message routing in such networks. This chapter presents a promising solution to route the messages while considering the *QoS* constraints.

5.2 Adaptive Message Routing protocol

This thesis is concerned with the efficient QoS routing that ensures minimum end-to-end delay while considering the QoS constraints which are: maintaining a threshold for connectivity probability and maintaining a threshold for hop count in each selected path. This chapter proposes an Adaptive Message Routing (AMR) approach [87].

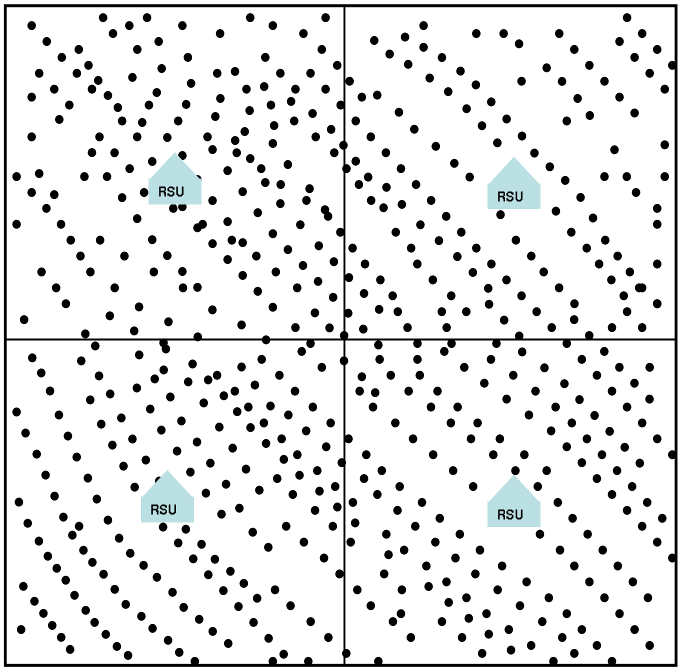


Figure 5.1: Network structure

5.2.1 System model of AMR

AMR models the network as a grid with fixed cell size across the area (see Figure 5.1). Each mobile node is aware of the location of the grid origin (X_M, Y_M), represented by longitude and latitude. Also, each cell has the origin (X_c, Y_c) with respect to the grid origin. The origin of each cell gives the cell a unique identifier (ID) that identifies its location in regard to the grid origin. In the center of the cell there is a fixed infrastructure, which is a Road Side Unit (RSU).

In AMR, the RSU located in the center of each cell acts as a *location server* (see Figure 5.1). Therefore, each RSU is responsible for maintaining detailed and up to date location information about the mobile nodes that reside in that cell. Specifically, this information consists of the mobile node ID, the transmission range T_r, $X-Y$ coordinates of the mobile node location, time of the last update, and velocity and direction of the movement of the mobile node.

Each vehicle reports its location information to the RSU each time it moves one transmission range faraway from its previous location. This enables the local RSU to have a view of the local network consisting of the nodes it manages. Doing so, a set of routes can be constructed between the RSU and the mobile nodes. These routes can be used by the mobile nodes to send location updating and querying messages to the local RSU. Nevertheless, if these routes consist of intermediate mobile nodes, these routes can not be considered to be stable due to intermediate nodes mobility. To increase the stability of the routes, AMR builds routes based on intermediate and adjacent road intersections towards the RSU as described in Figure 5.2; these routes are called backbone routes. Figure 5.2 shows two alternative backbone routes between a mobile node and the RSU.

As shown in Figure 5.2, AMR models the road map as a graph with a set of intersections and road segments. In this figure, each path on a backbone route begins from the intersection adjacent to the source node, adds another intersection that is in the way towards the RSU, and ends at the intersection adjacent to the RSU. Thus, AMR aims to determine the best backbone route between nodes and the RSU in order to forward location updating and querying messages.

The key building blocks of AMR are depicted in Figure 5.3, where the mobile nodes send updates about their detailed information to the local RSU, which in turn decides on the best backbone routes. Accordingly, the local RSU informs the nodes about the backbone routes that these nodes can use to forward their updating and querying messages.

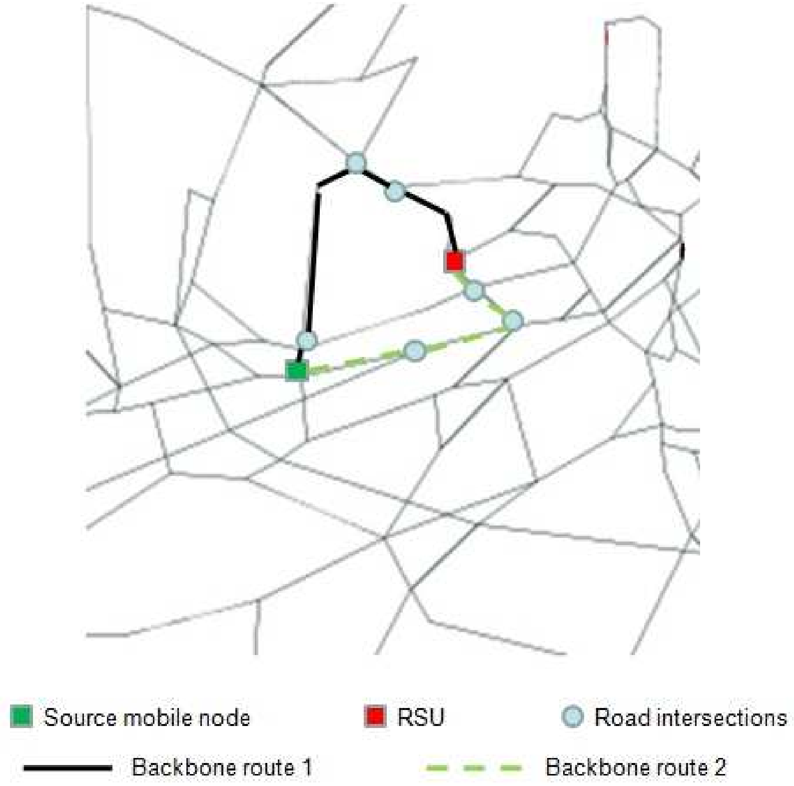


Figure 5.2: Message routing in VANETs

5.2.2 Functionality of AMR

It is worth noting that VANET exhibits a bipolar behavior, where it may have a connected topology with high traffic volume or sparse topology where the traffic volume is low [88]. This implies a variation in the traffic patterns, which AMR aims to mitigate. The variation in the traffic patterns results in variation in the data sent by the mobile nodes. In this case, each RSU will construct backbone routes that adapt to the changes in the data received from the mobile nodes in its vicinity. The RSU uses this data to evaluate statistics such as the average node density on each road segment and the average velocity in that road segment. The RSU uses these statistics to decide on the backbone routes, as will be explained

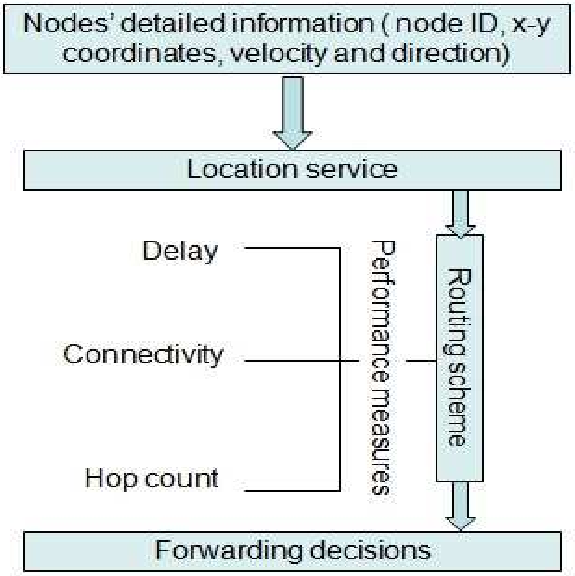


Figure 5.3: Block diagram of message routing

in Section 5.3.

When a mobile node enters for the first time a cell managed by a RSU, it queries its neighbouring vehicles about the optimal backbone route towards the RSU. To illustrate this, let us consider the example depicted in Figure 5.4. Assume that the red car, which enters the cell, moves southward. To send its location updating message to the RSU, there are three feasible backbone routes: A-B-D-F, A-C-D-F, or A-C-E-F. The neighbouring vehicles of the red car already know the optimal route which was selected by the local RSU. So, they reply to the red car's query. The red car uses that optimal backbone route to forward its detailed information to the RSU. This information includes node ID, transmission range, X-Y coordinates, time of last update, and velocity and direction of the mobile node's movement.

The following section presents the analytical framework which AMR uses to derive the end-to-end delay, hop count and connectivity probability.

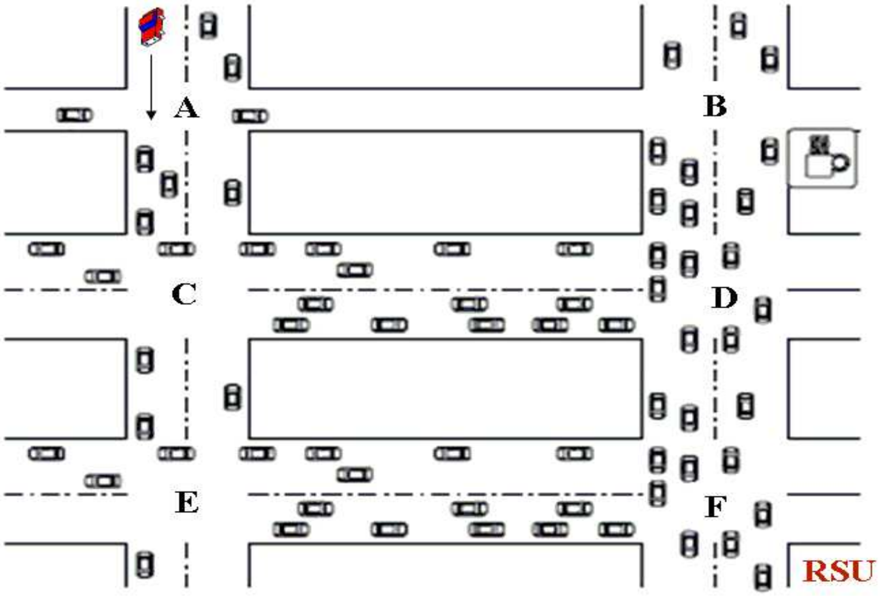


Figure 5.4: Message routing in VANETs using AMR

5.3 Analytical Framework

The road network is modeled as a road graph $G = (V, E)$ consisting of junctions (i.e., intersections of roads) $v \in V$ and road segments $e \in E$ connecting these junctions. AMR considers a two-way road scenario, where the vehicles are moving in two opposite directions on each road segment; the messages are relayed in the same direction of the vehicles' movement direction. Each road segment is modeled as shown in Figure 5.5. The road segment is divided into equal slots. Each slot corresponds to one transmission range T_r. That is, the two lane road is divided into slots according to the transmission range of the nodes.

AMR adapts to changes in the information received from the mobile nodes. In AMR, the local RSU needs to have an up to date view about the local network topology, so that it can update the estimated statistics about each segment in the road graph G. These statistics include:

Table 5.1: List of parameters

Parameter	Description
D	delay in the backbone route
G	road graph
H_{th}	thresholds on the hop count
L	road segment length
m	number of road intersections in one route
n	number of road segments in one route
N	number of nodes in the network
N_g	number of generations for the genetic algorithm
P_{cth}	thresholds on the connectivity probability
p_z	population size
t_p	the time needed for a node to process and transmit a message
t_s	the simulation time
T_r	transmission range
$\widehat{S}$	average speed of nodes on road segment j
α	the ratio between road segment length L and transmission range T_r
β	the portion of the road segment that does not have any node to forward the message
γ	node density
μ	mutation rate
θ	crossover probability

- The average speed of nodes on the segment j (denoted by $\widehat{S}$)

- The average spatial nodes density (denoted by γ_1 and γ_2 for lanes 1 and 2,

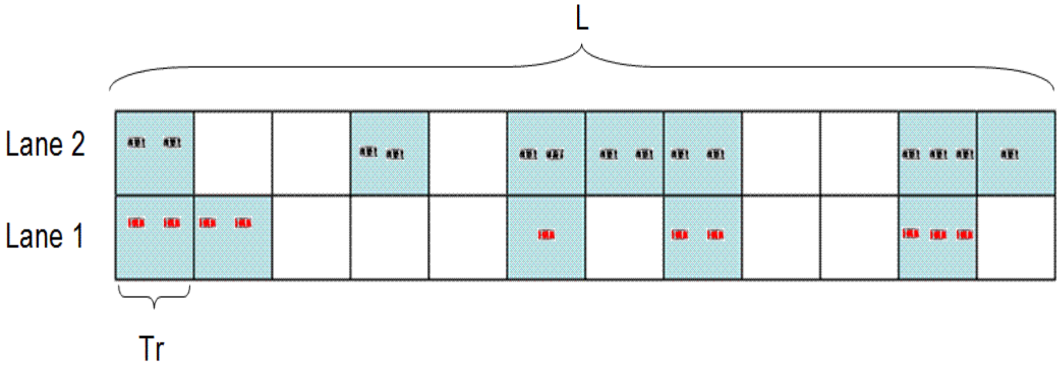


Figure 5.5: Two lane road segment

respectively).

In the following sections, analytical expressions for the delay D, the hop count H_c as well as the connectivity probability P_c of a backbone route y in a two-way road scenario is derived. The backbone route y consists of a number of intersections $v_1, v_2, ..., v_m$ which are connected by a set of road segments $e_1, e_2, ..., e_n$; $n = m - 1$. Table 5.1 describes the parameters used in the analysis.

5.3.1 Delay D

The end-to-end delay of a backbone route y, D, defines how long it takes for an updating or querying message to arrive at the RSU from the time it was sent out by the mobile node. Given the fact that the route y between the mobile node and the RSU consists of a total number of road segments n and each road segment j has an estimated delay D_j , then D can be expressed as:

$$D = \sum_{j=1}^{n} D_j \tag{5.1}$$

The delay D_j depends on the number of mobile nodes N_j traveling on road segment j, and on the time required for a message to be transmitted between the two mobile nods N_i and N_{i+1} which are traveling on road segment j. The time required for a message to travel from a node N_i to node N_{i+1} depends on the strategy N_i uses to forward the message. If N_i uses hop by hop greedy forwarding the delay will be the time needed to process and transmit the message which is denoted as t_p. On the other hand, if N_i uses carry and forward strategy, the message carried by N_i will travel with the same speed S_i as that of the mobile node N_i . In this case the delay depends on S_i and the distance traveled by N_i while carrying the message until it is able to forward the message to the next mobile node N_{i+1}; i.e, when it comes within the transmission range of N_{i+1}.

As a result, in order to estimate the delay D, two cases are considered.

case 1

One vehicle is allowed to forward the message along the road segment. This case occurs if the segment length L is less than one transmission range T_r. Let α be defined as $\alpha = \frac{L}{T_r}$. In this case, $\alpha \leq 1$. The delay on that segment D_j will be t_p, where t_p is the time that the vehicle needs to process and transmit the message on that segment.

case 2

This case occurs when the road segment length is larger than the transmission range (i.e., $\alpha \geq 1$), which is likely to be the case in real networks. In this context, more than one hop is needed to forward the message along that segment.

Let k_1 and k_2 be random variables denoting the number of vehicles that is present in an interval of length T_r on lanes 1 and 2, respectively (see Figure 5.6). Assuming that the vehicles on both lanes are uniformly distributed with the node spatial density γ_1 for lane 1 and γ_2 for lane 2, then k_1 and k_2 are Poisson

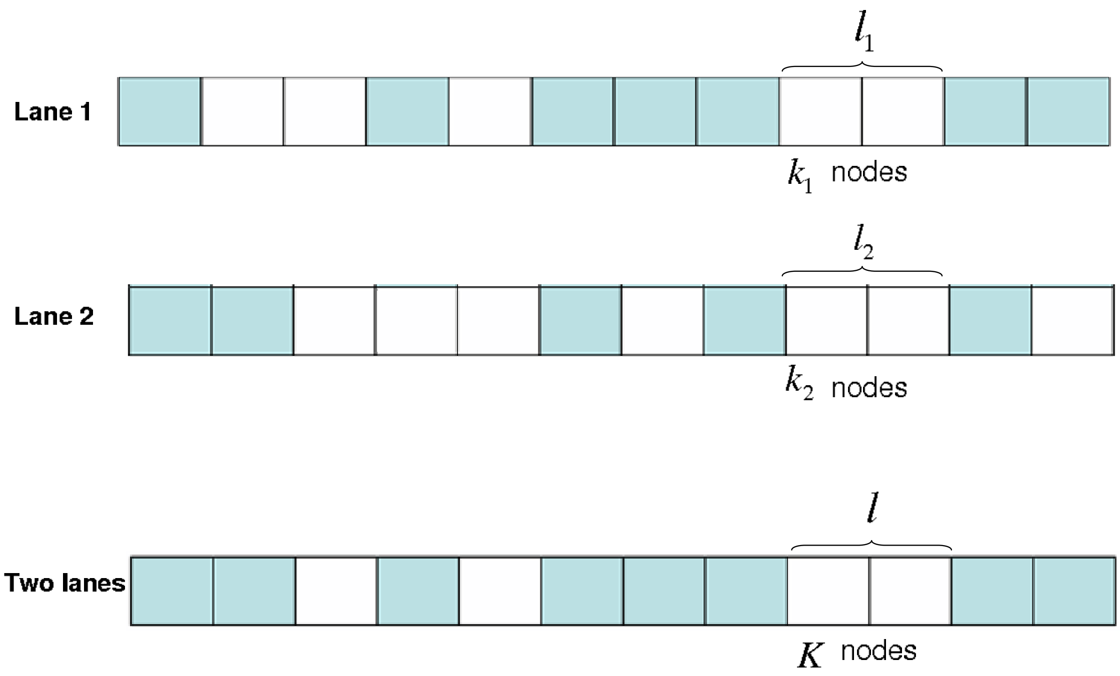


Figure 5.6: Two lane road segment

distributed with the probability mass functions given as follows:

$$f(k_1) = \frac{(\gamma_1 T_r)^{k_1}}{k_1!} e^{-\gamma_1 T_r} \tag{5.2}$$

$$f(k_2) = \frac{(\gamma_2 T_r)^{k_2}}{k_2!} e^{-\gamma_2 T_r} \tag{5.3}$$

Let K be a random variable denoting the number of vehicles present in the interval of length T_r on both lanes, as shown in Figure 5.6. Likewise, K follows a Poisson distribution with the following probability mass function:

$$f(K) = \frac{\big((\gamma_1 + \gamma_2) T_r\big)^K}{K!} e^{-(\gamma_1+\gamma_2) T_r} \tag{5.4}$$

In order to compute the delay on the road segment, the strategy that the mobile node uses to forward the message is considered. If the message is forwarded hop by hop, the delay on each link will be t_p as in the first case. On the other hand, if the message is carried and forwarded by nodes, an estimate of the portion β of the road segment that does not have any node to forward the message is needed. In this case, the last node on that portion receiving the message is allowed to carry and forward the message along that portion. The vehicle will not transmit the message until it comes within the transmission range of another vehicle. This portion (β) can be estimated as :

$$\beta = f(K = 0) = e^{-(\gamma_1+\gamma_2)T_r} \tag{5.5}$$

In this case, the average delay can be computed using the average speed of nodes on the segment (i.e., $\widehat{S}$).

Thus, the average delay on the road segment j can be given as:

$$D_j = \begin{cases} t_p & \text{if } \alpha \leq 1 \\ \alpha(1-\beta)t_p + \beta\frac{L}{\widehat{S}} & \text{otherwise.} \end{cases} \tag{5.6}$$

where,

$$\widehat{S} = \frac{\sum_{k=1}^{N_j} S_k}{N_j} \tag{5.7}$$

is the average speed on the road segment j, and N_j is the number of nodes on the road segment j.

5.3.2 Hop Count H_c

For a given backbone route y the number of hops the message travels on one road segment j is controlled by the length (L) of the road segment and the transmission range T_r of the nodes traveling on that road segment. If L is less than T_r ($\alpha \leq 1$), then one hop will be enough to transmit the message on that road segment. On the other hand, if L is larger than T_r ($\alpha \geq 1$) the message can be transmitted hop by hop or it can be carried and forwarded. Thus, the average hop count on road segment j can be given as:

$$H_{cj} = \begin{cases} 1 & \text{if } \alpha \leq 1 \\ \frac{L}{T_r}(1-\beta) + \beta N_j & \text{otherwise.} \end{cases} \tag{5.8}$$

Accordingly, the hop count of a backbone route y formed by n road segments is given by:

$$H_c = \sum_{j=1}^{n} H_{cj} \tag{5.9}$$

5.3.3 Connectivity Probability P_c

In this work, location updating and querying messages are relayed in the same direction of the vehicles' movement direction, as opposed to the strategy proposed in [89]. To increase the connectivity probability, one may be able to take advantage of the vehicles moving on the opposite direction on a two-way road (see Figure 5.5).

In this context, let us define a broken link between two consecutive vehicles N_i and N_{i+1} as a link with length $l = X_i > T_r$. This broken link is fixable if there are vehicles on the opposite direction within the transmission range of each other and connecting N_i to N_{i+1}. This implies that the distance between any two consecutive vehicles of the new path on lane 2 must be smaller than the transmission range T_r.

Consider k_2, which is a random variable denoting the number of vehicles on lane 2 that is present on an interval of length T_r. Under the assumption that the vehicles on this lane are uniformly distributed with the spatial density γ_2 and using (5.3), the probability P_f that a broken link between two consecutive vehicles N_i and N_{i+1} is fixable can thus be given by:

$$\begin{aligned} P_f &= \prod_{k=1}^{\lfloor X_i/T_r \rfloor} (1 - f(k_2 = 0)) \\ &= (1 - e^{-\gamma_2 T_r})^{\lfloor X_i/T_r \rfloor} \end{aligned} \tag{5.10}$$

Note that the number of vehicles on lane 1 follows a Poisson distribution and the distance X_i between N_i and N_{i+1} is exponentially distributed with parameter γ_1. To compute P_c, one should note that more than one broken link on lane 1 can occur. Let Q be a random variable denoting the number of broken links on lane 1. The road segment will be considered as connected if all the Q links are fixable. Let $P_{c|Q}$ be the conditional connectivity probability given that there are Q broken links. $P_{c|Q}$ can be written as:

$$\begin{aligned} P_{c|Q}(q) &= \prod_{i=1}^{q} P_f \quad \forall q = 0, 1, \ldots, N_j - 1 \\ &= (1 - e^{-\gamma_2 T_r})^{\sum_{i=1}^{q} \lfloor X_i/T_r \rfloor} \\ &= (1 - e^{-\gamma_2 T_r})^{(\alpha - \frac{(N_j - 1 - q)}{\gamma_1 T_r})} \end{aligned} \tag{5.11}$$

To obtain the total connectivity probability of the segment j, it is important to know the probability mass function of Q (i.e., $P_Q(q), \forall q = 0, 1, \ldots, N_j - 1$). Recall that, a link is broken if the distance between any two consecutive vehicles is larger than T_r. Let P_b be the probability that a link q is broken. Since the distance between any two consecutive vehicles is exponentially distributed, it follows that

$$P_b = Pr\{X_i > T_r\} = e^{-\gamma_1 T_r} \tag{5.12}$$

Hence,

$$P_Q(q) = \binom{N_j - 1}{q} \times P_b^q \times (1 - P_b)^{(N_j - 1 - q)} \tag{5.13}$$

Therefore, the total connectivity probability of the road segment j can be expressed as:

$$P_{cj} = \sum_{q=0}^{N_j-1} P_{c|Q}(q) \times P_Q(q) \tag{5.14}$$

Finally, the connectivity probability of the backbone route which is formed by n road segments is given by:

$$P_c = \prod_{j=1}^{n} P_{cj} \tag{5.15}$$

5.4 Minimum End-to-End Delay Routing

This section addresses the problem of finding the optimal or near optimal backbone route y which consists of a number of intersections $v_1, v_2, ..., v_m$ which are connected by a set of road segments $e_1, e_2, ..., e_n$; $n = m - 1$. v_1 is the first intersection in the backbone route that is connected to the source node and v_m is the last intersection in the route that is connected to the RSU. The optimal or near optimal backbone route minimizes the end-to-end delay while satisfying the constraints on both connectivity probability and hop count. The RSU uses this objective function to decide on the backbone routes used by the mobile nodes in its vicinity to route the updating and querying messages.

This approach is formulated as an optimization problem; the objective function is given as follows:

$$\min_{y} D(y) \tag{5.16}$$

$$D(y) = \sum_{j=1}^{n} D_j(y) \tag{5.17}$$

subject to

$$P_c(y) = \prod_{j=1}^{n} P_{cj}(y) \geq P_{cth} \tag{5.18}$$

$$H_c(y) = \sum_{j=1}^{n} H_{cj}(y) \leq H_{th}(5.19)$$

Where, $D(y)$ is the end-to-end delay of route y, and P_{cth} and H_{th} are thresholds on the connectivity probability and hop count, respectively.

The problem of finding a route subject to two constraints, one is additive and the other is multiplicative, is NP hard [90]. In AMR, the optimization problem is constrained by the hop count which is an additive constraint and the probability of connectivity which is a multiplicative constraint. Therefore, this problem is considered NP hard; thus classical optimization methods are not suitable to solve it. To solve this combinatorial nonlinear optimization problem, a heuristic method should be considered.

Simulated Annealing and Tabu Search are considered trajectory (single-solution based) methods that use local search to improve on an initial single solution. Genetic Algorithms, on the other hand, have the additional advantage of combining the good solutions in order to produce possibly better solutions [91]. In addition Genetic Algorithms have been shown to yield better results for routing problems [92, 93].

Therefore, in this thesis, Genetic Algorithm is proposed to address the adaptive message routing problem. The RSU uses the proposed genetic algorithm described in the following section, to decide on the backbone routes. Figure 5.7 displays the flow chart of the proposed algorithm which includes the following components: solution representation, initialization, evaluation, selection, crossover, mutation and termination.

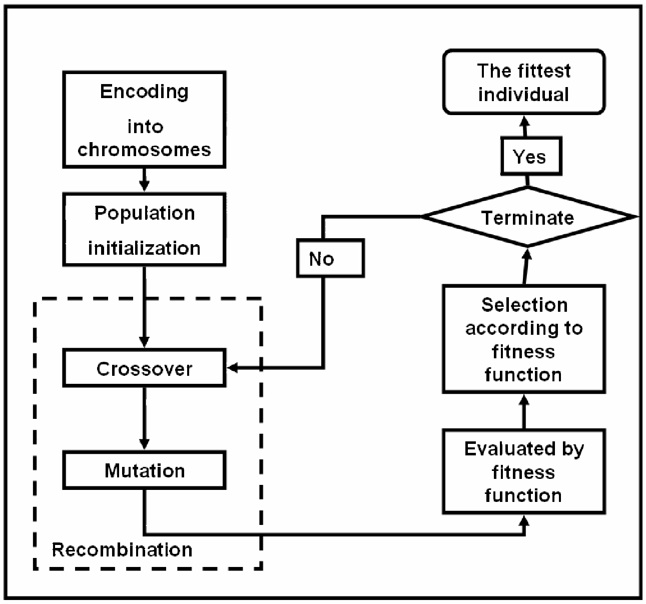


Figure 5.7: Flow chart of the proposed genetic algorithm

5.5 Genetic algorithms

Genetic algorithms (or GAs) are search techniques based on mechanics of natural selection inspired by evolutionary biology such as selection, crossover, mutation and reproduction. Genetic algorithms are used to find an optimal or near optimal solution to optimization problems [94, 95]. An implementation of genetic algorithms begins with a random population of abstract representations of candidate solutions (called individuals). The evolution happens in generations. In each generation, all the individuals are evaluated to allocate reproductive opportunities where the individuals, which represent a better solution to the optimization problem, have more chances to reproduce. Therefore, the goodness of a solution is defined with respect to the individuals in the current generation. In each generation, the selected individuals are modified using crossover and mutation to form a new population which is used in the next generation. The algorithm ends either when a satisfactory level of fitness has been reached for the population or when the maximum number of generations reaches a threshold specified by the user.

5.5.1 Solution Representation and Initialization

Choosing an appropriate representation to encode the feasible solutions is the first step in applying genetic algorithms. This representation should be suitable for the fitness function and the genetic operations. In this approach, a natural encoding scheme would be to define each intersection in the backbone route as a gene. The backbone route consists of the identification number of each selected intersection. Then, the ordered intersections in one route can be represented as a chromosome. Therefore, each feasible solution y consists of one chromosome, denoted as $v_1, v_2, ..., v_m$. For example, routes 1-2-7-8-25, 1-28-27-26-25 and 3-6-9-8-25 in Figure 5.8 are chromosomes. Thus, an individual (or chromosome) is a vector containing the ordered intersections.

GA conducts search from a population of solutions. The initial population is generated by randomly selecting feasible solutions. Each solution or chromosome begins with the intersection adjacent to the mobile node. The next gene is constructed from a randomly selected intermediate intersection. Then, the process randomly chooses the next intermediate intersection in the backbone route until the last intersection in the route is the intersection adjacent to the RSU. It is important to ensure that the solution is feasible. The solution is feasible if it satisfies two conditions: 1) each two consecutive intersections in the route are connected by a backbone link 2) The route satisfies the constraints on the connectivity probability and hop count which are represented in Equations 5.18 and 5.19 respectively. A population of individuals can be constructed by continuing this process until generating a certain number of chromosomes, called population size p_z independently.

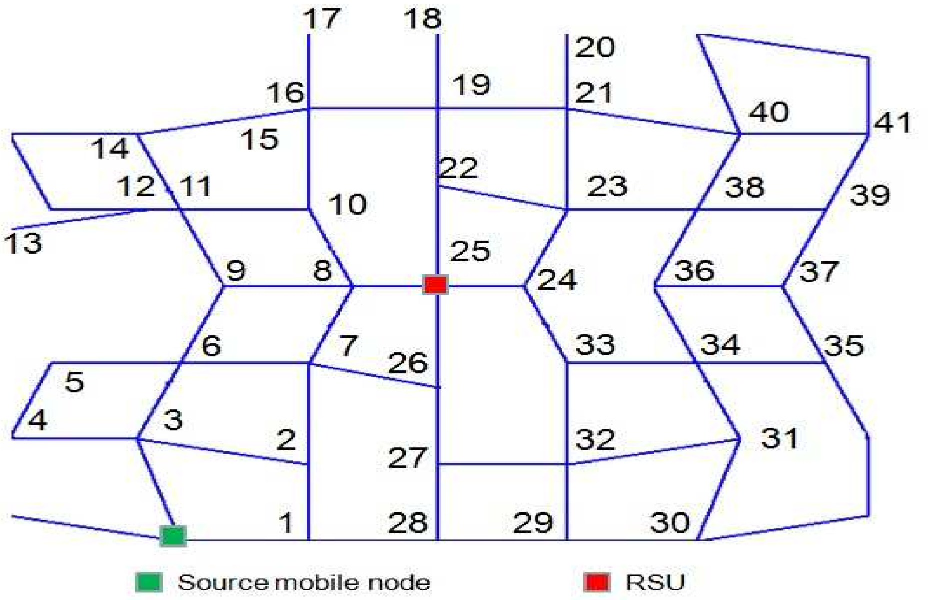


Figure 5.8: Road map used in the simulation

5.5.2 Evaluation

A value for the fitness function $f(y)$ is assigned to each chromosome y depending on how much it is close to solving the problem, and then the best individuals are selected depending on their fitness function. In this analysis, the objective is to minimize the end-to-end delay $D(y)$ given in Equation 5.17 when delivering messages along the backbone route y. Consequently, in order to minimize $D(y)$, the fitness function $f(y)$ is defined as:

$$f(y) = \frac{1}{D(y)} \tag{5.20}$$

5.5.3 Selection

During the selection operation, the quality of the population is improved by giving the high-quality solutions a better chance to produce offsprings that will be part of the next generation. In the implementation, the roulette wheel selection

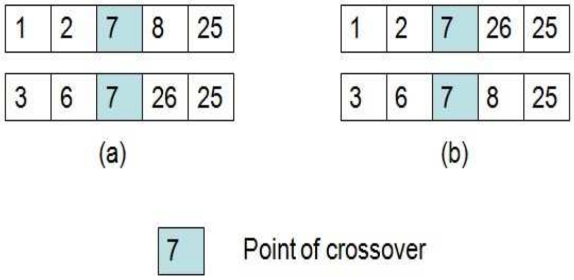


Figure 5.9: One point crossover operator: (a) two selected chromosomes with 7 as crossover point, (b) two new offsprings

strategy is used. Accordingly, the chromosomes are selected based on a probability which is proportional to its normalized fitness value, i.e., probability of choosing a chromosome y corresponds to

$$P_{selection} = \frac{f(y)}{\sum_{y=1}^{p_z} \left(f(y)/p_z \right)} \tag{5.21}$$

where p_z is the population size.

5.5.4 Crossover

The crossover operation is usually executed with a probability θ. One possible crossover operator is the *one point crossover* where two chromosomes are selected from the current population and then a common intermediate gene is randomly selected. That is, the *one point crossover* operator finds an intermediate intersection, called point of crossover, which is common to the two selected routes. Then, it swap the second part of each selected route beyond the point of crossover to form two new offsprings. Figure 5.9 (a) shows two chromosomes with 7 as crossover point and Figure 5.9 (b) shows new two offsprings. It is important to check the new individuals to make sure that they are feasible.

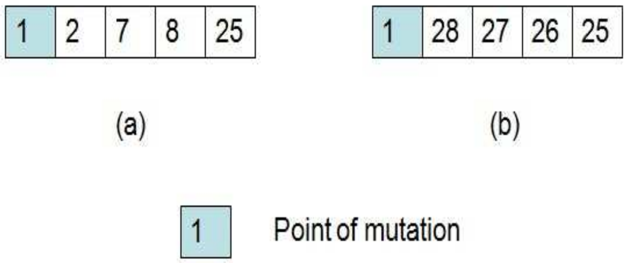


Figure 5.10: Uniform mutation operator: (a) the selected chromosome with 1 as a point of mutation, (b) new offspring

5.5.5 Mutation

Mutation is an operator which causes random changes in the genes inside one chromosome. Therefore, mutation causes diversion in the genes of the current population, which prevents the solution from being trapped in a local optimum. Mutation is performed on the current population with a rate μ. This implementation uses *uniform mutation* operator. So, after choosing any individual from the population with equal probabilities, *uniform mutation* operator randomly picks an intermediate gene (intersection) and then randomly chooses the adjacent intersection. Figure 5.10 (a) shows the selected chromosome and Figure 5.10 (b) shows the new offspring. It is important to make sure that the new individual is a feasible solution.

5.5.6 Termination

The termination criteria in Figure 5.7 can be based on the total number of generations, maximum computing time, or an acceptable threshold of the standard deviation between solutions in one population, or a hybrid termination criteria among them. This implementation uses the maximum number of generations.

Table 5.2: Parameter settings

Parameter	Value	Parameter	Value
t_s	1000 sec	T_r	250m
t_p	3 ms [96]	μ	0.3
P_{cth}	0.5 $\sim$ 0.8 (default 0.5)	θ	0.8
H_{th}	20 $\sim$ 30 (default 30)	N_g	20
S_k	50km/h	p_z	10

5.6 Analytical and Simulation Results

This section compares the proposed adaptive routing scheme AMR to the well known routing protocol (GPCR) [57] through both simulations and analytical approaches. To this end, a discrete-event simulator is developed using Matlab.

The simulation environment consists of a number of road segments which has a number of nodes distributed according to the statistical values of γ_1 and γ_2 for lane 1 and lane 2 respectively. The road map representing the mobility space of the simulated network is shown in Figure 5.8. The experiments consider different vehicle densities representing morning rush hours which is typically characterized by high vehicle density (i.e., dense network), noon time which has intermediate density and the night time which has low density (sparse network). The experiments use different number of mobile nodes, N, which is varied between 100 and 900 nodes, while the area of the simulated network is kept fixed. In addition, the mobility of nodes is modeled based on the given road map. Specifically, if a vehicle reaches an intersection, it chooses the next road segment according to a probability which is proportional to the road segment node density. The vehicle adjusts its velocity to the mean speed allowed on the new road segment. The parameter settings in the experiments are listed in Table 5.2, where t_s denotes the simulation time and N_g is the maximum number of generations for the genetic algorithm.

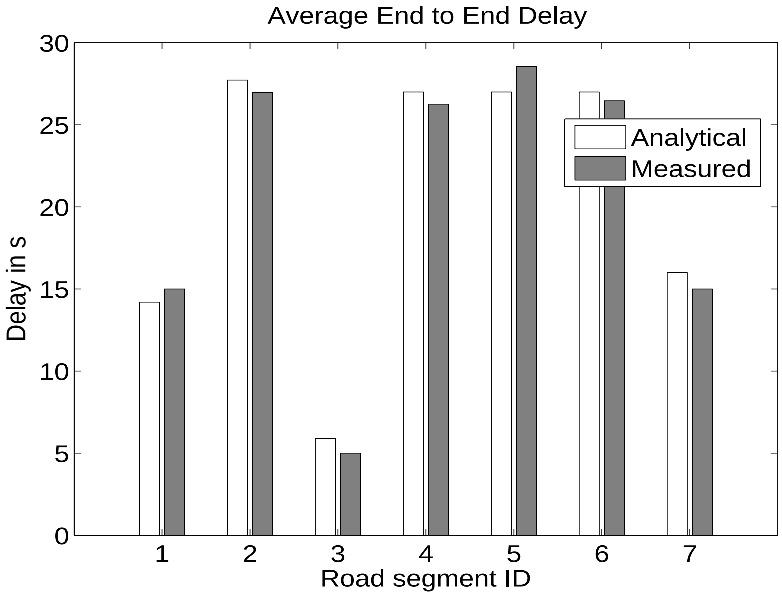


Figure 5.11: Analytical vs. simulation results of the delay in links along the backbone route

To get an insight into the analytical expressions for delay and connectivity probability, let us consider the results shown in Figures. 5.11 – Figure 5.13. It can be observed that the analytical and simulation results are in good agreement, which demonstrates the accuracy of the analytical model.

Let us now focus on the performance comparison of AMR with that of GPCR. Figure 5.14 depicts the end-to-end delay for both protocols as a function of N. As expected, AMR reduces the end-to-end delay, especially when the network is sparse, i.e., when the node density is low. AMR selects routes with higher number of nodes to meet the connectivity probability constraint; which results in less delay. On the other hand, GPCR selects the path with minimum number of intersections, without taking into consideration the connectivity degree. As such, in GPCR, more nodes are allowed to store and carry the packets, and then forward them when coming within the transmission range of the next node; which increases delay. It is worth mentioning that as the number of vehicles increases,

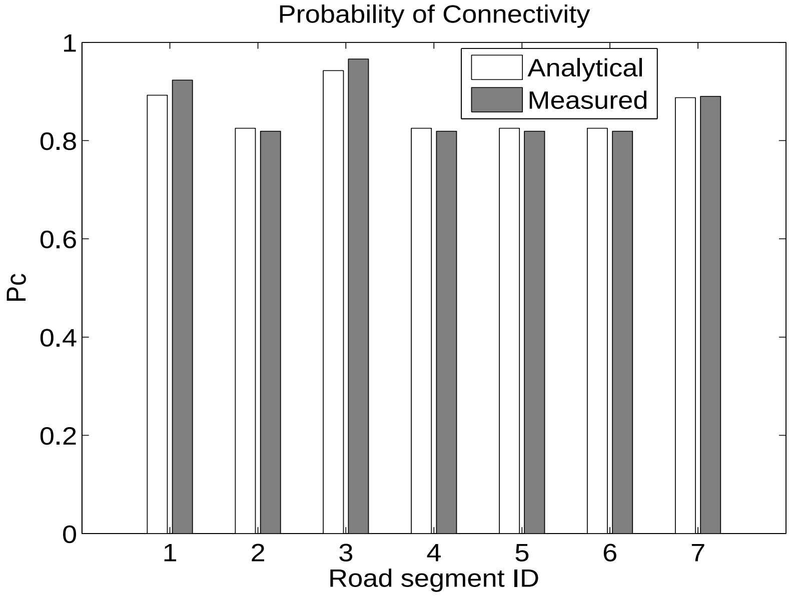


Figure 5.12: Analytical vs. simulation results of the connectivity probability for links along the backbone route

the performance of GPCR improves and becomes close to AMR as depicted in Figure 5.14. This is because vehicles tend to forward messages more using hop by hop forwarding than relying on the carry and forward strategy.

The curves in these figures are expected to have a smoother behavior if optimal solutions are achieved. Nevertheless, GA does not guarantee optimality, but rather gives optimal or near optimal solutions. When the solution of AMR is near optimal the curve experiences jumps from point to another instead of following a smooth behavior.

Figure 5.15 depicts the connectivity probability of the selected routes for the underlying protocols as a function of N. It shows that AMR builds routes with higher connectivity degree than that in GPCR. The performance gap between these two schemes is more significant in sparse networks, where AMR prefers longer paths. This tactic is persuade to attain minimum end-to-end delay where vehicles tend to forward messages more using hop by hop forwarding rather than

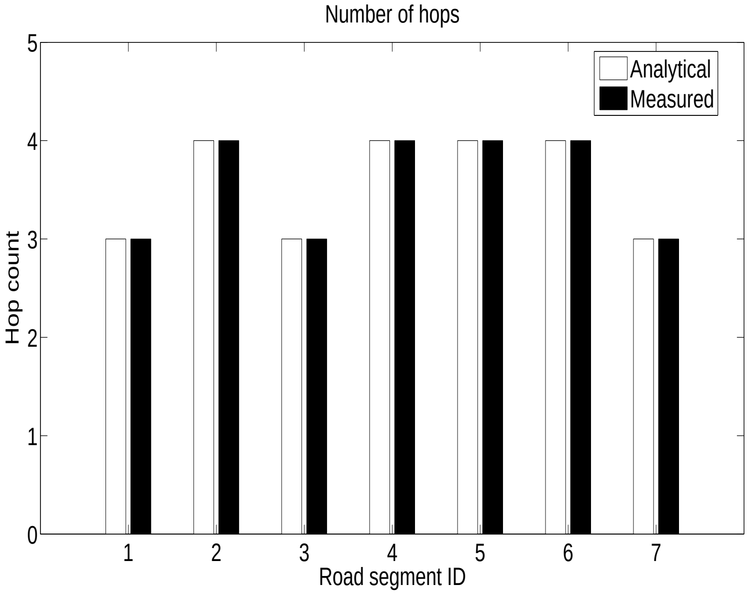


Figure 5.13: Analytical vs. simulation results of the hop count for links along the backbone route

relying on the carry and forward strategy, as explained above. In addition, it can be observed that the connectivity probability increases with the number of vehicles since the nodes become close to each other and can thus use hop by hop forwarding. However, this occurs at the cost of increasing the number of hops in the route in AMR as shown in Figure 5.16. Furthermore, as can be seen from the figure the hop count of routes generated by AMR is higher than that of GPCR.

Figures 5.17 and 5.18 show the effect of the connectivity threshold (i.e., P_{cth}) and the hop count threshold (i.e., H_{th}) on the end-to-end delay, respectively. It is noticed that the delay decreases as the threshold levels increase. This is due to the fact that routes with more and more vehicles are allowed to be selected. This enforces hop by hop forwarding and may result in a shorter delay. Note that variation of the threshold level (i.e., P_{cth} or H_{th}) does not affect the performance of GPCR, since it does not consider these parameters in the routing process.

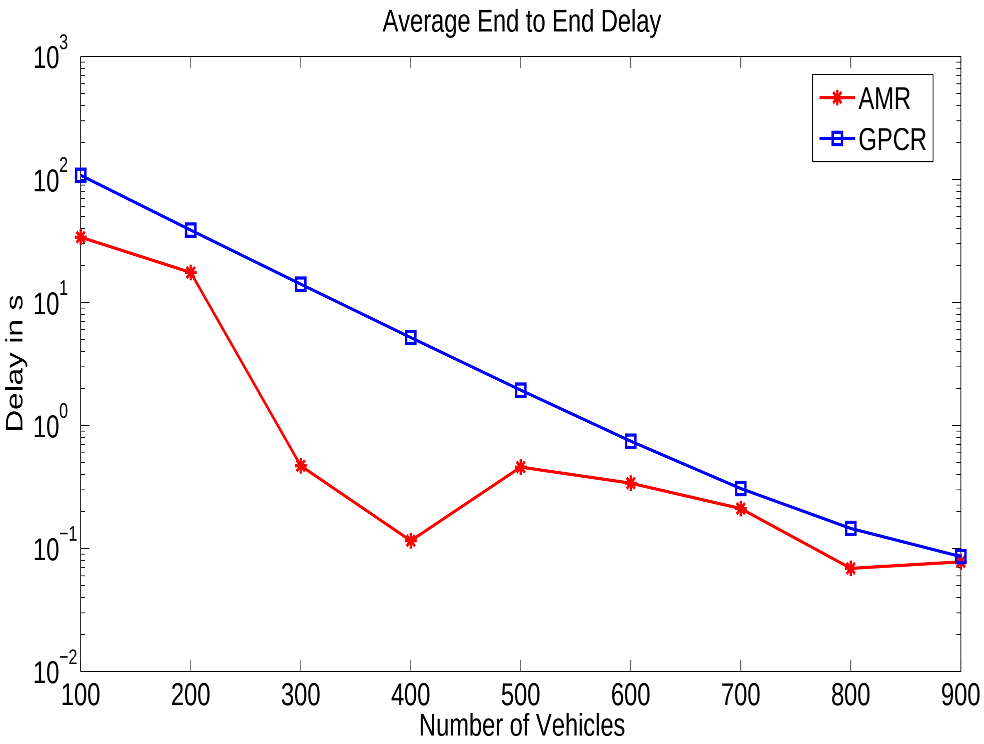


Figure 5.14: End-to-end delay

5.7 Summary

This chapter describes the proposed AMR protocol. It presents the analytical framework used to evaluate the QoS routing problem. In addition, this chapter presents the formulation of the QoS routing problem as an optimization problem and the proposed genetic algorithm to solve it. Analytical and simulation results are also presented.

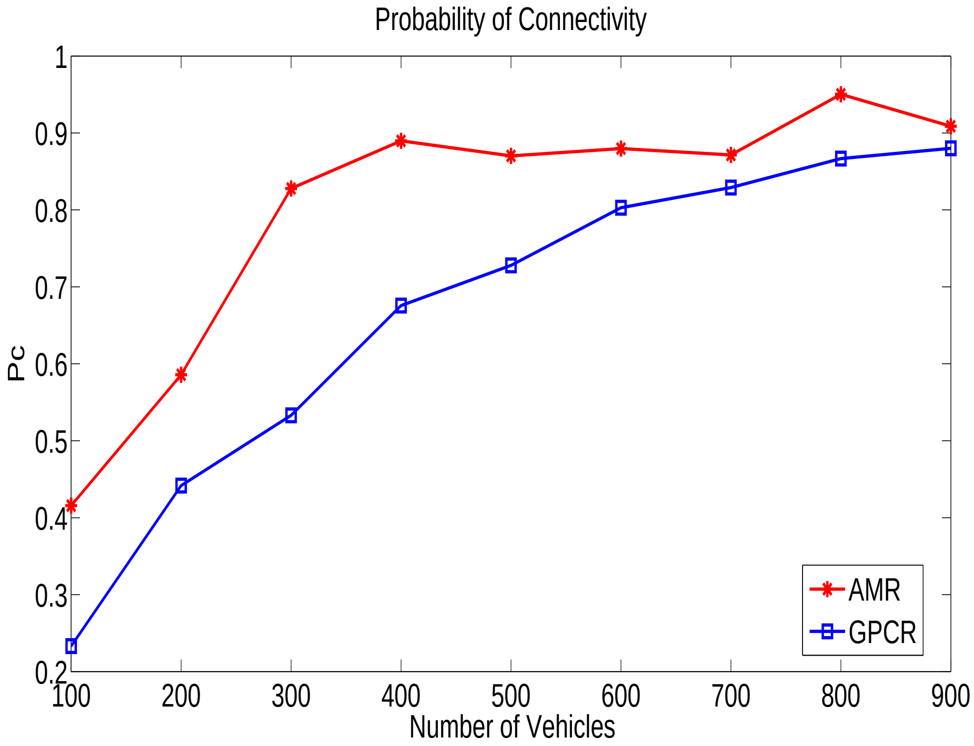


Figure 5.15: Connectivity probability

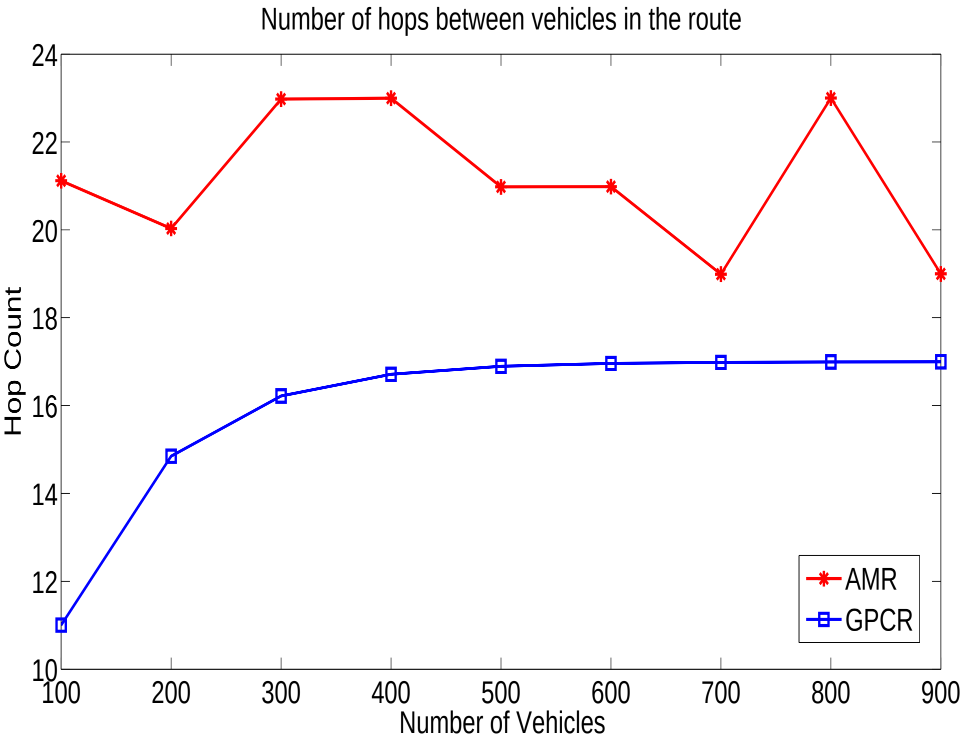


Figure 5.16: Hop count

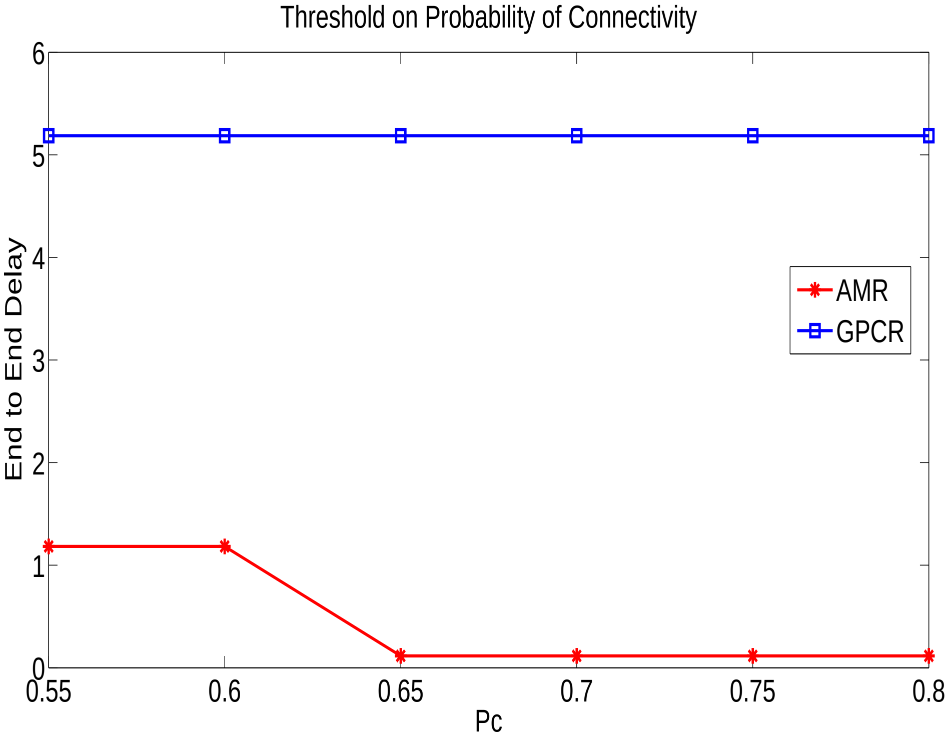


Figure 5.17: Effect of P_{cth} on the end-to-end delay

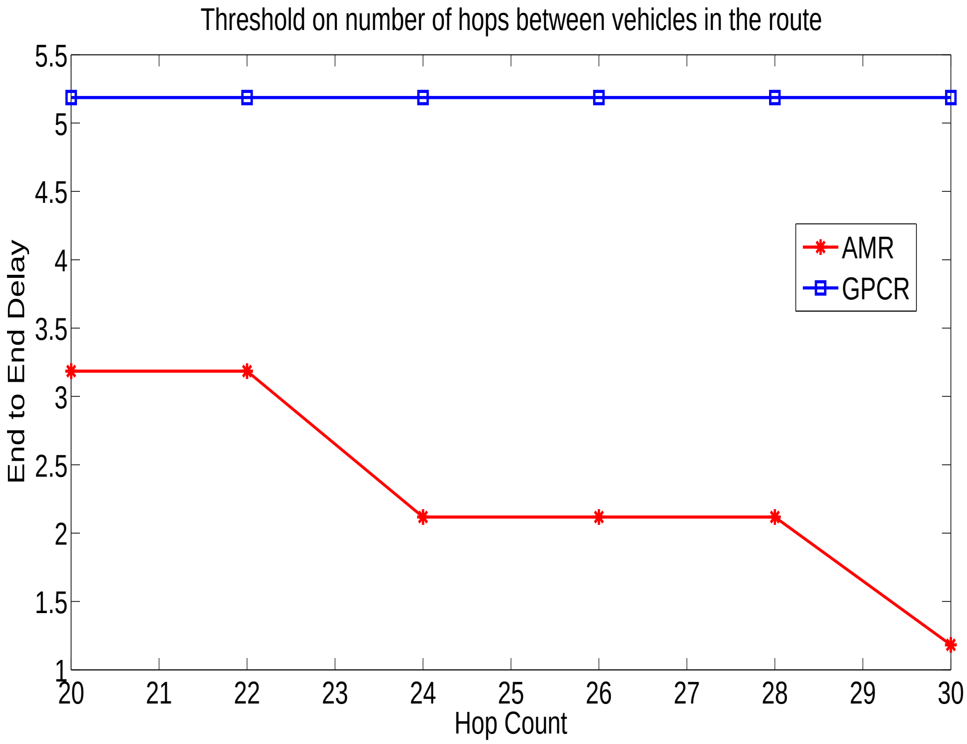


Figure 5.18: Effect of H_{th} on the end-to-end delay

Chapter 6

Conclusions and Future Work

Vehicular Ad-hoc Networks (VANETs) represent a rapidly emerging and challenging class of Mobile Ad Hoc Networks (MANETs). Vehicles form a decentralized communication network by means of wireless multi-hop routing and forwarding protocols. A wide range of applications can be enabled in VANETs. To support the different types of applications, the network must be able to efficiently locate its nodes. While moving, the mobile nodes send location updates to nodes located in specific regions (home regions) in the network called home regions. These nodes (referred to as location servers) are responsible for replying to location queries. An efficient location service management protocol is thus needed to track the locations of mobile nodes (location updates) and reply to location queries with minimum overhead. In this thesis, RLSMP capitalize on quorum based approaches to achieve efficient location service management, by means of node clustering and message aggregation. Thus, the updates and queries of nearby nodes are aggregated. Distinct groups of nodes called clusters are formed based on their geographical locations (i.e., nodes located in one geographical area are grouped in one cluster). In addition, since the communication patterns in VANETs are considered, to a great extent, to be local (i.e., vehicles inside one geographical cluster are more likely to communicate with each other), the proposed service management protocol should be locality aware. Consequently, location servers are assigned based on local traffic patterns considerations.

It is worth noting that VANET exhibits a bipolar behavior, where it may have a connected topology in high traffic volume situations, or sparse topology in low traffic volume situations. This implies variation in the traffic patterns. As such, this thesis aims to mitigate the impact of this variation using Adaptive Message Routing protocol (AMR). Variation in traffic patterns results in variation in the data sent by the mobile nodes to the location servers. In this case, each location server will construct backbone routes that adapt to the changes in the data received from the mobile nodes in its vicinity. The location server uses this data to determine backbone routes.

The rest of the chapter summarizes the major research contributions and proposes future works.

6.1 Major research contributions

The existing location service management protocols are not suitable to achieve high scalability in VANETs. The first main contribution of this research is a scalable, locality aware, and fault tolerant Region-based Location Service Management Protocol (RLSMP). RLSMP uses message aggregation and geographical clustering to reduce the signalling overhead in the network. It also resolves the localization of a destination node by using local search, which begins by exploring the vicinity of the source node. Doing so, it avoids the relatively long distance signalling incurred in other protocols in location updating and querying. In addition, RLSMP propose a mechanism to deal with the event when the current location server moves away from its location and another mobile node becomes the location server, so that the location server should not fail to answer the queries.

As the second main contribution, the total control overhead is formulated as an optimization problem. Analytical and simulation approaches are used to compare the proposed scheme with existing solutions (SLURP, HLS and XYLS). To achieve this, analytical models are developed to evaluate both the location updating and querying costs. In addition, the optimal configuration of the proposed protocol that minimizes the total signalling cost is investigated. It is concluded that RLSMP achieves substantial communication overhead reduction when increasing the network size. RLSMP scheme stands out, therefore, as a promising candidate for large scale wireless ad hoc networks such as VANETs.

As a third main contribution, a new approach for routing delay-sensitive updating and querying messages in vehicular ad hoc networks is proposed. The proposed protocol called AMR uses paths that minimize the end-to-end delay while maintaining a threshold on connectivity degree and a threshold on the hop count along each route. To achieve this, the QoS routing in a two-way road scenario is formulated as a constrained optimization problem. Then, a genetic algorithm is proposed to solve the optimization problem. Using both analytical and simulation approaches, AMR is compared with GPCR. It is concluded that AMR achieves substantial end-to-end delay reduction, especially in sparse networks, a scenario that is particularly challenging in VANETs compared with the task of routing messages in dense networks. The proposed solution stands out, therefore, as a promising candidate for large-scale ad hoc networks such as VANETs.

6.2 Future Work

This research explores the problem of node localization in VANETs and explains how to route updating and querying messages in a timely manner. The node localization problem is mitigated using an efficient location service management protocol. In addition, a routing protocol that builds backbone routes between the nodes and the location server is proposed. Although the research has realized its main objectives, the challenging nature of message routing in VANETs, including the need to enable different applications requires investigation of several issues in order to extend this research work. These issues are as follows:

- Message routing between a source node and a destination node other than the location server. This thesis investigates the problem of routing updating and querying messages between nodes and the location servers. The routing between nodes is planned to be investigated in the future work. The criteria for building these routes depend on the intended application [99] – [104]. For example, if the two nodes exchange text messages, delay can be tolerated. On the other hand, if the two nodes download audio or video multimedia, time is critical and delay is not tolerable. Message routing scheme will depend on the performance metric (application oriented) to be optimized and the QoS constrains that should be taken into consideration.

- Wireless mesh networks: This research considers vehicular ad hoc networks as a vehicle-to-vehicle or a vehicle-to-road side unit network architecture that can be easily deployed without relying on expensive network infrastructure. On the other hand, enabling the communication between the vehicles and pre-existing fixed infrastructure such as gateways to the Internet is an interesting direction of research [97, 98]. One proposed solution is to use wireless mesh networks as a communication backbone that may make it possible to connect the vehicles to the Internet. Wireless mesh networking is a new direction gaining interest today because it can be rapidly deployed.

References

[1] Y. Toor, P. Muhlethaler, A. Laouiti and A. Fortelle, “Vehicular Ad Hoc Networks: applications and related technical issues”, IEEE Communications Surveys and Tutorials, 3rd Quarter 2008.

[2] B. Ducourthial, Y. Khaled, and M. Shawky, “Conditional Transmissions: Performance Study of a New Communication Strategy in VANET”, IEEE Transactions On Vehicular Technology, Vol. 56, No. 6, November 2007.

[3] M. Rozenberg, ”ROVER a Car 2 Car communication protocol”, master thesis, 2006.

[4] W. Kiess and M. Mauve, ”A survey on real-world implementations of mobile ad-hoc networks”, Ad Hoc Networks, vol. 5, no. 3, pp. 324-339, 2007.

[5] V. Naumov, R. Baumann, T. Gross, ”An evaluation of inter-vehicle ad hoc networks based on realistic vehicular traces”, MOBIHOC 2006, Proceedings of the Seventh ACM International Symposium on Mobile Ad Hoc Networking and Computing,2006.

[6] C. Schroth, F. Dotzer, T. Kosch, B. Ostermaier, and M. Strassberger, ”Simulating the traffic effects of vehicle-to-vehicle messaging systems”, In Proceedings of the 5th International Conference on ITS Telecommunications, June 2005.

[7] Q. Liang, Q. Ren, ”Energy and mobility aware geographical multipath routing for wireless sensor networks”, In Proceedings of IEEE Wireless Communications and Networking Conference 2005 (WCNC 2005), New Orleans, 2005.

[8] W. Kiess, J. Rybicki, M. Mauve, “On the nature of inter-vehicle communication”, ACM Workshop on Mobile Ad-Hoc Networks, WMAN 2007.

[9] L. Wischhof, “Self-Organizing Communication in Vehicular Ad Hoc Networks”, PhD. thesis, Hamburg University of Technology, 2007.

[10] C. Lochert, H. Hartenstein, J. Tian, H. Fler, D. Hermann, and M. Mauve, ”A routing strategy for vehicular ad hoc networks in city environments”, In Proceedings of the IEEE Intelligent Vehicles Symposium, 2003.

[11] J. Zhao and G. Cao, “VADD: Vehicle-Assisted Data Delivery in Vehicular Ad Hoc Networks”, IEEE Transactions on Vehicular Technology, Vol. 57, No. 3, May, 2008.

[12] T. Arnold, W. Lloyd, J. Zhao and G. Cao, “IP Address Passing for VANETs”, IEEE International Conference on Pervasive Computing and Communications (PerCom’08), March 2008.

[13] M. Jerbi, R. Meraihi, S. Senouci, Y. Ghamri-Doudane, ”GyTAR: improved greedy traffic aware routing protocol for vehicular ad hoc networks in city environments”, Proceedings of the 3rd international workshop on Vehicular ad hoc, Pages: 88 - 89, 2006.

[14] S. Panichpapiboon, and W. Pattara-atikom, “Connectivity Requirements for Self-Organizing Traffic Information Systems”, IEEE Transactions On Vehicular Technology, Vol. 57, No. 6, November 2008.

[15] World Health Organization, “http://www.who.int/features/factfiles/roadsafety/04_en.html.“

[16] J. Zhao, Y. Zhang, and G. Cao, “Data Pouring and Buffering on the Road: A New Data Dissemination Paradigm for Vehicular Ad Hoc Networks”, IEEE Transactions on Vehicular Technology, Vol. 56, No. 6, November 2007.

[17] C. Adler, ”Information Dissemination in Vehicular Ad Hoc Networks”, PhD thesis, Technische Universitt Mnchen, Fakultt fr Informatik, 2006.

[18] F. Bai, T. Elbatt, G. Hollan, H. Krishnan, and V. Sadekar, ”Towards characterizing and classifying communication-based automotive applications from a wireless networking perspective”, 1st IEEE Workshop on Automotive Networking and Applications, 2006.

[19] D. Sormani, G. Turconi, P. Costa, D. Frey, M. Migliavacca and L. Mottola, ”Towards lightweight information dissemination in inter-vehicular networks”, Proceedings of the 3rd international workshop on vehicular ad hoc networks Pages: 20 - 29 , 2006

[20] S. Philip, J. Ghosh, and C. Qiao, ”Performance evaluation of a multilevel hierarchical location management protocol for ad hoc networks”, Elsevier Computer Communications, Special issue on Performance Issues of Wireless LANS, PANs, and Ad Hoc Networks.

[21] R. Morris, J. Jannotti, F. Kaashoek, J. Li, D. Decouto, ”CarNet: a scalable ad hoc wireless network system”, 9th ACM SIGOPS European Workshop, 2000.

[22] C. Cheng, H. Lemberg, S. Philip, E. Berg, and T. Zhang, ”SLALoM: A scalable location management scheme for large mobile ad-hoc networks”, In Proceedings of IEEE WCNC, March 2002.

[23] S. Das, H. Pucha, and Y. Hu, ”Performance comparison of scalable location services for geographic ad hoc routing”, In Proceedings of IEEE INFOCOM, Miami, FL, 2005.

[24] R. Melamed, I. Keidar, and Y. Barel, ”Octopus: a fault-tolerant and efficient ad-hoc routing protocol”, In the 24th IEEE Symposium on Reliable Distributed Systems (SRDS 2005), October 2005.

[25] I. Stojmenovic, ”Location updates for efficient routing in ad hoc networks, ”Handbook of Wireless Networks and Mobile Computing, Wiley, pp. 451-71, 2002.

[26] W. Wang, F. Xie; M. Chatterjee, “TOPO: Routing in Large Scale Vehicular Networks”, Vehicular Technology Conference, IEEE VTC- Fall, 2007.

[27] Y. Yu, ”Efficient and scalable lookup services in large scale p2p and mobile ad hoc networks”, Master thesis, University of Minnesota, 2004.

[28] Y. Yu, G. Lu, and Z. Zhang, ”Enhancing Location Service Scalability With HIGH-GRADE”, University of Minnesota, Technical Report TR-04-002, 2004.

[29] R. Friedman and G. Kliot, ”Location Services in Wireless Ad Hoc and Hybrid Networks: A Survey”, Technical Report CS-2006-10, Department of Computer Science, The Technion. April 2006.

[30] B. N. Karp and H. T. Kung, “GPSR: Greedy Perimeter Stateless Routing for Wireless Networks”, ACM MOBICOM’00, pp. 243-254, Boston, Aug. 2000.

[31] C. Perkins and E. Royer, Ad-hoc on demand distance vector routing, in Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications, pp. 90100, Feb. 1999.

[32] D. Johnson and D. Maltz, “Dynamic source routing in ad hoc wireless networks”, Mobile Computing, 1996.

[33] S. Basagni, I. Chlamtac, V. Syrotiuk, and B. Woodward, “A distance routing effect algorithm for mobility (DREAM)”, ACM MOBICOM 1998.

[34] J. Bernsen and D. Manivannan, ”Routing Protocols for Vehicular Ad hoc Networks that ensure Quality Of Service”, In Proc. ACM ICWMC, 2008.

[35] F. Bai, N. Sadagopan, and A. Helmy, ”The IMPORTANT framework For analyzing the impact of mobility on performance of routing protocols for ad hoc networks”, Ad Hoc Networks Journal-Elsevier, 1:383-403,”November” 2003.

[36] H. Fler, ”Position-based packet forwarding for vehicular ad-hoc networks” PhD thesis, Universitt Mannheim, 2007.

[37] T. Taleb, M. Ochi, N. Kato, and Y. Nemoto, ”An efficient vehicle-heading based routing protocol for vanet networks”, Proc. of WCNC, 2006.

[38] C. Lochert, ”A routing strategy for vehicular ad hoc network in the city environments”, Proc. of IEEE Intelligent Vehicles Symposium, 2003.

[39] K. C. Lee, J. Haerri, U. Lee, and M. Gerla, ”Enhanced perimeter routing for geographic routing protocols in urban vehicular scenarios”, IEEE Autonet Workshop, 2007.

[40] M. Mauve, H. Fuessler, J. Widmer, and T. Lang, ”Positionbased multicast routing for mobile ad-hoc networks”, Technical Report TR-03-004, Department of Computer Science, University of Mannheim, 2003.

[41] D. Johnson and D. Maltz, ”Dynamic source routing in ad hoc wireless networks”. In Imielinski and Korth, editors, Mobile Computing, volume 353. Kluwer Academic Publishers, 1996.

[42] CAR 2 CAR communication consortium, “http://www.car-2-car.org“

[43] R. Morris, J. Jannotti, F. Kaashoek, J. Li, D. Decouto, ”CarNet: a scalable ad hoc wireless network system”, 9th ACM SIGOPS European Workshop, 2000.

[44] FleetNet project, “http://www.et2.tuharburg.de/fleetnet.“

[45] C3 Project, “http://www.aqualab.cs.northwestern.edu/ projects /C3.html.“

[46] T. Camp, J. Boleng, and L. Wilcox, “Location information services in mobile ad hoc networks”, IEEE ICC 2001, pp. 3318.3324.

[47] M. Ksemann, H. Fler, H. Hartenstein, and M. Mauve, “A reactive location service for mobile ad hoc networks”, Technical report, Department of Science, University of Mannheim, TR-02-014, November 2002.

[48] S. Woo, S. Singh, “Scalable routing protocol for ad hoc networks”, Wireless Networks, vol. 7, pp. 513-529, Kluwer Publishers, 2001.

[49] R. Flury and R. Wattenhofer, “MLS: an efficient location service for mobile ad hoc networks”, ACM MOBIHOC, Italy, May 2006.

[50] S. Das, H. Pucha, and Y. Hu, “Performance comparison of scalable location services for geographic ad hoc routing”, IEEE INFOCOM, Miami, FL, 2005.

[51] D. Liu , I. Stojmenovic and X. Jia, “A scalable quorum based location service in ad hoc and sensor networks,” IEEE MASS, 2006.

[52] W. Kiess, H. Fssler, J. Widmer, M. Mauve, “Hierarchical Location Service for Mobile Ad-Hoc Networks”, ACM MC2R, vol. 8, no. 4, Oct. 2004.

[53] P. Basu, N. Khan, and T.D.C. Little, ”A mobility based metric for clustering in mobile ad hoc networks,” In Proceedings of Distributed Computing Systems Workshop, pp.413-418, April 2001.

[54] R. Santos, R. M. Edwards, N. Seed and A. Edwards. A , ”Location-based routing algorithm for vehicle to vehicle communication”. The 13th IEEE International Conference on Computer Communication and Networks (ICCCN 2004), pp. 221-226, 2004.

[55] C. Perkins and P. Bhagwat, ”Highly dynamic destination-sequenced distance-vector routing (DSDV)”, In Proceedings of the Conference on Communications architectures, protocols and applications (SIGCOMM ’94), 1994.

[56] Z. J. Haas and M. R. Haas, ”The performance of query control schemes for the zone routing protocol”, IEEE/ACM Transactions on Networking (TON), 9(4):427-438, August 2001.

[57] C. Lochert, M. Mauve, H. Fusler, and H. Hartenstein. ”Geographic routing in city scenarios”, ACM SIGMOBILE Mobile Computing and Communications Review, 2005.

[58] Z. Mo, H. Zhu, K. Makki, N. Pissinou, ”a multi-hop routing protocol for urban vehicular ad hoc networks”, Third Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, 2006.

[59] A. Skordylis, N. Trigoni, ”Delay-bounded routing in vehicular ad-hoc networks”, MOBIHOC 2008.

[60] A. Nandan, S. Das, B. Zhou, G. Pau, and M. Gerla, "AdTorrent: digital billboards for vehicular networks", IEEE/ACM International Workshop on Vehicle-to-Vehicle Communications (V2VCOM), 2005.

[61] C. Perkins, "Ad hoc networking", Addison Wesley, 2001.

[62] K. Xu, M. Gerla, and S. Bae, "How effective is the IEEE 802.11 RTS/CTS handshake in ad hoc networks?", In Proc. GLOBECOM 2002, Taipei, November 2002.

[63] A. Tanenbaum, "Computer networks", Third edition, Prentice Hall PTR, 1996.

[64] J. Li, C. Blake, D. S.J. De Couto, H.I. Lee and R. Morris, "Capacity of ad hoc wireless networks", ACM MOBICOM 2001, pp. 61-69, Rome, Italy, July 2001.

[65] I. Abraham, D. Dolev, and D. Malkhi, "LLS: a locality aware location service for mobile ad hoc networks", In Proc. of the Joint Workshop on Foundations of Mobile Computing ($DIALM - POMC$), pages 75-84. ACM Press, 2004.

[66] T. Camp, J. Boleng, and L. Wilcox, " Location information services in mobile ad hoc networks", In Proceedings of the IEEE International Conference on Communications (ICC)), pages 3318.3324, 2001

[67] J. Li, J. Jannotti, D. De Couto, D. Karger, and R. Morris, "A scalable location service for geographic ad-hoc routing", In Proceedings of ACM MobiCom, pages 120.130, August 2000.

[68] Y. Xue, B. Li, and K. Nahrstedt, "A scalable location management scheme in mobile ad-hoc networks", In Proceedings of the IEEE Conference on Local Computer Networks, 2001.

[69] C. Perkins, "Ad hoc networking", Addison Wesley, 2001.

[70] H. Saleet, R. Langar, O. Basir, and R. Boutaba, "A Distributed Approach for Location Lookup in Vehicular Ad Hoc Networks", IEEE ICC'09, Germany, 2009.

[71] C. Lochert, B. Scheuermann, et al., "Data Aggregation and Roadside Unit Placement for a VANET Traffic Information System", In Proc. ACM VANET 2008.

[72] L. Briesemeister, G. Hommel. "Role-based multicast in highly mobile but sparsely connected ad hoc networks", In Proc. ACM MOBIHOC 2000.

[73] H. Hartenstein, B. Bochow, A. Ebner, M. Lott, M. Radimirsch, and D. Vollmer. "Position-aware ad hoc wireless networks for inter-vehicle communications: the Fleetnet project", In Proc. ACM MOBICOM 2001.

[74] H. Wu, R. Fujimoto, R. Guensler, and M. Hunter. "MDDV: a mobilitycentric data dissemination algorithm for vehicular networks", In Proc. of ACM VANET 2004.

[75] H. Saleet, O. Basir, R. Langar, and R. Boutaba, "Region-Based Location Service Management Protocol for VANETs", IEEE Trans. Vehicular Technology, to appear.

[76] H. Saleet, R. Langar, O. Basir, and R. Boutaba, "Proposal and Analysis of Region-based Location Service Management Protocol for VANETs ", In Proc. IEEE GLOBECOM 2008.

[77] H. Hartenstein, K. Laberteaux, "A Tutorial Survey on Vehicular Ad Hoc Networks", IEEE Communications Magazine, June 2008.

[78] Yunpeng Zang, L. Stibor, B. Walke,H. Reumerman,A. Barroso, "Towards Broadband Vehicular Ad-Hoc Networks - The Vehicular Mesh Network (VMESH) MAC Protocol", IEEE WCNC'07, Hong Kong, March 2007.

[79] I. F. Akyildiz, Y.-B. Lin, W.-R. Lai, and R.-J. Chen "A New Random Walk Model for PCS Networks" IEEE JSAC, vol. 18, no. 7, pp. 1254-1260, July 2000.

[80] K-H. Chiang, N. Shenoy, "A 2-D Random-Walk Mobility Model for Location-Management Studies in Wireless Networks", IEEE Transactions on Vehicular Technology, Vol. 53, No. 2, pp. 413-424, Mar. 2004.

[81] D. N. Alparslan, K. Sohraby, "Two-Dimensional Modeling and Analysis of Generalized Random Mobility Models for Wireless Ad Hoc Networks", IEEE/ACM Trans. Networking, vol. 15, no. 3, pp. 616-629, June 2007.

[82] T. Spyropoulos, K. Psounis, and C. Raghavendra, “Efficient Routing in Intermittently Connected Mobile Networks: The Single-Copy Case”, IEEE/ACM Trans. Networking, vol. 16, no. 1, pp. 63-76, February 2008.

[83] F. Bai , N. Sadagopan and A. Helmy , “The IMPORTANT framework for analyzing the Impact of Mobility on Performance Of RouTing protocols for Adhoc NeTworks”, Ad Hoc Networks, Elsevier, 2003.

[84] G. Rubino and B. Sericola, “A finite characterization of weak lumpable Markov processes. Part II: the continuous time case”, *Stochastic Processes and their Applications*, vol. 45, 1993.

[85] J.A. Barria, R.M. Lent, “MANET Route Discovery using Residual Lifetime Estimation”, IEEE ISWPC 2006.

[86] M. Piorkowski, N. Sarafijanovoc-Djukic and Matthias Grossglauser , “A Parsimonious Model of Mobile Partitioned Networks with Clustering”, in Proc. IEEE/ACM COMSNETS’09, India, Jan. 2009.

[87] H. Saleet, R. Langar, O. Basir, and R. Boutaba, “Adaptive Message Routing with QoS Support in Vehicular Ad Hoc Networks ”, In Proc. IEEE GLOBECOM 2009.bipolar

[88] N. Wisitpongphan, F. Bai, P. Mudalige and O. Tonguz, ” On the Routing Problem in Disconnected Vehicular Ad Hoc Networks”, In Proc. IEEE INFOCOM 2007.

[89] S. Panichpapiboon and W. Pattara-atikom, ”Connectivity Requirements for Self-Organizing Traffic Information Systems”, Vehicular Technology, IEEE Transactions, 2008.

[90] Z. Wang, J. Crowcroft, ”Quality-of-Service Routing for Supporting Multimedia Applications”, IEEE Journal on Selected Areas in Communications 14(7): 1228-1234 (1996).

[91] X. Cheng, ”Multiple Description Video Communications in Wireless Ad Hoc Networks”, master thesis, 2005.

[92] W. Mann and G. Smith, ”A Comparison of Heuristics for Telecommunications Traffic Routing”, In Rayward-Smith et al. (editors), Modern Heuristic Search Methods. John Wiley Sons, 1996.

[93] J. Schaffer and L. Eschelman, ”Combinatorial Optimization by Genetic Algorithms: The Value of the Genotype/Phenotype Distinction”, In Rayward-Smith et al. (editors), Modern Heuristic Search Methods. John Wiley Sons, 1996.

[94] D. Goldberg, ”Genetic algorithms in search, optimization, and machine learning”, Addison-Wesley, 1989.

[95] J. Lopez, M. Imine and O. Madsen, “Network Planning Using GA For Regular Topologies”, In Proc. IEEE ICC 2008.

[96] S. Haykin and M. Moher, ”Modern wireless communication”, Prentice Hall, 2005.

[97] C. Hung, H. Chan, and E. H. Wu, ”Mobility Pattern Aware Routing for Heterogeneous Vehicular Networks”, WCNC 2008.

[98] V. Namboodori, M. Agarwal, and L. Gao, ”A study on the feasibility of mobile gateways for vehicular adhoc networks”, Proc. of VANET, 2004.

[99] Network Coding Homepage ”http://www.icsi.berkeley.edu/l̃uby/”

[100] ”http://www.ifp.illinois.edu/k̃oetter/NWC/index.html”

[101] ”http://www.mit.edu/ medard/coding1.htm”

[102] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, ”Network Information Flow”, (IEEE Transactions on Information Theory, IT-46, pp. 1204-1216, 2000).

[103] A. H. Mohsenian-Rad, J. Huang, V. W.S. Wong, S. Jaggi, and R. Schober, ”A Game-Theoretic Analysis of Inter-Session Network Coding”, ICC’09, Germany, 2009.

[104] T. Ho, R. Koetter, M. Mdard, D. R. Karger and M. Effros, ”The Benefits of Coding over Routing in a Randomized Setting”, IEEE International Symposium on Information Theory, 2003.